PERCEPTION AND PRODUCTION OF EMOTIONAL PROSODY IN THE SPEECH OF MANDARIN-SPEAKING ADULTS WITH COCHLEAR IMPLANTS

by

Cecilia Liu Pak

APPROVED BY SUPERVISORY COMMITTEE:

____

William F. Katz, Chair

____

Peter F. Assmann

____

Andrea Warner-Czyz

____

Jun Wang

Copyright 2018

Cecilia Liu Pak

All Rights Reserved

Dedicated to my parents, Peilin Liu and Miao Lin.

PERCEPTION AND PRODUCTION OF EMOTIONAL PROSODY IN THE SPEECH OF MANDARIN-SPEAKING ADULTS WITH COCHLEAR IMPLANTS

by

CECILIA LIU PAK, MS

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

COMMUNICATION SCIENCES AND DISORDERS

THE UNIVERSITY OF TEXAS AT DALLAS

December 2018

ACKNOWLEDGMENTS

Pursuing a PhD has always been my dream, and I would like to express my gratitude to many people for their support and contributions in helping me make this dream come true. First, I would like to acknowledge Professor William Katz, my research advisor, for his professional guidance, expertise, patience and support in advising an international student to conquer the difficulties of this dissertation journey. I would also like to thank Professor Emily Tobey, for inspiring me to pursue my PhD in the fields of hearing science and cochlear implant research at The University of Texas at Dallas. Many thanks to my committee members, Professor Andrea Warner-Czyz, Professor Peter Assmann and Professor Jun Wang, for their valuable insight and feedback throughout this project. Thank you to Professor Andrea Warner-Czyz for providing clinical insight on cochlear implants, to Professor Peter Assmann for helping with acoustic cues synthesis using MATLAB, and to Professor Jun Wang for guidance in machine learning techniques. I would also like to thank Professor Yang Liu for providing the emotional database to support my research. Many thanks are owned to Dr. Hua Zhang and Wen Sun from the Beijing Institute of Otolaryngology, for their support and help in recruiting Mandarin-speaking CI adults. I also thank the members of Professor Katz's lab in the graduate program in Communication Sciences and Disorders, including Sonya Mehta for helping analyze data and providing training in new techniques using DirectRT and Praat, and to Evdoxia Doli, Kelly Fazel, Sujin Shin, and Annette Glotfelty for helpful feedback and friendship. I would also like to thank my friends, Jenson Lam for his statistical help, other Chinese students at UT Dallas for their support and participation, and especially, all the Mandarin-speaking adults with CIs for their participation. Thanks also to the PhD Research Small Grant Award and the Jerger

Fellowship in Audiology for providing research funding. Finally, my deepest gratitude goes to my parents, Miao Lin and Peilin Liu, for always believing in me. This accomplishment would not have been possible without their unconditional love, unwavering support, and encouragement.

October 2018

PERCEPTION AND PRODUCTION OF EMOTIONAL PROSODY IN THE SPEECH OF MANDARIN-SPEAKING ADULTS WITH COCHLEAR IMPLANTS

Cecilia Liu Pak, PhD
The University of Texas at Dallas, 2018

Supervising Professor: William F. Katz

Emotional prosody, which refers to the process of expressing emotions through spoken language, is essential for correctly recognizing speakers' emotional states during spoken communication. Previous research has shown that non-tonal language-speaking individuals with cochlear implants (CIs) demonstrate deficits in perceiving and producing emotional prosody, as compared to their typical hearing (TH) counterparts. These research, however, did not explore how well tonal language-speaking individuals (e.g., Mandarin) with CIs perceive and produce emotional prosody in speech, in comparison to their TH counterparts. Additionally, no data are available to clarify whether CI adults who speak tonal languages differ from those who speak non-tonal languages with respect to the extent of emotional prosodic processing. These concerns were addressed in this dissertation through four experiments. The first experiment explores the differences in emotional prosody perception between 15 TH adults and 15 CI adults. All were native Mandarin-speaking adults. The TH listeners were required to listen to natural speech stimuli and noise-vocoded speech stimuli designed to stimulate CI input. The CI adults were only asked to listen to natural speech. The results showed that overall emotional prosody recognition by TH and CI listeners for natural speech is 72.8% and 50.3%, respectively. These findings suggest that CI adults demonstrate deficits in perceiving emotional prosody. TH listeners performed better with natural speech than with noise-vocoded speech, and their intelligibility was lower when the number of noise-vocoded filter channels was reduced. In addition, the performance of CI listeners in natural speech was similar to that of TH listeners at a lower channel setting (at 4-channel), in contrast to 8-channel shown in previous comparable studies of non-tonal languages (e.g., English). This finding provide evidence consistent with a "functional view" hypothesis, which claims that Mandarin (a tonal language that uses pitch for purposes of linguistic tone) has relatively little prosodic space to signal emotional prosody through the pitch dimension. The second experiment was intended to determine whether enhancement of secondary cues (duration and amplitude) can benefit CI listeners to perceive emotions. This was explored by modifying the prosodic cues for two contrasting emotions, "happy" and "sad", and observing how the CI listeners perceived these modifications. The result showed that increased duration cues can slightly improve recognition of the "sad" emotion and increased amplitude can improve identification of the "happy" emotion. These findings suggest that the selected enhancement of secondary cues could potentially benefit CI listeners. The third experiment investigated whether TH and CI talkers differ in terms of acoustic cue production and examined which acoustic measures are most predictive of emotions produced by these talkers. This was done by analyzing the fundamental frequency (F_0), intensity, and duration patterns of short sentences spoken by the TH and CI talkers in the "angry", "happy" and "sad" emotional contexts. The results suggested that CI talkers showed decreased mean intensity, increased intensity range, and sentence duration values in their emotional prosody productions compared to their TH counterparts. In addition, a machine learning (decision tree) model of emotion classification was used to analyze which acoustic measures were most predictive of three emotions produced by TH and CI talkers. The results indicated that TH talkers utilize intensity as the most important classifier, followed by F_0 to predict the three emotions, while CI talkers used duration as the most important classifier, followed by intensity. The findings of model indicated that the secondary cues (duration and intensity) are most predictive in classifying the three emotions in the CI talkers' productions. The fourth experiment examined the production data in a perceptual manner to determine whether the deficits of CI talkers described acoustically in Experiment 3 could be perceived by TH listeners. The results confirm that CI users show impaired emotional prosody production, and this deficit is reflected in lowered perception scores by TH listeners. In addition, CI talkers received more judgments for the "neutral" emotion than did TH talkers, even though these produced sentences were not intended to express a "neutral" emotion. This pattern of result suggests that the CI users produced speech with impaired F_0, resulting in a less perceptible (and therefore more monotone or "neutral") judgement. Finally, there was a significant correlation (r=0.524, $p < 0.05$) between the emotional prosody perception ability of CI individuals and TH listeners' rating scores of the same CI individuals' productions. This implies that difficulties with emotional prosody perception contributes to imprecise speech prosody production, through a reduced ability to form correct speech internal models and/or by problems in monitoring auditory feedback relating to prosodic cues in the speech.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Prosody (or the “melody of language”) carries not only linguistic, but also paralinguistic function. “Emotional prosody” refers to the process of expressing emotions through spoken language (Buchanan et al., 2000; Raithel & Hielscher-Fastabend, 2004). Correctly recognizing speakers’ emotional states is essential for spoken communication (Soto & Levenson, 2009). Particularly for those individuals with severe-to-profound hearing loss, deficits in emotional state recognition may have a negative influence on social relationships and social networks (Jiam, Caldwell, Deroche, Chatterjee, & Limb, 2017; Schorr, 2005; Su, Galvin, Zhang, Li, & Fu, 2016). For example, individuals with hearing impairment (HI) who experience inaccurate perception of emotional expressions may receive insufficient and incorrect feedback from their peers, leading to low self-esteem, isolation and rejection affecting their life quality (Schorr, 2005; Schorr, Roth, & Fox, 2009).

Cochlear implants (CIs) function as sensory aids that transmit sound energy via electrical coded stimulators to the auditory nerve. CIs can provide remarkably restored hearing function to listeners with severe-to-profound hearing loss and should therefore help them to better identify emotions from speech signal. Regardless of this remarkable success, research has shown that current CI technology does not perform well in encoding the spectro-temporal fine structure of speech (e.g., fundamental frequency and harmonics) (Chen, Wong, Chen, & Xi, 2014; Jiam et al., 2017; Moore, 2003; Tan, Dowell, & Vogel, 2016; Xu et al., 2004). For example, in CI systems, a bandwidth of 5000-10,000 Hertz (Hz) is typically encoded by 6-22 spectral bands. This limited temporal information (i.e., band-specific envelopes comprised of a few hundred Hz) may result in only 6-10 functional channels and, therefore, may lead to poor pitch perception in CI users (Nie, Barco, & Zeng, 2006). In addition, TH listeners have an input dynamic range of ~120 decibels (dB), whereas CI recipients receive a much narrower range (i.e., ~30-60 dB) of sound. This intensity resolution limitation may also contribute to poor emotional prosody identification for CI users.

Most studies of vocal emotion processing in individuals with CIs have been conducted in non-tonal languages (e.g., English) (Chatterjee et al., 2015; Chatterjee et al., 2016; Gilbers et al., 2015; House, 1994; Luo, Fu, & Galvin III, 2007; Most & Aviner, 2009; Pereira, 2000; Pereira, 2000b). Findings show that English-speaking CI users demonstrate significant deficits in vocal emotional prosody recognition and production, in comparison to their typical hearing (TH) counterparts (Chatterjee et al., 2015; Chatterjee et al., 2016; Gilbers et al., 2015; Luo et al., 2007; Most & Aviner, 2009; Pereira, 2000b; Wang, Trehub, Volkova, & van Lieshout, 2013). These deficits for individuals with CIs may be caused by difficulty detecting subtle acoustic features (e.g., pitch and duration cues) involved in vocal emotion, which in turn can lead to insufficiency in perceiving and producing prosody (Gandour et al., 2003; Jiam et al., 2017).

Little is known about how speakers of tonal languages, such as Mandarin, process emotional prosody. In Mandarin, tone serves as a linguistic cue to distinguish the meaning of words. Beside the use of fundamental frequency (F_0) in signaling lexical stress (e.g., record [verb] vs record [noun]), non-tonal languages like English do not frequently utilize pitch change to differentiate lexical meaning. However, in languages like Mandarin, lexical tone and emotional prosody are signaled together with pitch. Therefore, speakers of tonal languages and non-tonal languages may differ in the way acoustic information (e.g., F_0) is used to process emotional prosody. A recent study by Su et al. (2016) examined 26 Mandarin-speaking CI recipients (adults and children) in sentence recognition task using a single interrogative sentence on a variety of conditions, including whispered speech, emotional speech (“happy”) and non-emotional speech produced by one female talker. Major findings include that adult CI recipients performed worse on identifying emotional speech (71.9%) than non-emotional speech (81.0%). Aside from this study, little research has been conducted on vocal emotion perception and production in Mandarin-speaking individuals with CIs. Also, there are no data comparing emotional prosody perception by Mandarin-speaking CI users and English-speaking CI users.

Because current CI technology results in highly degraded complex pitch perception (Kong, Lee, Yuan, & Yu, 2012), tonal language (e.g., Mandarin) speaking individuals with CIs may have difficulty in processing tone. Therefore, a key question is whether CI recipients who are speakers of tonal languages differ from CI recipients who are speakers of non-tonal languages with respect to the extent of emotional prosody processing. Zhu (2013), citing Ross et al. (1986), proposed a “functional view” (FV) hypothesis claiming that spoken languages have a limited prosodic space in their use of pitch, which may influence emotional prosody processing. According to this hypothesis, speakers of tonal languages such as Mandarin use pitch for lexical purposes and therefore have less space remaining for expressing emotional prosody through the pitch dimension. This hypothesis implies that using F_0 for one linguistic function (i.e., word meaning) in Mandarin inhibits its use for the communication of another function (e.g., emotional prosody) (Figure 1.1). In contrast, Xu (in press) has expressed skepticism concerning claims that tonal languages have limited “room” left for intonation. Instead, Xu emphasizes that multiple aspects of prosody appear to be encoded by different coding mechanisms that presumably rely on

F_0 for various purposes. The current study will provide evidence concerning the FV hypothesis by obtaining data for Mandarin (a tonal language) that may be compared with previous results of English (a non-tonal language).

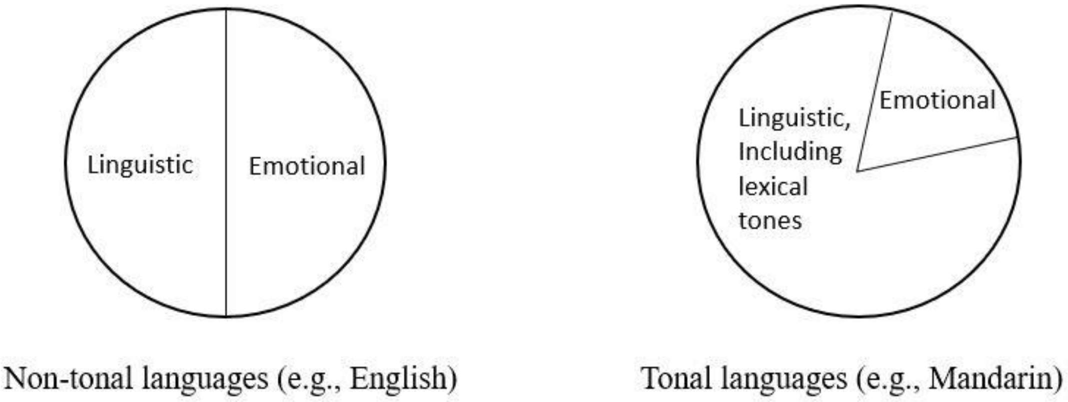

Linguistic
Emotional
Non-tonal languages (e.g., English)
Linguistic, Including lexical tones
Emotional
Tonal languages (e.g., Mandarin)

Figure 1.1. Idealized model of Functional View (FV) for different usage of pitch dimension in prosody "space" for non-tonal and tonal languages (based on Ross, Edmondson, & Seibert, 1986; Zhu, 2013)

The most detailed research to date in this area appears to be that of Chatterjee and colleagues (2015), comparing vocal emotional recognition by English-speaking TH individuals and CI users. Participants with CIs included both children and adults. TH listeners performed a five-alternative forced-choice vocal emotion recognition task ("happy", "sad", "angry", "scared", and "neutral") with four conditions of synthetic speech: Full spectrum and 16-channel, 8-channel and 4-channel noise vocoded, while CI users only heard full spectrum stimuli. A series of studies demonstrated that the sentence recognition of adults with CIs improves with an increasing number of electrodes in both noise and quiet, peaking at eight channels (Fishman, Shannon, & Slattery, 1997; Friesen, Shannon, Baskent, & Wang, 2001; Garnham, O'driscoll, Ramsden, &

Saeed, 2002; Kong, Winn, Poellmann, & Donaldson, 2016; Nie et al., 2006). Based on these studies, Chatterjee and colleagues (2015) hypothesized that English-speaking TH adult listeners' performance with 8-channel noise-vocoded speech would demonstrate the same approximate performance as that of CI listeners. Their results confirmed this hypothesis for English-speaking adults and children with CIs. Using a similar paradigm with natural and noise-vocoded Mandarin speech, Lu et al. (2017) found that Mandarin-speaking children with CIs showed deficits in vocal emotion recognition in natural speech, compared to TH Mandarin-speaking children. They also found that TH children showed a range of performance with 8-channel and 4-channel noise-vocoded speech that was similar to that of children with CIs listening to the natural speech.

The first aim of this study is to explore the perception of emotional prosody by Mandarin-speaking TH and CI listeners. These data contribute to a growing body of studies describing how emotional prosody may be particularly vulnerable in individuals with hearing impairment, particularly those with CIs. In addition, the data will permit testing of the FV hypothesis (Zhu, 2013) and help determine whether the emotional prosody processing differences observed between adult TH and CI individuals described in the English-speaking literature will obtain for Mandarin, a tone language. This can be done by implementing the noise-vocoded speech paradigm of Chatterjee et al. (2015) to test Mandarin-speaking adult listeners using a similar methodology.

A second and related aim of this study is to explore how Mandarin-speaking CI listeners perceive the acoustic characteristics (F_0, amplitude, and duration) of emotional prosody. These data can be used to investigate whether the enhancement of secondary cues (amplitude and duration) for emotional prosody can benefit Mandarin-speaking CI listeners. The data can also be used to test whether a common pattern of emotional recognition obtains among tonal and non-tonal language-speaking individuals with TH and with CIs, which is that “angry”, “sad” and “neutral” are easier to perceive than “happy”, and that “happy” often confused with “angry”. Lastly, this experiment addresses a controversy concerning how individuals with CIs who speak English (a non-tonal language) use acoustic cues to determine the prosodic information in speech. Some studies have shown that non-tonal language-speaking individuals with HI rely more heavily on secondary acoustic cues (e.g., amplitude) than F_0 (Moore & Moore, 2003; Moore & Carlyon, 2005), while other studies suggest that CI listeners attend primarily to F_0 in prosody perception (Meister, Landwehr, Pyschny, Walger, & Wedel, 2009) and do not critically attend to secondary cues.

The third aim of this study is to explore the relationship between the perception and production of acoustic cues for emotional prosody in Mandarin-speaking CI users. This will test the hypothesis that the poor perception of emotional prosody contributes to imprecise speech prosody production through a reduced ability to form correct speech internal models and/or by problems in monitoring auditory feedback. To date, several studies support these types of explanations. For instance, studies have used the “device-on and device-off” paradigms to test on both adults and children with CIs (Svirsky, Lane, Perkell, & Wozniak, 1992; Tobey, et al., 1991; Tye-Murray, Spencer, Bedia, & Woodworth, 1996). In these paradigms, speech production measures of CI users are obtained with the speech processor turned on versus off. The results show that there are significant decreases in speech production including prosodic deficits, when the processor turned off (Tye-Murray et al., 1996). The findings suggest that auditory feedback is crucial for ongoing speech production. Similarly, other studies emphasize that an internal model of speech is essential for accurate speech production. With maturation of a healthy central auditory system, the internal model increases its accuracy in reproducing more precise sequences of speech sounds (Perkell et al., 2000). Therefore, long term use of CIs results in eventual improvement of prosodic quality, suggesting CIs function to strengthen the internal representation of this aspect of speech.

By examining these aims, this thesis will contribute to the literature concerning emotional prosody in tonal-language speakers, including those individuals with hearing impairments. Given the increasing population of CI recipients who speak tonal languages (e.g., Mandarin), understanding the role of vocal emotional prosody perception and production can potentially contribute to better rehabilitation benefits and sound processing strategies for these individuals.

CHAPTER 2

LITERATURE REVIEW

Prosodic characteristics of speech produced by individuals with typical hearing (TH)

Prosody (or the “melody of language”) carries not only linguistic but also paralinguistic function. Much of what we know about prosody has been studied in English (Bänziger & Scherer, 2005; Dellaert, Polzin, & Waibel, 1996; Luo et al., 2007; Mozziconacci & Hermes, 1999; Pell, 2000; Peters, 2006; Raithel & Hielscher-Fastabend, 2004; Rodero, 2011; Williams & Stevens, 1972). This review will therefore begin by describing prosody in English, both linguistic and emotional. The behavioral of TH individuals will be reviewed, serving as the basis for a comparison with individuals who have hearing loss (HL).

Linguistic prosody in English

Linguistic prosody is suprasegmental information which serves as a grammatical function, such as word stress (e.g., object vs object) (Fry, 1958; Gay, 1978), sentence focus (e.g., the BIRDS are singing vs the birds are SINGING) (Lambrecht & Polinsky, 1997), marking boundaries (Lehiste, Ilse & Wang, 1976; Lehiste, Ilse, 1979) and intonation (Pierrehumbert, 1980). English sentence-level intonation patterns are typically classified into three or four major types: declarative, two question types (those beginning with interrogative words and those requiring a Yes/No answer), commands, and exclamations (Chun, 2002). In English, simple declarative sentences and exclamations are traditionally described as having falling patterns, and Y/N questions as having rising patterns.

In addition, linguistic prosody plays an essential role in providing a platform for early language acquisition by children. For example, stress patterns in English sentences contain both strong and weak syllables, providing a credible and helpful tool to separate words in speech (Whalley & Hansen, 2006). In addition, Epstein (2002) examined the effects of stressed or prominent words for English sentence intonation compared to non-prominent word, based on speaker's voice quality. The results showed that speakers do use a tenser voice quality to present prominent words and this may be useful for listeners in distinguishing phonological tone differences in statement and question sentences. Furthermore, less intonation variability may cause speakers sound "foreign" or "disordered."

In general, prosody represents the melody and rhythm in speech, and linguistic prosody expresses linguistic content in speech. Here, I shall first briefly review prosody and linguistic prosody to provide background for understanding emotional prosody.

Emotional prosody is a second important function of prosody and will be the focus of this dissertation. Signaling the emotional state of the speaker is, in fact, a paralinguistic (rather than linguistic) skill. Emotional prosody refers to the process of expressing emotions through spoken language (Buchanan et al., 2000; Raithel & Hielscher-Fastabend, 2004). To correctly perceive how others feel, individuals must be able to utilize and integrate emotional cues from various sources, such as face, body language and tone of voice (Uskul, Paulmann, & Weick, 2016). Although talkers' emotional states can be clearly indicated by facial expressions, recognition of another's emotional state using only auditory cues is crucial for many types of communication (e.g., phone calls, listening to radio) (Luo et al., 2007).

Various terms are used to describe emotion, including “basic emotions” (Izard, 1992; Ortony & Turner, 1990; Stein & Oatley, 1992), “fundamental emotions” (Gray, 1994) and “primary emotions” (Plutchik, 1984). Researchers commonly investigate sets of emotions, including “happy”, “angry”, “sad”, and “neutral” (Cowie & Cornelius, 2003; Greasley, Sherrard, & Waterman, 2000; Juslin & Laukka, 2003; Juslin & Scherer, 2005). To study how emotional states are expressed in speech, studies typically examine how well listeners perceive emotional cues from semantically meaningful sentences (Pell, Monetta, Paulmann, & Kotz, 2009) or from pseudo-sentences (i.e., sentences without meaning) (Scherer, Banse, & Wallbott, 2001). Interestingly, investigations have shown that listeners with TH generally perceive some emotions (e.g., “angry”) better than others (e.g., “happy” or “sad”) (Banse & Scherer, 1996; Juslin & Laukka, 2003). For example, Banse and Scherer (1996) conducted a recognition study that examined 12 participants responding to 14 emotions. The results showed that “angry” is better recognized than “happy” and “sad”. Juslin and Laukka (2003) featured a total of 60 listening experiments to decode accuracy for vocal emotion expression, showing “angry” and “sad” have higher accuracy than “happy”. Based on this rather well-described literature, “happy”, “angry”, “sad”, and “neutral” are selected as the basic emotions discussed in this review.

A growing body of research has reported acoustic analyses of emotional prosody over the past few years (Bachorowski & Owren, 1995; Banse & Scherer, 1996; Johnstone & Scherer, 2000; Scherer, 2003; Schuller, Rigoll, & Lang, 2004; Sidorov, Brester, Ultes, & Schmitt, 2017; Sobin & Alpert, 1999; Williams & Stevens, 1972). These studies typically examine fundamental frequency (F_0), amplitude, and duration (Lehiste, I., 1970). F_0 is defined as the frequency of vocal fold vibrations, expressed in Hertz (Hz). In speech, varying the F_0 produces changes in the perceived pitch. Amplitude is a description of the magnitude of a sound wave and is typically perceived as loudness, measured in decibels (dB). Duration describes the extent of a sound perceived as length, typically measured in milliseconds (ms) (Peters, 2006).

In general, the pitch ranges of adult males and females are distinct. For example, most adult males will have average F_0 in the range of 85 to 185 Hz, while a typical adult female average is from 165 to 255 Hz (Titze, 2000). Williams and Stevens (1972) measured acoustic cues in the speech signal reflecting vocal emotion and found that “angry” has a higher F_0 than “neutral”. Sobin and Alpert (1999) examined a total of 152 emotional sentences, showing that “angry” also has a high F_0 with a short duration, and “sad” has a low F_0 with a long duration. In line with these findings, Banse and Scherer (1996) found that “angry” shows a high increase in intensity, while “sad” exhibits decreased intensity.

The acoustics of emotional states of speech can also be influenced by arousal and valence dimensions (Gilbers et al., 2015; Posner, Russell, & Peterson, 2005; Russell & Mehrabian, 1977; Russell, 1980; Schröder, Cowie, Douglas-Cowie, Westerdijk, & Gielen, 2001). Valence concerns the difference between positive (e.g., “happy”) and negative emotions (e.g., “sad”), while arousal concerns the difference between high-stimulation (e.g., “angry”) and low-activation states (e.g., “bored”) (Gilbers et al., 2015; Russell, 1980). These factors are shown in Fig. 2.1. Literature related to valence dimension and voice pitch is inconsistent. Some studies found positive valence was associated with low mean values of F_0 and large variability of F_0 (Pereira, 2000a; Scherer & Oshinsky, 1977; Uldall, 1960), while other research fails to observe relationship of vocal cues and valence (Apple, Streeter, & Krauss, 1979; Schröder et al., 2001). As for duration and intensity cues, Schröder et al. (2001) investigated a database of 100 English speakers’ spontaneous emotional speech and observed that “angry” speech shows short pause duration, is high intensity, and is associated with high activation level, while negative valence such as “sad”, presents longer pause duration, is low intensity, and is related to low activation level. As Jiam et al. (2017) note, the relationship between vocal intensity and voice emotion is important, the data are few, and this issue needs further exploration.

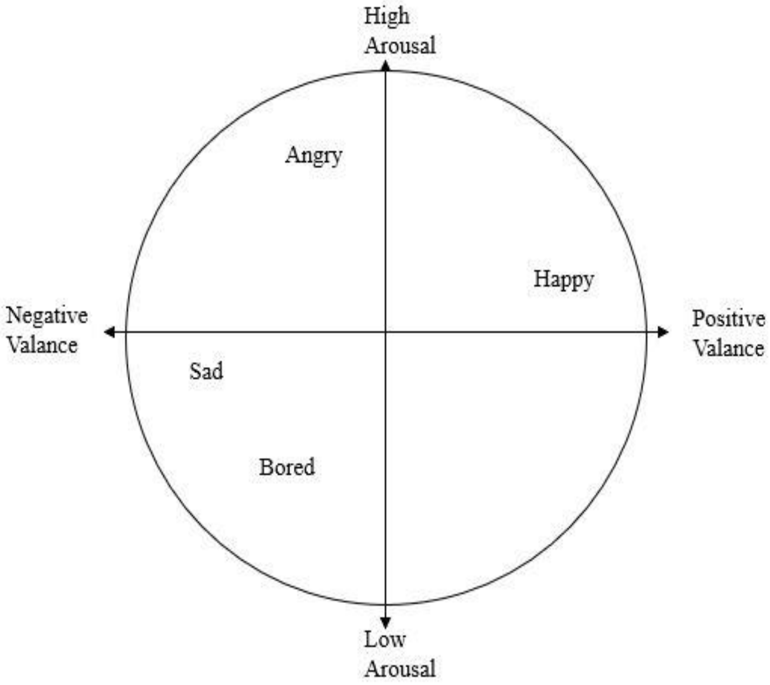


Figure 2.1. Examples of arousal and valence dimensions in four emotions (based on Russell, 1980).

Linguistic prosody in Mandarin

Mandarin Chinese, referred to as “Mandarin” is spoken by over 1 billion speakers. Mandarin tone is considered a suprasegmental phenomenon, having a wider pitch range than English, and serving as an abstract linguistic property to distinguish the meaning of words (Zhu, 2013). Mandarin has four distinct tone patterns: (1) flat and high, (2) rising, (3) falling-rising, and (4) falling, respectively (Xu et al., 2004). For instance, the syllabus /ma/ expressed with tone 1-4, can mean “mother”, “hemp”, “horse”, and “curse” respectively (Fu & Zeng, 2000; Jongman, Wang, Moore, & Sereno, 2006; Wei, Cao, & Zeng, 2004) (Figure 2.2). Tone contours carry important information for Mandarin speech perception (Chen, Wong, & Hu, 2014). It is easiest for TH Mandarin speakers to identity Tone 1 and Tone 4, while it is hardest for them to identify Tones 2 and 3, and they often confuse Tone 2 with Tone 3 (Liu, Tien-Chen, Hsu, & Horng, 2000).

The primary and sufficient acoustic parameter to characterize Mandarin tone is fundamental frequency (F_0) of the speech signal, especially, the F_0 height and F_0 contour (Fu & Zeng, 2000; Howie, 1976; Jongman et al., 2006; Liu, 1924; Luo & Fu, 2004; Massaro, Cohen, & Tseng, 1985; Moore & Jongman, 1997; Wong, 2012; Zhu, 2013). The F_0 characteristics of a word at the syllabic level can have an effect on the phonological level (Duanmu, 2007). Also, some tones (e.g., Tone 2 and 3) show similar F_0 contours (Jongman et al., 2006; Moore & Jongman, 1997). In addition, other acoustic cues that covary with F_0 are amplitude, vowel duration, and vocal quality. For instance, the four tones in Mandarin have different duration values, with Tones 3 and 4 showing the longest and the shortest lengths, respectively (Xu & Zhou, 2011). Duration and amplitude can provide additional information concerning tone perception when F_0 cues are absent (Fu & Zeng, 2000; Liang, 1963; Ryant, Yuan, & Liberman, 2014; Whalen & Xu, 1992; Xu & Zhou, 2011). Fu and Zeng (2000) generated several signal-correlated-noise stimuli given to four young healthy adult listeners, in order to identify what types of temporal envelope cues (e.g., duration, amplitude contour, and periodicity) contribute to tone identification. The perceptual results showed that the duration cue mainly used to recognize Tone 3, and the amplitude cue distinguishes Tone 3 from Tone 4 (Fu & Zeng, 2000; Fu & Zeng, 2013).

Some research provides evidence that duration and amplitude cues may play a stronger role in Mandarin tone perception than was previously expected. For example, Gao (2002) states that with the absence of F_0 cues in natural whispered speech, duration and amplitude contour can moderately benefit tone identification. Liu and Samuel (2004) examined 10 Chinese listeners perceiving 80 Mandarin monosyllables in whispered speech. The results showed that amplitude contour cues are essential for Tone 3 recognition. In support of the flexible nature of cues for Mandarin tone perception, in recent research by Ryant et al. (2014) a deep neural network (DNN) was trained to classify Mandarin tone in the absence of pitch information using cepstral mean-variance normalization and produced highly accurate Mandarin tone classification. Although it should be noted that deep learning is a computational model and the relevance to human speech perception remains unknow, these findings may still provide indirect evidence to support the role of secondary cues involved in lexical tone perception. In general, some (but not all) studies suggest a complementary role of duration and amplitude in cuing Mandarin tone when F_0 information is minimal.

Beside the use of F_0 to signal stress, non-tonal languages (e.g., English) do not use pitch change to differentiate lexical meaning. In contrast, in tonal languages like Mandarin, tone and emotional prosody are signaled together with pitch. Therefore, there may be differences in the ways that the speakers of tonal languages and non-tonal languages use F_0 to process emotional prosody.

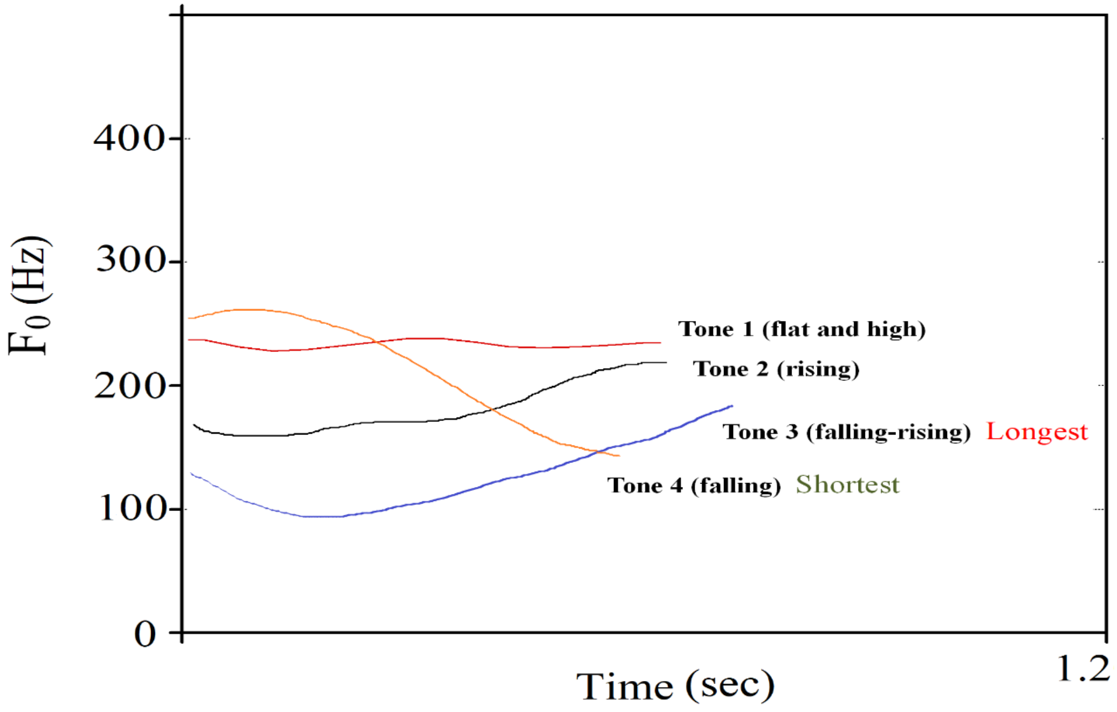


Figure 2.2. The syllable *ma* produced in four Mandarin tones by a Mandarin-speaking female talker (author's voice).

Researchers have studied emotional prosody in Mandarin using various methods, for example by examining 1) emotional sentences from TV shows and movies (Tao, Kang, & Li, 2006; Yu, Chang, Xu, & Shum, 2001; Zhang, Shiqing, 2008); 2) language-like pseudo-utterances emotionally expressed by speakers (Liu, Pan & Pell, 2012) and (3) sentences elicited from emotionally-biasing contexts (You, Chen, & Bu, 2005). Yuan and colleagues (2002) conducted acoustic analyses of 288 sentences from a Chinese emotion database. The researchers concluded that "angry" is mainly cued by the amplitude difference of the first and second harmonics, a phonation factor, while "joy" is mostly signaled by fluctuation of sentence F_0. Li and colleagues (2011) examined 504 monosyllabic utterances for emotional production. The findings showed that Mandarin-speaking individuals express some emotions like "angry" by using some degree of falling intonation with a reduced pitch range, and "happy" by a type of rising intonation with a higher pitch range (Li et al., 2011). For example, Figure 2.3 shows the words "happy" and "angry" marked with arrows in the final word of Mandarin sentence (da3 gao1 er3 fu1) "Play golf". From Liu and Pell's research (2012), acoustic analysis at the word level shows the following pattern: "angry" exhibits relatively high F_0 values and large amplitude variation, while "sad" presents low F_0 values and small amplitude variation (Liu, Pan & Pell, 2012). At the sentence level, F_0 values and amplitude variation show the same kind of patterns in "angry" and "sad" productions. In addition, "angry" shows the shortest sentence length, while "sad" exhibits the longest duration (Zhang, Sheng, Ching, & Kong, 2006). The relations of the three acoustic cues for words and sentences are summarized in Table 2.1.

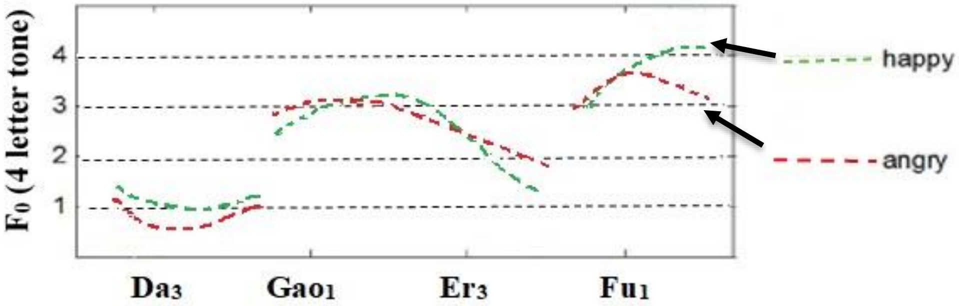


Figure 2.3. Mandarin sentence (da3 gao1 er3 fu1) "Play golf" spoken in two emotions by a male speaker (based on Li et al., 2011). F_0 contours have been time-normalized by linear time compression. "Happy" is marked with an arrow in the final word, showing a higher pitch range, whereas "angry" has a reduced pitch range in the final word.

Table 2.1. *Acoustical characteristics of prosody for four emotions in Mandarin (based on Zhang, Sheng et al., 2006).*

		Acoustic cues		
Emotion		F_0	amplitude variation	duration
word-level	angry	high	large	short
	sad	low	small	long
	happy	relatively high	relatively large	relatively short
	neutral	(happy>neutral>sad)		
sentence-level	angry	high	large	short
	sad	low	small	long
	happy	relatively high	relatively large	relatively short
	neutral	(happy>neutral>sad)		

Emotional prosody perception and production by individuals with CIs

The effect of auditory deprivation on the processing of emotional prosody

Auditory deprivation, such as that which occurs in hearing-impaired (HI) individuals with CIs, could cause difficulty processing emotional prosody. These issues were addressed by Schorr (2005), who investigated the accuracy of vocal expressions of emotion by English-speaking children with CIs (age range 5-14 years). In this study, 39 children with CIs and 37 TH age-matched peers were required to identify the emotional valence of sounds ("positive", "negative" or "neutral"). The results showed that children with CIs demonstrated less accuracy than TH peers in identification of emotional sounds. In addition, Schorr (2005) stated that the effect of auditory deprivation during early childhood could limit positive relationships with parents and peers and cause problems interpreting the emotional expressions of others accurately. In support of negative effects of auditory deprivation on emotional prosody perception, Luo et al. (2007) investigated the ability of vocal emotion recognition (e.g., “happy”, “sad”, “neural” and “anxious”) by testing 16 English-speaking listeners (age range 22-73 years), including eight TH individuals and eight CI users. The results suggested that listeners with CIs exhibit poorer performance (~ 45% accuracy) in vocal emotional perception than TH individuals (~ 90% accuracy). Together, these data imply that lack of auditory input could have a negative effect on processing emotional prosody in English-speaking HI individuals with CIs.

Non-tonal languages: Evidence from English and Japanese

A growing body of research focusing on vocal emotion recognition and production in TH and CI individuals in non-tonal languages, such as English and Japanese (Chatterjee et al., 2015; Chatterjee et al., 2016; Gilbers et al., 2015; Luo et al., 2007; Nakata, Trehub, & Kanda, 2012; Pereira, 2000b; Wang et al., 2013; Zhu, Miyauchi, Araki, & Unoki, 2016). Most of these studies use emotionally recorded productions of semantically neutral sentences (e.g., “it takes 2 days”; Luo et al., 2007), in order to test the perceptual accuracy of targeted emotions (e.g., “sad”, “happy”, “neutral”, “angry” and “scared”). Such studies have shown that English-speaking CI users demonstrate significant deficits with vocal emotional perception and exhibit a disadvantage at recognizing emotions, when compared with TH individuals (Chatterjee et al., 2015; Gilbers et al., 2015; Luo et al., 2007; Peters, 2006). Luo et al. (2007) investigated the ability of vocal emotion recognition (e.g., “happy”, “sad”, “neutral” and “anxious”) by testing 16 English-speaking listeners, including eight TH and eight CI individuals. In general, they found that TH listeners exhibit better performance (~ 90% accuracy) in vocal emotional perception, while listeners with CIs only showed 45% accuracy. In addition, they stated that CI users identified “sad”, “angry” and “neutral” emotions more easily, rather than “happy”.

In line with these findings, Chatterjee et al. (2015) tested 9 adults with CIs and 10 adults with TH on sentences with different emotions (e.g., “he wore his yellow shirt.”). Results showed that adults with CIs have a deficit with vocal emotion recognition relative to adults with TH. In addition, adult listeners with TH performed better in emotional prosody recognition than adult listeners with CIs, even with synthetic stimuli designed to resemble CI input (Chatterjee et al., 2015; Gilbers et al., 2015).

In general, non-tonal language speaking CI recipients show higher identification in “sad” and “neutral”, rather than “happy”, and often confuse “angry” with “happy” (House, 1994; Luo et al., 2007; Pereira, 2000b). These patterns could be for at least three reasons. Firstly, CI users may have failed to identify emotional acoustic cues in the speech samples and therefore tend to mostly respond with “neutral”, by default. Secondly, the acoustic cues of “sad” are quite distinct from “happy” and “angry” (i.e., lowest mean F_0 value, lowest amplitude, and longest duration), which should make “sad” more easily recognized. Thirdly, the small differences between acoustic features of “happy” and “angry” likely make CI recipients confuse these two emotions.

Although obvious impairment in emotional prosody production by CI recipients is reported to persist (Jiam et al., 2017), few studies have investigated emotional prosody production in CI recipients. Recent emotional speech production tasks conducted by Chatterjee et al. (2015) demonstrated that child CI recipients generate much smaller intensity, pitch range, and mean pitch difference than children with TH in “happy” and “sad” emotions, as well as smaller spectral change. Nakata et al. (2012) found that Japanese-speaking children with CIs performed poorer and received lower ratings on the production of emotional prosody, especially in expression of “disappointment”, compared with their TH peers. In addition, they reported a strong correlation between sentence prosody perception and production in the speech of Japanese children with CIs.

Taken together, these data suggest that non-tonal language-speaking individuals with CIs demonstrate deficits in vocal emotion recognition and production. The data also suggest a correlation between emotional prosody perception and production in these individuals at the sentence level.

Tonal languages: Evidence from Mandarin

As previously mentioned, acoustic cues for emotional prosody and linguistic tone share the same phonetic features. For example, tone is cued by F_0, and paralinguistic information (such as a speaker’s emotion state) is also typically conveyed through F_0. By experiencing auditory impairment and having difficulty encoding pitch information, Mandarin-speaking individuals with CIs may demonstrate a deficit in vocal emotion perception and production. Recently, Su et al. (2016) studied 26 CI recipients (15 adults and 11 children) with emotional speech from Mandarin Hearing in Noise Test (MHINT; Wong et al., 2007) database. Ten sentences, produced by a female talker speaking in “happy” and “excited” manners were selected from this database. The researchers found that the Mandarin-speaking CI recipients performed more poorly with emotional speech than with normal speech, and that these subjects exhibited a wide range of variability. Except for this study, very little research has been conducted on vocal emotion perception and production in Mandarin-speaking adults with CIs. Furthermore, Su et al. (2016) did not examine other common emotions in daily life, such as “angry” and “sad”. It is unclear from these data how variability in F_0 and amplitude affects the recognition of emotional speech by Mandarin-speaking CI users (Su et al., 2016).

Although this dissertation focuses on adults with CIs, Lu and colleagues (2017) recently examined Mandarin-speaking children with CIs. This voice emotion recognition study used the same MHINT database and a similar paradigm to the natural and noise-vocoded speech experiment conducted by Chatterjee et al. (2015). Participants included 41 Mandarin-speaking children with CIs and 47 Mandarin-speaking children with TH. During the testing procedure, all participants heard emotionally neutral Mandarin sentences in five emotions (“happy”, “scared”, “neutral”, “sad” and “angry”) spoken by a male and a female in a child-directed way. TH participants heard the natural and noise-vocoded Mandarin speech, whereas CI participants only heard the natural speech. The results showed that Mandarin-speaking children with CIs experienced deficits in vocal emotion recognition in natural speech, as compared to Mandarin-speaking children with TH. In addition, TH children showed a range of performance with 8-channel and 4-channel noise-vocoded speech that was similar to the performance of children with CIs listening to the natural speech.

Overall, compared to the growing numbers of studies that focus on processing of emotional prosody by non-tonal language speakers with CIs, much less data on Mandarin-speaking CI recipients have been reported. To date, no data exploring the difference between the performance of vocal emotion perception and production by adult Mandarin-speaking CI users and English-speaking CI users. A chart summarizing the two most relevant studies of children and adults with TH and with CIs in English and Mandarin is shown in Appendix A.

A “Functional View” model for prosody and tone (Zhu, 2013)

Ross et al. (1986) investigated emotional prosody in tone languages and compared them with English. Five male native speakers of Mandarin and five male native speakers of English were asked to produce sentences in their native languages with five emotions: “neutral”, “happy”, “sad”, “angry”, and “surprise”. This study aimed to evaluate how well speakers of a tonal language used acoustical parameters for signaling overall emotional prosody across the spectrum, in comparison to English speakers. The results suggested that tone in a language limits the free use of F_0 for signaling emotions. Zhu (2013) further interpreted the findings reported by Ross et al. (1986), using additional experimental approaches to examine the ability of perceiving vocal emotions in native and non-native listeners of a tonal language. In Zhu’s first experiment, 20 native Mandarin-speaking listeners and 20 native Dutch-speaking listeners were asked to identify six emotional prosodies in Mandarin (“neutral”, “happy”, “angry”, “surprise”, “sad”, and “sarcasm”) produced by four native Mandarin-speaking talkers. The results showed no significant difference between native Mandarin-speaking listeners (mean recognition rate = 45.9%) and Dutch-speaking listeners (mean recognition rate = 45.6%) in identifying various emotions expressed in Mandarin. In addition, native Mandarin-speaking listeners showed less confidence (mean = 1.49) in their emotional identifications than the Dutch-speaking listeners (mean = 1.96). The results of this first perception study suggested that native Mandarin-speaking listeners failed to recognize emotions in their native language more correctly and confidently than native Dutch-speaking listeners. In a following perceptual test, another group of 20 native

Mandarin-speaking listeners and 20 native Dutch-speaking listeners participated, using the same emotional prosodic stimuli, but portrayed in Dutch. The results showed that the mean correct identification rate for the native Mandarin-speaking listeners was 37%, which was significantly lower than the rate for the native Dutch-speaking listeners (66%). That is, native Dutch-speaking listeners "identified emotional prosody portrayed in their native language (Dutch) significantly better than they did with the unknown language – Mandarin" (Zhu, 2013, p. 109). These findings supported Zhu's prediction that "listeners of a non-tonal language are generally better at perceiving emotional prosody than listeners of a tonal language" (Zhu, 2013, p. 109). According to Zhu's "functional view model" (2013), spoken languages have a limited prosodic space (affecting pitch, loudness, and length) and this can affect performance in processing emotional prosody. Speakers of Mandarin, a tonal language, use pitch for lexical purposes and will therefore have relatively little space for expressing emotional prosody through pitch. In general, Zhu's study suggests that "native listeners of a tonal language are less intent on (and in fact less experienced in) decoding this paralinguistic use of prosody than listeners of a non-tonal language" (Zhu, 2013, p. 6). Zhu's hypothesis implies that use of a certain type of acoustic feature (e.g., F_0) in Mandarin inhibits or "limits its use for the communication of emotion" (Zhu, 2013, p. 25). In contrast, there is reason to suspect that Mandarin speaking individuals do not have limited "space" in F_0 use for prosody (Xu, in press). For instance, Xu suggest that the multidimensionality of emotional prosody cannot be only explained by the pitch dimension alone, but instead, prosody appears to be determined by different coding mechanisms (e.g., formant dispersion, voice quality, and body size projection). According to this view, the speech of Mandarin-speaking individuals with CIs will resemble that of English-speaking individuals with CIs with respect to the use of prosody.

The present study can further test the FV hypothesis through a comparison between Mandarin and another non-tonal language, such as English. Specifically, we examine data from Mandarin and English-speaking CI recipients.

Summary and Introduction to Research Questions

In this section, studies addressing the prosody of individuals with TH in English (a non-tonal language) and Mandarin (a tonal language) were reviewed. In addition, the emotional prosody capability of the HI population with CIs was addressed. Individuals with TH can easily distinguish basic emotions based on acoustic characteristics, in contrast to HI individuals with CIs, who do not perform as well (Pereira, 2000). Because of deprivation of auditory stimulation and deficits in tone perception and production by Mandarin-speaking CIs users, regaining the ability to perceive and produce tone is likely to be difficult. Moreover, in current CI speech-processing strategies, the coding of spectro-temporal cues for incoming speech signals is limited (Tan et al., 2016). Particularly, the current CI systems do not strongly and explicitly present pitch information of complex acoustic stimuli (Moore, 2003; Tan et al., 2016; Xu et al., 2004). Therefore, Mandarin-speaking individuals with CIs likely have ongoing difficulty encoding pitch information for perception of lexical tones and for recognizing emotional prosody in sentences.

According to the literature on non-tonal languages, CI users have obvious differences with voice emotional production, in comparison to their TH counterparts (Chatterjee et al., 2015; Chatterjee et al., 2016; Nakata et al., 2012). In addition, Zhou et al. (2013) found a correlation between overall tone perception and production across Mandarin-speaking CI users, noting that tone production depends on accurate perception. Since tone and emotional prosody are indicated along with pitch, therefore, deficits in emotional prosody perception may cause deficits in emotional prosody production by Mandarin-speaking CI recipients. These issues will be explored further in the dissertation.

In summary, this literature review underscores the necessity of obtaining more data on emotional prosody perception and production in Mandarin, as the current literature contains very few studies on this topic. Fully understanding the communicative intent and emotional state of speakers is critical in order to contribute to positive spoken communication (Chatterjee, Kulkarni, Christensen, Deroche, & Limb, 2015). Schorr et al. (2009) assessed the speech perception (single words) and emotion identification of 37 children with CIs, using a list from a self-reported quality of life questionnaire. The results showed that children with CIs perceived an increase in their quality of life as predicted by their performance on emotion identification tasks, rather than their word recognition scores. Thus, the findings of the present research could potentially be used to improve the spoken communication skills of Mandarin-speaking CI recipients. Few studies have investigated the rehabilitation of emotion perception of CI users, and most of those studies have focused on using facial emotion cues, rather than on vocal emotion cues (Dyck & Denver, 2003; Jiam et al., 2017). Although facial expressions may play an important role in navigating difficult listening conditions, the vocal emotional state of speakers provides critical information in social communication when facial expressions cannot be detected (Chatterjee et al., 2015; Luo et al., 2007). Therefore, the acoustic cues analyzed in this research study could potentially benefit rehabilitative efforts for Mandarin-speaking CI recipients, and this could further support the development of hearing devices/processors.

Research questions and predictions

I. Perception of emotional prosody in Mandarin-speaking TH and CI listeners

Question 1. Do Mandarin-speaking TH and CI listeners differ in the accuracy with which they detect emotional prosody in short Mandarin sentences?

Prediction 1: Yes. Mandarin-speaking adults with CIs are expected to show significant deficits in detecting emotion in natural speech, as compared to TH Mandarin-speaking adults. Also, in accordance with the FV hypothesis (Zhu, 2013), Mandarin-speaking TH adult listener's performance with reduced spectral resolution, synthesized speech will demonstrate the approximate performance of Mandarin-speaking CI listeners. That is, compared to the amount of channel degradation noted for synthesized speech in English (Chatterjee et al., 2015; Fishman et al., 1997; Friesen et al., 2001), i.e., 16-channel → 8-channel, Mandarin-speaking TH users will show a higher degree of degradation (16-channel → 4-channel) in order to approximate CI listeners.

Question 2. Will the enhancement of secondary cues for emotional prosody (i.e., amplitude and duration) benefit Mandarin-speaking CI listeners?

Prediction 2: Yes, the enhancement of secondary cues (amplitude and duration) will benefit Mandarin-speaking CI listeners in perceiving emotional prosody.

II. Production of sentence-level emotional prosody by Mandarin-speaking CI listeners: Acoustic and perceptual measures

Question 3. Does the proportion of acoustic cues (F_0, amplitude and duration) significantly related to emotional prosody production in the speech of Mandarin-speaking CI users differ from that of Mandarin-speaking TH individuals?

Prediction 3: Yes. Because problems monitoring auditory feedback (Selleck & Sataloff, 2014) and/or construction of internal model, Mandarin-speaking CI recipients are expected to demonstrate decreased mean values of F_0, intensity and increased sentence duration values in their sentence production, compared to Mandarin-speaking TH individuals.

Question 4. To what extent is the emotional prosodic information produced by CI speakers perceptually salient to Mandarin-speaking TH listeners?

Prediction 4: TH Mandarin-speaking listeners will show decreased accuracy in detecting emotions (“angry”, “happy”, and “sad”) in sentences produced by Mandarin-speaking CI talkers, as compared with TH Mandarin-speaking talkers. That is, the acoustic patterns noted in Q3 will be perceptually validated with Mandarin-speaking TH listeners.

Question 5. Is there a significant relationship between the perception of emotional prosodic information (I1) and the production of emotional prosody (II2) by Mandarin-speaking CI users?

Prediction 5: Yes. Deficits in emotional prosody perception will correlate with deficits in emotional prosody production by Mandarin-speaking CI recipients. This may suggest that CI recipients having imprecise speech production due to a reduced ability to form correct speech internal models and/or problems with monitoring auditory feedback.

CHAPTER 3

EXPERIMENT 1: PERCEPTION OF EMOTIONAL PROSODY IN NATURAL AND NOISE-VOCODED SPEECH: TH VERSUS CI LISTENER GROUPS

This chapter presents an experiment that seeks to determine whether the perception of different channels of vocoded speech approximate the level at which emotional prosody is represented by Mandarin-speaking CI listeners. The methodology involves synthesized speech designed to resemble a degraded signal similar to that heard by individuals with CIs. This research involves the channel vocoder, a process which using a bank of bandpass filters, breaking down the incoming speech signal into several contiguous frequency channels (Loizou, 2006). Noise-vocoded speech uses random noise and excitation signals, which are frequently utilized as an acoustic simulation of CIs (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Zhu, et al., 2016), and many assume that the performances of TH listeners with different channel settings indicate how well CI users can process emotional prosody. Despite the 16 or more electrodes present in current CI technology, CI users receive the quantity of input information is limited (Nie et al., 2006). These limitations result in there being only 6 to 10 functional channels for a typical cochlear implant user (Chatterjee et al., 2015; Fishman et al., 1997; Garnham, O'driscoll, Ramsden, & Saeed, 2002; Kong, Winn, Poellmann, & Donaldson, 2016; Nie et al., 2006). Studies have compared the performance of adult TH listeners were attending to 4-,8-, and 16-spectral channels synthesized speech, with CI adult listeners' performance in sentence recognition tasks in both quiet and noise. The results suggest that the best CI adults' performance is similar to TH adults' performance with 8-

channel synthesized speech (Chatterjee et al., 2015; Fishman et al., 1997; Garnham et al., 2002; Kong et al., 2016; Nie et al., 2006).

As noted previously, the FV theory claims that speakers of different languages have a limited capacity for processing prosodic information, particularly in F_0. According to this theory, we predict that Mandarin-speaking adults with CIs will perform with relatively poorer emotional prosody perception than English-speaking adults with CIs. Following this line of reasoning, Mandarin-speaking TH adult listeners are predicted to perform at a level similar to Mandarin-speaking CI listeners when those TH listeners hear noise-vocoded speech at a lower band setting of 4-channel, instead of 8-channel (which was reported as the best performance level for English-speaking TH listeners; see above).

Participants

Fifteen Mandarin-speaking adults with typical hearing (TH) (Group 1) and fifteen Mandarin-speaking adults with cochlear implants (CIs) (Group 2) participated. Of the 15 TH participants, 5 were male and 10 were female. Of the 15 CI participants, 6 were male and 9 were female. A Texas cohort (group 1) of TH participants was recruited at the main campus of University of Texas at Dallas, in order to ensure they have a middle-school educational level. TH participants were native speakers of Mandarin Chinese with no self-reported history of speech, hearing, emotional or language impairment. The age range of participants with TH was from 19 to 39 years (mean = 27.3 years, SD = 5.6). A Chinese cohort (group 2)[1] of CI participants was recruited and tested at the Beijing Institute of Otolaryngology in China. The age range of the CI group was from 20 to 30 years (mean = 24.1 years, SD = 3.6), providing a close age match to their counterparts with TH. Participants with CIs all had a middle-school educational level and no self-reported history of emotional or speech/language disorders, other than those associated with their hearing loss. Of the 15 CI users, 2 had bilateral CIs and 13 participants had unilateral CI. For bilateral CI users, the first implanted ear or the one with better hearing level was chosen for the listening test. These participants with bilateral CIs were asked to unplug the non-tested device or wear a disposable earplug in the non-tested ear during the testing period. All participants with CIs had more than two years of experience with the device. Of the 15 CI users, 4 were postlingually deafened and 11 were prelingually deafened. Characteristics of the participants with TH and CIs are presented in Table 3.1. Detailed information about the adult CI group is presented in Table 3.2.

Table 3.1. *Participant descriptive statistics for Experiment 1.*

Characteristic	**Adults with CIs (n= 15)**	**Adults with TH (n=15)**
Age at testing, mean (SD)	24.1 (3.6)	27.3 (5.6)
Duration of CI use, mean (SD)	9.6 (6.9)	NA
Duration of deafness, mean (SD), mo	6.2 (5.1)	NA

[1] Note: a possible concern about groups 1 and 2 is that the Texas cohort of Mandarin-speaking participants with TH might be "bilingual" while the Chinese cohort participants with CIs is not. However, all Chinese individuals with TH and with CIs had reported at least of 12 years of English training experience, suggesting that prior English experience is not a confound.

Age of implantation, mean (SD), mo	14.47 (8.3)	NA
Male, No. (%)	6 (40)	5 (33)
Female, No. (%)	9 (60)	10 (67)
Education, HS graduate, No. (%)	15 (100)	15 (100)

Note. HS= high school

Table 3.2. *Detailed information about adult CI participants for Experiment 1.*

CI ID	**Device**	**Coding**	**Pre/Postlingual deafness**	**Etiology**
CI_1	Cochlear Nucleus 5	ACE	Prelingual	Congenital
CI_2	Cochlear Nucleus 5	ACE	Prelingual	Congenital
CI_3	MED-EL OPUS 2	FS4	Prelingual	Congenital *
CI_4	Nurotron	APS	Postlingual	Drug-induced: HL onset at 15 years
CI_5	Cochlear Nucleus 24	ACE	Prelingual	Congenital *
CI_6	Advanced Bionics	HiRes F120	Postlingual	Drug-induced: HL onset at 23 years
CI_7	Cochlear Nucleus 6	ACE	Prelingual	LVAS: HAs until implantation
CI_8	Cochlear Nucleus 24	ACE	Prelingual	Pneumonia Complication
CI_9	Cochlear Nucleus 24	ACE	Postlingual	Fever: HL onset at 10 years
CI_10	Cochlear Nucleus 5	ACE	Prelingual	LVAS: HAs until implantation
CI_11	Cochlear Nucleus 5	ACE	Prelingual	LVAS: HAs until implantation

CI_12	Cochlear Nucleus 24	ACE	Postlingual	LVAS: HL onset at 6 years; HAs until implantation
CI_13	MED-EL OPUS 2	FS4	Prelingual	Meningitis: HAs until implantation
CI_14	Cochlear Nucleus 24	ACE	Prelingual	Unknown cause
CI_15	Cochlear Nucleus 24	ACE	Prelingual	LVAS

Note. ACE=advanced combination encoder; APS=advanced peak selection; FS=fine Structure; LVAS=large vestibular aqueduct syndrome; HL=hearing loss; HAs=hearing aids; *: bilateral otherwise unilateral

Procedure

Stimuli: The materials were three Mandarin sentences selected from the Affective Speech Recognition Database (Chinese Corpus Consortium, 2006) (see Appendix B), and spoken by one male and one female (professional actors) with four common emotional categories (“angry”, “happy”, “sad”, and “neutral”), giving a total of 24 natural stimuli (3 sentences × 4 emotions × 2 talkers). Three sentences were selected to contain little semantic emotional bias, according to the investigator’s judgement. For example, the sentence “Go do your own stuff” expresses very little emotional bias. Vocoded stimuli were created using a noise-vocoder (Angelsim^{TM}, Emily Shannon Fu Foundation, www. Tigerspeech.com) with four conditions for testing: full-spectrum speech, 16-chanel, 8-channel and 4-channel noise-vocoded speech. The strategy for noise vocoding followed the same method described by Chatterjee and her colleagues (2015). The procedure utilized the Greenwood frequency-place function, using sinewave as the carrier type for the vocoder and low-pass filtering (24 dB/octave filter, 160 Hz cutoff) for extracting speech envelope from each band. In line with comparable research reported by Chatterjee et al. (2015) for English, set numbers of spectral channels (i.e., 4-channel, 8-channel and 16-channel) were selected to create the final noise-vocoded output to simulate the level of hearing impairment experienced by the CI participants hearing natural speech.

Data collection: All stimuli were presented at an average level of 65 dB SPL (A-weighting) via a single Altec-Lancing Bluetooth speaker (IMW475), located approximately two feet from the listeners. Because this research was designed to determine which levels of TH listeners' performances in hearing noise-vocoded speech correspond to those of CI individuals, TH adults performed the task with both natural and noise-vocoded speech, while CI adults only heard the natural (non-vocoded) speech. Stimuli were blocked by stimulus type and randomized across listeners. Prior to testing, TH listeners heard a "warmup" playout of all stimuli to minimize the chance of surprise when hearing noise-vocoded speech for the first time. During testing, participants were seated in front of a computer screen as they listened to each emotional sentence. By means of a four-alternative, forced-choice task, they indicated which of the emotions they heard. Listeners' accuracy was recorded using DirectRT experiment monitoring software (Jarvis, 2014).

Method of analysis

Data Analysis: An analysis of variance (ANOVA) examined TH listeners' accuracy across natural and three noise-vocoded speech conditions (16-channel, 8-channel and 4-channel). Planned comparisons further explored the effects of speech condition. Independent *t*-tests were performed to examine differences between the CI and the TH listener groups, since the CI listeners heard natural speech only. For the natural speech data, a mixed design, two-way ANOVA (Listener group × Emotion) was conducted, followed by confusion matrices of intended and perceived emotions to further analyze the error patterns.

Results

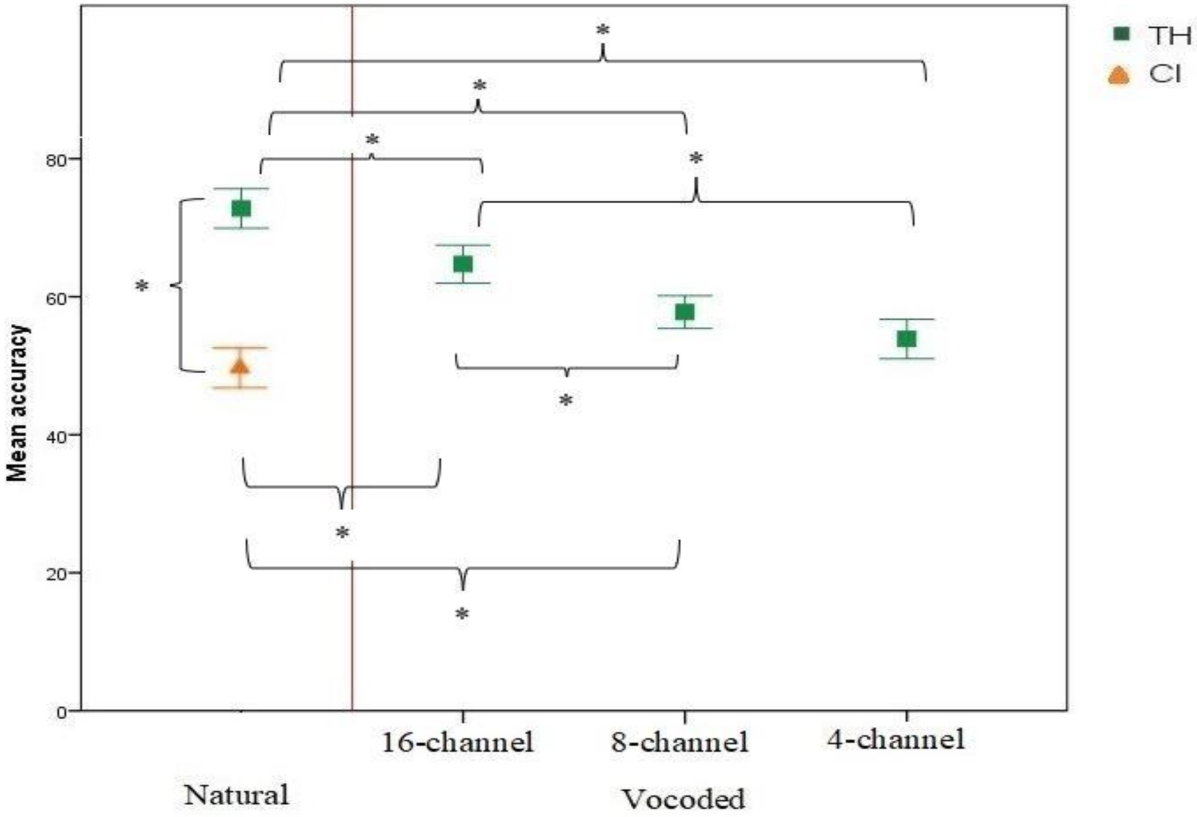


Figure 3.1. Average accuracy of TH listeners (*n=15*) (green square) separated by a red line from CI listeners *(n=15)* (orange triangle) for natural and vocoded speech stimuli, with standard error bars included. Note: * indicates $p< 0.05$.

Figure 3.1. shows the mean accuracy of emotional prosody recognition by TH listeners and CI listeners. The overall emotional prosody recognition by TH and CI listeners for natural speech is 72.8% and 50.3%, respectively. As expected, TH listeners showed higher mean accuracy than CI listeners for natural speech. In addition, TH listeners showed a nearly linear pattern of decrease in accuracy from the natural to the three vocoded speech conditions. The ANOVA results showed a significant main effect at the $p < 0.05$ level for TH listeners across the natural speech and the three noise-vocoded speech conditions [F (3, 56) =9.339, $p < 0.05$, $\eta p^2 = 0.333$]. Planned comparisons at $p < 0.05$ conducted for levels of stimulus type, indicated significant differences between natural speech and the three noise-vocoded versions (i.e., 16-channel, 8-channel and 4-channel), as well as significant differences between the 16-channel speech and the two lower channel conditions (8-channel and 4-channel). However, the 8-channel and 4-channel versions did not differ significantly from each other.

Independent-sample *t*-tests were separately used to compare the mean accuracy of CI listeners in the natural condition and TH listeners across different stimulus types (natural, 8-channel and 4-channel stimuli). Recall that this experiment was to explore whether CI listeners demonstrate deficits in emotional prosody perception in natural speech and to determine which channel setting of noise-vocoded speech heard by the TH listeners best approximates the CI natural speech level. More specifically, we test whether 8-channel or 4-channel noise-vocoded speech of TH adults' performance is similar to CI adult listeners' level. The results showed a significant difference between the CI listeners and TH listeners for the natural speech condition ($t = 4.015$, $p < 0.05$). Next, a modified Bonferroni test (Keppel, 1991) was applied for multiple comparisons between the CI listeners' natural speech accuracy scores and the TH listeners 16-,8-, and 4-channel noise-vocoded speech accuracy values, respectively. The result showed significant differences for the 16-channel noise-vocoded speech, and the 8-channel noise-vocoded speech ($t = 6.201$, $p < 0.033$ and $t =$

2.261, $p < 0.033$, respectively). There was no significant difference between CI listeners' natural speech and TH listeners' 4-channel noise-vocoded speech accuracy values ($t = 0.975$, $p = 0.338$, ns). These findings suggest that the performance of Mandarin-speaking CI listeners in natural speech is similar to that of TH adults at a lower channel setting, at 4-channel of noise-vocoded CI-stimulated speech, as compared to 8-channel Noise-vocoded speech. Recall that 8-channel noise-vocoded speech had been previously reported to be comparable to non-tonal language-speaking CI users (e.g., in English and Japanese).

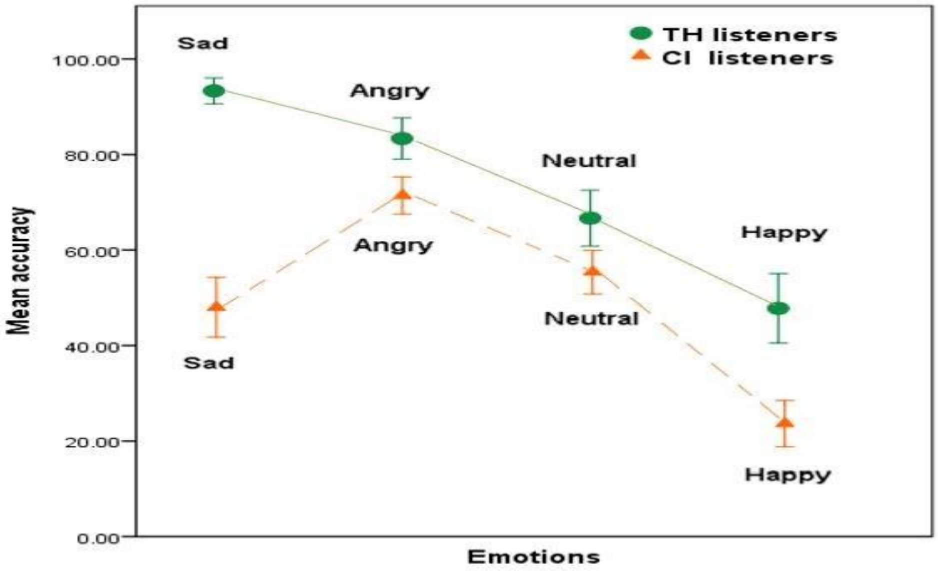


Figure 3.2. Accuracy of the four emotions in sentences grouped by listeners with TH *(n=15)* (green solid line) and with CIs *(n=15)* (orange dashed line) for natural speech, in order of highest accuracy, for TH listeners.

Figure 3.2. displays the mean accuracy of TH and CI listeners considered by four emotions in sentences. Visual inspection suggests the two listener groups showed a rather similar accuracy ranking for the "angry", "neutral", and "happy" emotions, but not the "sad" emotion, for which the two groups were clearly different. Nevertheless, similar to previous studies of emotional prosody in non-tonal languages, the “sad” emotion was more accurate judged than the “happy” emotion for both the TH and CI listeners. In addition, both listeners groups were better at identifying “angry” than “neutral”, with “happy” the most difficult to identify. A two-way ANOVA revealed a significant main effect for Listener group [F (1, 112) = 38.435, p < 0.05, ηp^2= 0.255] and Emotion [F (3,112) = 26.106, p < 0.05, ηp^2= 0.412]. In addition, there was a significant Listener group × Emotion interaction, indicating that TH listeners and CI listeners reacted differently across the four emotions [F (3,112) = 5.341, p < 0.05, ηp^2= 0.125].

Table 3.3. *Confusion matrices of intended and responded emotions by adult Mandarin-speaking TH listeners (panel A) and CI listeners (panel B). Listeners' intended emotions are presented vertically, and listeners' response emotions are organized horizontally. Numbers on the main diagonal (shown in bold and shaded) indicate the percentage of correct recognition accuracy. The row totals add to 100%.*

A

Intended	Response: TH listeners (N=15)			
	Angry	Happy	Sad	Neutral
Angry	**83.3**	12.2	0	4.5
Happy	18.9	**47.8**	1.1	32.2
Sad	2.2	0	**93.3**	4.5
Neutral	1.1	0	32.2	**66.7**

B

	Response: CI listeners (N=15)			
	Angry	Happy	Sad	Neutral
Angry	**73.3**	15.6	1.1	10.0
Happy	41.1	**23.3**	3.3	32.3
Sad	6.7	3.3	**47.8**	42.2
Neutral	2.2	2.2	38.9	**56.7**

Table 3.3. Panels A and B show error matrices (in % mean recognition) for each emotion, for the TH and CI listener groups. Values along the diagonal represent correct responses and the row totals add to 100%. The groups differences between TH and CI listener groups can be summarized by the following two error patterns: First, for the “happy” emotion, TH listeners tended to confuse “happy” sentences with “neutral” responses, while CI listeners showed a high degree of confusing “happy” sentences with “angry” responses. Second, for the “sad” emotion, CI listeners showed a high degree of judging “sad” sentences as “neutral”, whereas TH listeners did not show this pattern. There was little difference in the “angry” and “neutral” error patterns shown by the two listener groups.

In summary, the results of Experiment 1 suggest that Mandarin-speaking listeners with CIs show deficits in perceiving emotional prosody compared to their TH counterparts. There was decreased accuracy among the TH listeners from the natural speech to the three noise-vocoded speech conditions, with lower levels of accuracy corresponding to less spectral resolution in synthesis. In addition, the performance of Mandarin-speaking CI listeners in natural speech is similar to that of TH adults at a lower channel setting which is 4-channel, in contrast to 8-channel shown in previous comparable studies of non-tonal languages, e.g., English (Chatterjee et al., 2015; Chatterjee et al., 2015; Friesen et al., 2001; Zhu, et al., 2016). In terms of comparing TH and CI listeners perceiving different emotions in natural speech, CI listeners showed more difficult identifying the “sad” and “happy” emotions, as compared to TH listeners, although CI listeners’ accuracy order of “sad” and “happy” was similar to that of the TH listeners (“sad” > “happy”). In addition, CI listeners showed a high degree of confusing “happy” with “angry” and confusing “sad” with “neutral”, while TH listeners did not show these patterns.

Discussion

As expected, the results in the present study show that Mandarin-speaking adults with CIs have deficits in perceiving emotional prosody at the short sentence level, compared to their TH counterparts. The findings are consistent with previous studies on the vocal emotional recognition of CI users who speak non-tonal languages (e.g., English and Japanese) (Chatterjee et al., 2015; Gilbers et al., 2015; Luo et al., 2007; Pereira, 2000b; Zhu, et al., 2016), and are in general agreement with some limited studies of emotional prosody perception in Mandarin (Lu et al., 2017; Su et al., 2016). At least two interrelated mechanisms maybe involved in CI users’ deficits in emotional prosody processing: limitations of CI technology and hearing loss associated deficits in emotional processing.

First, although CI development has provided significant benefits for individuals with HI to access auditory information, some technological limitations of CI users remain. As previously mentioned, CI users cannot fully access the spectral-temporal fine structure information (e.g., F_0) provided by electrodes in the device, causing highly degraded acoustic cues that may result in difficulty processing prosodic cues (Chen, et al., 2014; Friesen et al., 2001; Jiam et al., 2017; Moore, 2003; Tan et al., 2016; Xu et al., 2004). In addition, CI users experience a narrower dynamic range compared to TH individuals, a limitation that may cause restricted intensity resolution and poor emotional prosody identification. A second possible problem for CI users is that the impaired auditory input may have a negative effect on social and emotional functions, which are important factors affecting their quality of life (Arlinger, 2003; Pereira, 2000b; Schorr, 2005). For example, Schorr (2005) stated that the effect of auditory deprivation during early childhood could limit positive relationships with parents and peers and cause problems interpreting the emotional expressions of others accurately. In summary, problems caused by CI technological limitations and negative effect of impaired auditory input, resulting in deficits in processing emotional prosody by CI users.

Accuracy was lower among TH listeners for the three noise-vocoded synthesized speech conditions, compared to those with natural speech, with lower levels of accuracy corresponding to less spectral resolution in synthesis. This pattern supports previous findings that TH listeners perform better with natural speech than with noise-vocoded speech, and there intelligibility is lower when spectral resolution is reduced (Chatterjee et al., 2015; Chatterjee et al., 2015; Friesen et al., 2001; Henry & Turner, 2003; Zhu, et al., 2016).

Across both listener groups, “angry” was better recognized than “neutral” and “happy”, with “happy” being the most difficult emotion to identify. However, for the “sad” emotion, CI listeners showed relatively more deficits in identifying this emotion than their TH counterparts. A possible explanation for this is that CI listeners have limited access to low-frequency signals (Chang, Bai, & Zeng, 2006) and pitch acuity (Thompson, William Forde, 2014), leading to potential difficulties in perceiving the “sad” emotion, which presents lower F_0 and smaller pitch variation compared to “angry” and “happy” emotions. The error patterns shown by CI listeners further confirmed this finding. CI listeners had a tendency to confuse “sad” with “neutral”, possibly because these two emotions present similar patterns of F_0 variations (Yildirim et al., 2004). In addition, consistent with the findings of previous literature (House, 1994; Luo et al., 2007; Pereira, 2000b), CI listeners showed a relatively high degree of misidentifying “happy” sentences as “angry,” while TH listeners did not show these patterns. As noted previously, “angry” and “happy” are both produced with relatively high F_0, amplitude variation with similar duration values.

Lastly, if the current results are compared with previous studies of English, some preliminary findings are noted that are consistent with the FV hypothesis. Recall that the FV hypothesis claims that that Mandarin (a tonal language) uses pitch for purposes of linguistic tone, and therefore, has less prosodic space to signal paralinguistic information (e.g., emotional states). Experiment 1 showed that the performance of CI listeners in natural speech is similar to that of TH adults at a low channel setting (i.e., 4-channel), as compared to the 8-channel setting noted in previous studies of non-tonal languages (e.g., English/Japanese) (Chatterjee et al., 2015; Chatterjee et al., 2015; Friesen et al., 2001; Zhu, et al., 2016).

In contrast, it is possible that the different channel performances of TH listeners in Mandarin and in English may be due to lack of temporal fine structure (TFS) information in the current CI technology, which is essential for Mandarin-speaking TH listeners to perceive speech perception (Wang, Xu, & Mannell, 2011). In addition, differences between the current study and the previous two studies conducted by Chatterjee (2015, 2016) in participants, testing materials, and study designs may well explain the different outcomes. Thus, while the current data are consistent with the FV hypothesis, it would be important to eventually compare tonal and non-tonal languages within a signal experimental design and with more complex stimuli in order to further substantiate any such claims.

CHAPTER 4

EXPERIMENT 2: PERCEPTION OF EMOTIONAL PROSODY IN STIMULI WITH MODIFIED ACOUSTIC CUES: TH VERSUS CI LISTENER GROUPS

Several studies have explored the role of acoustic cues in lexical tone perception by Mandarin-speaking CI listeners (Han et al., 2009; Liu, Tien-Chen et al., 2000; Wang, Liu, Zhang et al., 2012; Xu & Pfingst, 2003). While all findings suggest that the Mandarin-speaking CI users demonstrate deficits in lexical tone identification, there is controversy concerning the exact role of different acoustic cues in perceiving tones. Some studies suggest that secondary cues (duration and amplitude) play a more important role than expected. For example, some indicate that CI individuals use duration to rule out other tones (Wang, et al., 2012) and that amplitude plays a necessary role in helping them recognize lexical tones (Fu & Zeng, 2000; Luo & Fu, 2004; Wei et al., 2004). Other studies emphasize that F_0 is a primary cue playing a salient role in tone perception for CI users (He, Deroche, Doong, Jiradejvong, & Limb, 2016; Peng et al., 2017) and by implication suggest secondary cues are not so important.

There is similar controversy concerning how non-tonal language speaking CI listeners use acoustic cues to perceive prosodic information. Some studies find that individuals with HL rely more on temporal envelope cues (e.g., timing of the envelope) for complex tone perception, as compared with their TH counterparts (Moore & Moore, 2003; Moore & Carlyon, 2005). Relatedly, Kalathottukaren et al. (2015) state that the ability to resolve subtle changes over time, both in frequency and amplitude, is the key to recognizing prosodic patterns accurately for CI users. In contrast, some studies indicate that F_0 is crucial for prosody perception. Meister et al. (2009) examined the effects of modified F_0 on the perception of linguistic prosody by TH listeners and CI listeners. The results reveal that TH and CI listeners have increased accuracy in prosody perception when mean values of F_0 increase, while temporal and amplitude related features remain unchanged. In line with these findings, studies addressing the relationship between music and lexical tone perception suggest that CI users have great difficulty in perceiving melodic pitch due to poor tone perception (Hsiao, 2008; Tao, et al., 2015; Wang, Liu, Dong et al., 2012; Wang, Zhou, & Xu, 2011).

Although studies have addressed the role of acoustic cues in processing prosodic information, the results have focused only non-tonal language. To date, no research investigates how Mandarin-speaking CI users process different acoustic aspects of emotional prosody. In addition, there are no data addressing whether Mandarin-speaking CI listeners attend to acoustic cues for emotional prosody in a similar manner to TH listeners.

This chapter describes an experiment designed to explore whether enhancement of secondary cues for emotional prosody (amplitude and duration) could benefit Mandarin-speaking CI listeners. This is examined by modifying the prosodic cues for two contrasting emotions, “happy” and “sad”, and observing how CI listeners perceive these modifications.

Participants

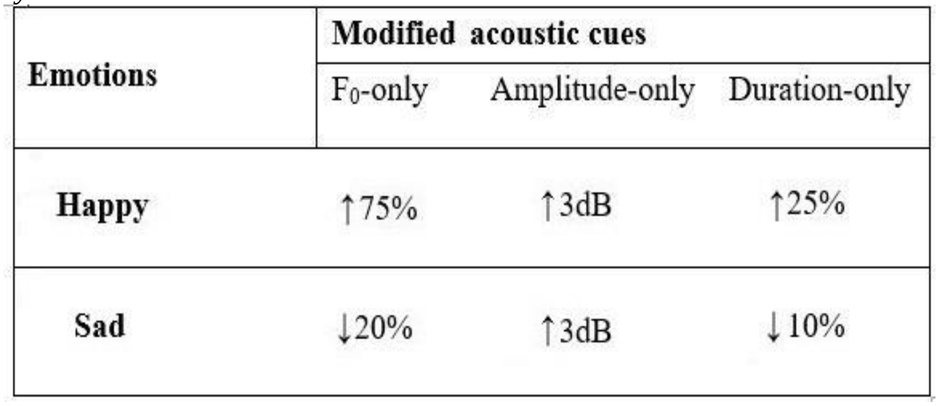
The same groups of participants from Experiment 1 took part in this experiment: Fifteen Mandarin-speaking adults with TH from the Texas cohort (group 1) and 15 Mandarin-speaking adults with CI from the China cohort (group 2).

Procedure

Stimuli: Spoken (audio) data were collected using the same three Mandarin sentences from Experiment 1. The stimuli were spoken by one male and one female (professional actors) with four common emotional categories ("angry", "happy", "sad", and "neutral"). However, the acoustic cues (F_0, amplitude, and duration) for each sentence were modified using a speech synthesizer (Kawahara, Masuda-Katsuse, & De Cheveigne, 1999). The aim was to create separate conditions of enhanced $F_{0,}$ duration, and amplitude cues for both the "happy" and "sad" emotions. "Happy" and "sad" emotions were chosen to be modified in this experiment based on their extremes of acoustic characteristics at the sentence level: "sad" exhibits lower $F_{0,}$ smaller amplitude variation, and longer duration than "happy" (Zhang, Sheng et al., 2006). In addition, emotional recognition accuracy by TH Mandarin-speaking listeners listening to natural speech and noise-vocoded speech suggest that "happy" is the most difficult emotion to recognize and the easiest to confuse with "angry" (Pak & Katz, 2017).

Based on spectral envelope scaling synthesis constraints described by Assmann & Nearey (2008) for English, F_0 was increased approximately 75% for "happy" and decreased 20% for "sad". In order to maximize the naturalness of the modified stimuli, the duration of "happy" sentences were increased by 25%, while "sad" sentences were decreased by 10%. The amplitude of "happy" and "sad" were both increased to enhance audibility. Detailed information concerning the way each emotion was modified is shown in Table 4.1. A total of 72 stimuli (3 sentences × 3 acoustic cues × 4 emotions × 2 talkers) were presented to the TH and CI Mandarin-speaking listeners. In the stimuli set, the sentences for "happy" and "sad" were modified, while those for “angry” and “neutral” were not. However, “angry” and “neutral” (unmodified) were provided as choices in order to remain consistent with the forced choice identification paradigm used in Experiment 1.

Table 4.1. *Acoustic modifications for the two emotions: “happy” and “sad” in three separate synthesis conditions.*



Emotions	**Modified acoustic cues**		
	F_0-only	Amplitude-only	Duration-only
Happy	↑75%	↑3dB	↑25%
Sad	↓20%	↑3dB	↓10%

Data collection: Participants were tested individually in a quiet room. All stimuli were presented at an average level of 65 dB SPL (A-weighting) via a single Altec-Lancing Bluetooth speaker (IMW475), located approximately two feet from the listeners. Stimuli were blocked by modified cue type (i.e., F_0, amplitude, and duration) and randomized across sentences and emotions for listeners. The participants were seated in front of a computer screen and first listened to a playout of a sentence. Next, by means of a four-alternative, forced-choice task, they indicated which of the emotions they heard. Listeners’ accuracy was recorded using DirectRT experiment monitoring software (Jarvis, 2014).

Method of analysis

Data Analysis: To analyze the mean accuracy of four emotions (“angry”, “happy”, “sad”, and “neutral”) judged by TH and CI listeners across the natural speech and three separate modified conditions (F_0-modified-only, amplitude-modified-only and duration-modified-only), a mixed-design Listener × (Emotion × Cue type) repeated measures ANOVA was conducted. Listener groups (TH and CI) were a between-subjects variable, and Emotion and Cue type (natural, F_0-modified-only, amplitude-modified-only and duration-modified-only) were within-subject, repeated-measured variables. In addition, confusion matrices of intended and perceived emotions were computed for the three modified conditions.

Results

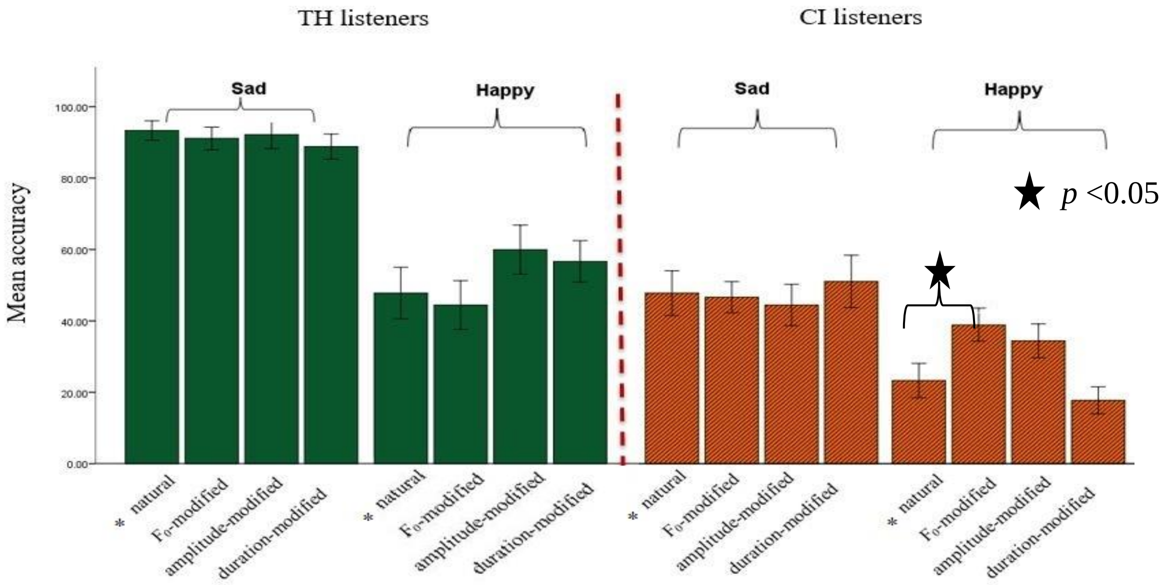


Figure 4.1. Judgment by TH listeners *(n=15)* (green solid bar) is separated by a dashed red line from CI listeners *(n=15)* (orange dashed bar) for “natural” and “modified” stimuli (i.e., having F_0, amplitude and duration enhanced cues), for “sad” and “happy” emotions. Note: * from Experiment 1.

Figure 4.1. shows the mean accuracy of TH and CI listeners identifying "happy" and "sad" emotions. Visual inspection suggests that TH listeners were better at identifying both emotions than CI listeners, especially for the "sad" emotion. In addition, TH listeners showed different patterns than the CI listeners in perceiving the "sad" and "happy" sentences. That is, TH listeners remained unchanged accuracy for "sad" emotion and with an increased accuracy in both the amplitude- and duration-modified conditions compared to natural (unmodified) condition for "happy" sentences. In contrast, CI listeners only showed minimal gain (3%) for duration-modified condition over natural in perceiving the "sad" emotion, while F_0-modified and amplitude-modified conditions provided notable increases in the accuracy of the "happy" emotion (16% and 11%, respectively).

Results of repeated-measure ANOVA revealed a significant main effect for Listener group [F (1, 28) = 73.142, $p < 0.05$, $\eta p^2 = 0.723$] and Emotion [F (3, 84) = 30.511, $p < 0.05$, $\eta p^2 = 0.521$]. Bonferroni-corrected pairwise analysis of the Emotion main effect showed that "sad" judgements were higher in accuracy than "happy". In addition, there was a significant Listener × Emotion × Cue type interaction [F (9,252) = 2.527, $p < 0.05$, $\eta p^2 = 0.083$], suggesting that groups responded differently to four different emotions across four different cue types. Planned comparisons at $p < 0.05$ were conducted for the Listener × Emotion × Cue type interaction. For CI listeners perceiving the "happy" emotion under natural speech and the three modified conditions, there was a significant difference ($t = 2.456$, $p < 0.05$) only between natural speech and the F_0-modified condition. This difference ($p < 0.05$) is indicated with a "star" in Figure 4.1.

Of key importance to this study were two, significant, two-way interactions (Emotion × Listener group and Cue type × Listener group). The results indicated that listener groups reacted differently across different Emotions [F (3, 84) = 4.874, $p < 0.05$, $\eta p^2 = 0.148$] and across different Cue types (both natural and modified conditions) [F (3, 84) = 4.585, $p < 0.05$, $\eta p^2 = 0.141$].

Confusion matrices were computed to further analyze these data. CI listeners' intended and responded emotions under natural speech and three of the modified conditions (F_0-modified-only, amplitude-modified-only and duration-modified-only) for only the "happy" and "sad" emotions are shown in Table 4.2.

Table 4.2. *Confusion matrices of intended and responded emotions by CI listeners under natural speech (panel A) and three modified conditions (F_0-modified-only, amplitude-modified-only and duration-modified-only) (panel B to D). CI listeners' intended emotions are presented vertically, and response emotions are organized horizontally. Numbers on the main diagonal (shown in bold and shaded) indicate the percentage of correct recognition accuracy. The row totals add to 100%.*

A

Intended	Response: CI listeners (N=15)			
	Natural speech			
	Angry	Happy	Sad	Neutral
Angry	**73.3**	15.6	1.1	10.0
Happy	41.1	**23.3**	3.3	32.3
Sad	6.7	3.3	**47.8**	42.2
Neutral	2.2	2.2	38.9	**56.7**

B

Intended	F_0-modified-only			
	Angry	Happy	Sad	Neutral
Angry	**63.3**	16.7	1.1	18.9
Happy	46.7	**38.9**	2.2	12.2
Sad	5.6	0	**46.7**	47.8
Neutral	3.3	3.3	47.8	**45.6**

C

Intended	Amplitude-modified-only			
	Angry	Happy	Sad	Neutral
Angry	**64.4**	22.2	0	13.3
Happy	36.7	**34.4**	0	28.9
Sad	3.3	3.3	**44.5**	48.9
Neutral	2.2	0	56.7	**41.1**

D

Intended	Duration-modified-only			
	Angry	Happy	Sad	Neutral
Angry	**60.0**	22.2	1.1	16.7
Happy	57.8	**17.8**	2.2	22.2
Sad	8.9	0	**51.1**	40.0
Neutral	3.3	2.2	50.0	**44.4**

Table 4.2. (panels A- D) shows the mean recognition rates (in %) for each emotion by CI listeners. Visual inspection of the numbers on the diagonal (shaded) suggest that “angry” sentences received the highest identifications, while “happy” sentences were the lowest. As for both the “happy” and “sad” emotions, the same error patterns were observed in the modified

speech conditions (panels B, C, and D). First, CI listeners showed a high degree of confusing “happy” with “angry”, especially under the duration-modified-only condition. Second, CI listeners exhibited a high degree of confusing “sad” with “neutral”, across the three modified conditions. In terms of using the duration cues for the “happy” and “sad” sentences, CI listeners demonstrated different patterns. That is, enhanced duration cues only minimally increased the accuracy of their “sad” judgements, while it decreased the accuracy of their “happy” judgements.

In summary, the results of Experiment 2 suggest that Mandarin-speaking TH better identifying “happy” and “sad” emotions under each modified condition than the CI listeners. For both the TH and CI listeners, “happy” was the harder emotion to identify while “sad” had higher accuracy. In addition, enhancement of duration cues minimally increased the accuracy of “sad” judgements (3%), while enhancement of amplitude cues provided notable increased in accuracy of “happy” emotion (11%).

Discussion

Experiment 2 addressed the ability of CI listeners in perceiving modified acoustic cues for emotional prosody. Given that Experiment 1 showed that CI listeners had unexpectedly low identification of “sad’ and “happy” sentences, this experiment used modified prosodic cues (i.e., F_0, amplitude, and duration) for those two emotions. This experiment was designed to determine whether modification of secondary cues could benefit CI listeners in perceiving emotional prosody.

The results show that for both TH and CI listeners, the “happy” emotion was the most difficult emotion to perceive while the “sad” emotion has the higher accuracy. Compared with the TH listeners who only showed improvements in the “happy” sentences under the amplitude- and duration-modified conditions, the CI listeners showed increased accuracy in both the “sad” and “happy” sentences, yet with small improvements (arrows in Figure 4.2). For the sentences produced with the “sad” emotion, the CI listeners showed a slight improvement in accuracy (3%) when listening to stimuli synthesized with longer duration, as compared with unmodified stimuli. For the “happy” emotion, the CI listeners showed notable increased accuracy (11%) in the amplitude-modified condition, compared with unmodified stimuli. Listeners’ error patterns (see Table 4.2) were helpful in further understanding these results. Compared to the TH listeners, the CI listeners slightly increased the judgements of “sad” sentences with the duration-modified condition, while they reduced the number of their “neutral” judgements. This supports the view that enhancement of secondary cues (in this case, duration) can benefit CI listeners to perceive the “sad” emotion. Since amplitude cues for prosody closely follow patterns of F_0 (Liu, Azimi, Tahmina, & Hu, 2012), it is expected that accuracy in identifying the “happy” emotion may increase under the F_0-modified condition. The data supported this expectation, as CI listeners’ accuracy increased from 23.3% in unmodified condition to 38.9% in the F_0-modified condition (panels A and B in Table 4.2).

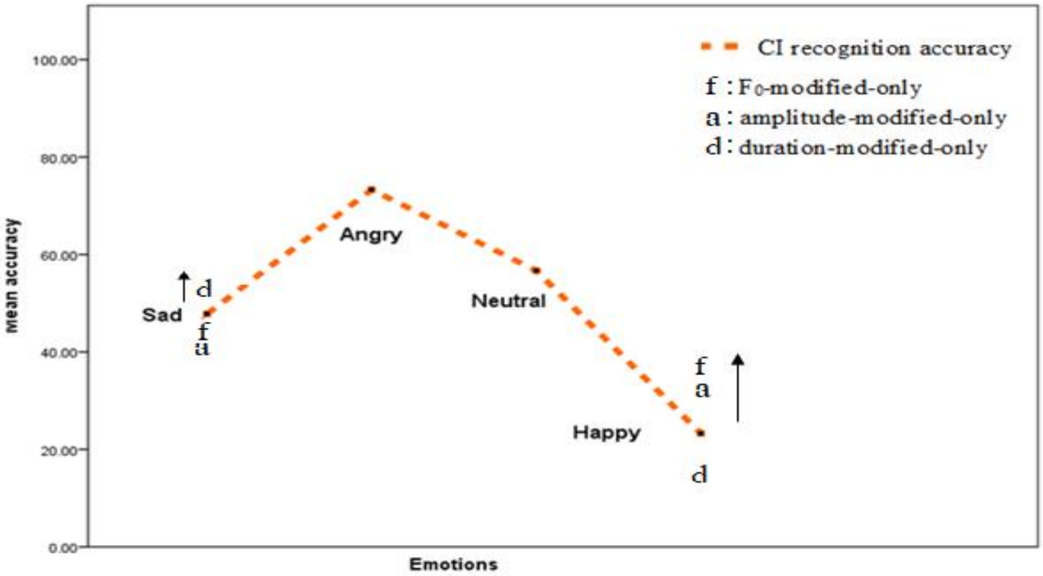


Figure 4.2. Accuracy of the four emotions in sentences, grouped by CI listeners *(n=15)* (orange dashed line) for natural and modified stimuli (i.e., having F_0, amplitude, or duration enhanced cues). The F_0-modifed-only condition is presented with a letter “f”, amplitude-modified-only condition is presented with a letter “a”, and duration-modified-only condition is presented with a letter “d”. Two black arrows represent increased accuracy.

The results of a mixed-design (Emotion × Cue type) repeated measures ANOVA showed a number of significant main effects and interactions, but critically no significant main effect of acoustic modifications (Cue types). Statistical analyses of the data indicated a significant difference between CI listeners’ accuracy for the natural and the F_0-modified stimuli of the “happy” sentences. Although we expected identification of the “sad” emotion under the duration-modified condition to be significantly different from the unmodified condition, this was not the case. While there are many potential reasons why increasing the acoustic cue of duration did not result in a pronounced increase in the perceptibility of the “sad” sentences by CI listeners, we suggest that issues with stimuli design are involved. Recall that duration cues of the “sad” sentences were decreased by 10% to maintain naturalness. An unintended consequence may have been that CI users were not provided sufficient duration cues such that they could perceive the “sad’ emotion. Alternatively, the “sad” stimuli are modified with low F_0. Since low frequency resolution is a known problem for CI users, lowering these already-low F_0 values in the current experimental design might not provide much benefit to CI listeners.

Table 4.3. *Confusion matrices of intended and responded “happy” emotion by TH listeners (panel A) and CI listeners (panel B), under natural speech and duration-modified conditions. The row totals add to 100%.*

A

Intended	Response: TH listeners (n=15)			
	Natural speech			
	Angry	Happy	Sad	Neutral
Happy	18.9	**47.8**	1.1	32.2
	Duration-modified-only			
	Angry	Happy	Sad	Neutral
Happy	35.6	**56.7**	0	7.7

B

Intended	Response: CI listeners (n=15)			
	Natural speech			
	Angry	Happy	Sad	Neutral
Happy	41.1	**23.3**	3.3	32.3
	Duration-modified-only			
	Angry	Happy	Sad	Neutral
Happy	57.8	**17.8**	2.2	22.2

Table 4.3. Panels A and B show error matrices expressed (mean % recognition rate) for the "happy" emotion for the TH and CI listener groups. Values along the diagonal represent correct responses, with row totals adding to 100%. Visual inspection of panels A and B suggest that both TH and CI listener groups show increased misjudgments of "angry" sentences under the duration-modified condition, with different patterns in identifying the "happy" sentences. As can be seen in panel A, TH listeners showed increased accuracy in "happy" judgements in the shorter duration-modified condition, compared to natural speech. However, this was not the case for CI listeners. Panel B shows that CI listeners demonstrated decreased accuracy in the "happy" judgements, with an even higher degree of confusing the "happy" with "angry" sentences under the duration-modified condition, as compared to natural speech. These findings suggest that modification of duration cues alone is not enough to help CI users to distinguish "happy" and "angry", whereas modification of spectral cues and amplitude cues which provide more high frequency information, can help CI users to identify the "happy" emotion.

To sum up, the present study suggests that the role of acoustic cues in processing prosodic information for Mandarin-speaking CI users differs for different emotions. In the current study, increased duration cues were found to provide minimal gain in perceiving the "sad" emotion for CI listeners, while increased amplitude and F_0 information provided more appreciably increased accuracy for the "happy" emotion. These findings suggest that individual cues for different emotions vary for CI listeners. In addition, these findings have implications for emotional prosody conversion speech synthesis programs designed for Mandarin which currently only focus on F_0 patterns (i.e., enhanced duration cues for "sad" and enhanced amplitude cues for “happy” for Mandarin-speaking CI listeners) (Jiang, Zhang, Shen, & Cai, 2005; Tao, et al., 2006; Wen, Wang, Hirose, & Minematsu, 2011). Clinical implication will be further discussed in Chapter 8.

CHAPTER 5

EXPERIMENT 3: PRODUCTION OF EMOTIONAL PROSODY BY TH AND CI TALKER GROUPS: ACOUSTIC ANALYSIS AND EMOTION CLASSIFICATION

Although research on emotional prosody has mainly focused on perception, a few studies investigating non-tonal language (e.g., English and Japanese) have shown that CI users demonstrate significant impairments in emotional prosody production, as compared with their TH counterparts (Chatterjee et al., 2015; Chatterjee et al., 2016; Jiam et al., 2017; Nakata et al., 2012; Wang, et al., 2013). Nakata el al. (2012) report that Japanese-speaking children with CIs (age range 5-13 years, SD = 1.97) imitate “surprise” and “disappointment” in speech significantly more poorly than their TH peers. Wang et al. (2013) conducted studies using a 10-year-old native English-speaking female’s speech as a model to compare the imitated production results between English-speaking children with TH (mean age = 5.2, SD = 0.8) and with CIs (mean age = 6.0, SD = 0.7). The results show that children with CIs demonstrated less accurate imitations of “happy” and “sad” sentences, and produced these sentences with smaller F_0 values differences, compared with their TH peers. Recent studies by Chatterjee and colleagues (2015, 2016) are also support these findings. The results suggest that English-speaking children with CIs produce smaller “happy”/ “sad” contrasts in mean intensity, mean F_0, and spectral centroid values compared to their TH peers.

Little is known about how tonal language-speaking adults with CIs produce emotional prosody, as compared with their TH counterparts. For instance, there has been no research addressing the acoustic cues related to potential differences in emotional prosody production in the speech of Mandarin-speaking adults with CIs.

This chapter presents data analyzing and comparing acoustic cues produced in emotional prosody by Mandarin-speaking CI adults and TH adults. The purpose of the acoustic analysis was to determine whether production of those acoustic cues (F_0, amplitude, and duration) differs for talkers with CIs compared to their TH counterparts. Mandarin-speaking CI talkers are expected to demonstrate decreased mean values of F_0, intensity and increased sentence duration values in their sentence production, compared to Mandarin-speaking TH individuals. In addition, it was of interest to know which acoustic measures in the data are most predictive of three different emotions (“angry”, “happy”, and “sad”) produced by TH and CI talkers, and whether CI talkers use the same acoustic measures as TH talkers to predict emotions.

Participants

A total of 30 talkers participated in this experiment: fifteen CI participants from the same China cohort in Experiment 1, and fifteen Mandarin-speaking adults with typical hearing (TH) (Group 3), recruited from the campus of University of Texas at Dallas. Of the 15 TH participants, 6 were male and 9 were female, matching the gender pattern in the CI group. TH participants were native speakers of Mandarin Chinese with no self-reported history of speech, hearing, emotional or language impairment. Similar to the China cohort, the TH participants all reported at least a middle school education level. The age range of participants with TH was from 23 to 39 years (mean≈ 26.8 years, SD = 4.3), providing an age match to CI group.

Procedure

Stimuli: Stimuli were the same three Mandarin sentences described in Experiment 1(see Appendix B). Each talker was asked to produce the sentences with three emotions (“angry”, “happy”, and “sad”). Thus, nine sentences were generated by each talker, yielding a total of 270 sentences (9 sentences × 15 talkers × 2 talker groups).

Data collection: Talkers were recorded while seated in a double-walled, sound attenuating booth. The primary investigator (PI) led each talker through the recording protocol using introductory instructions. To minimize PI influence on the recording, written instructions in Chinese characters were presented through slides shown on a computer monitor at eye-level, approximately 18” (45cm) from each talker. A practice dialogue script then followed the instructions to ensure that talkers fully understood the protocol before the experiment begun. Talkers’ productions were audio recorded using an Audio-Technica directional microphone (Model: ATR6550) located approximately 15 cm from the mouth, and a TASCAM portable digital recorder (DR-05R). All recordings were stored using a 16-bit and 44.1 KHz (mono) sample rate format.

Each talker produced the target sentences heard in the perceptual experiment (e.g., the semantically neutral phrase: $\mathit{Mang}_2\mathit{Ni}_3\mathit{De}_1\mathit{Shi}_4$ $\mathit{Qing}_2\mathit{Ba}_{1.}$*忙你的事情吧. “You go and do your own stuff”*) in the following emotion contexts: “angry”, “happy” and “sad”. All emotions were elicited using scenarios for target sentences (Busso et al., 2017; Hubbard, Faso, Assmann, & Sasson, 2017). Additional instructions with target sentences in Chinese characters were used. First, talkers were notified which emotion they needed to produce, followed by emotional scenarios described in written paragraphs containing highlighted target sentences. Talkers were given enough time to experience each emotional scenario. Next, they were asked to produce the target sentences with their best expressive ability. The experiment was self-paced, with each talker scrolling through the slides themselves. Different randomized orders were used across participants. As an example, the emotional scenario for the sentence *"You go and do your own stuff"* in "angry" context is detailed in Table 5.1.

Table 5.1. *Example emotional scenario dialogue for the sentence "You go and do your own stuff" in the "angry" emotion context.*

Emotional scenario Dialogue:	Target sentence	*Mang$_2$Ni$_3$De$_1$Shi$_4$ Qing$_2$Ba$_1$.* *忙你的事情吧.* *"You go and do your own stuff."*
	Target emotion	"Angry"
	Scenario	I have a deadline tomorrow. However, my laptop has some problems and needs to be fixed. My cousin said he can fix it with no problem. I trusted him, but he accidentally formatted the whole hard drive. I cannot believe it! All my data was not backed up! I am very angry: "*You go and do your own stuff*! If you can't fix it, don't touch my laptop!"

Method of analysis

Data Analysis: To determine how individuals with TH and CI use acoustic cues for emotion during speech, a series of acoustic measures (mean F$_0$, F$_0$ range, mean intensity, intensity range, and duration) were obtained. These measures were selected based on previous studies showing that these acoustic measures for emotions vary meaningfully between TH and CI talkers (Chatterjee et al., 2015; Chatterjee et al., 2016; Jiam et al., 2017; Nakata et al., 2012; Wang, et al., 2013). Relevant patterns were then analyzed using statistical means, including machine learning techniques.

Acoustic analysis: All audio recorded data were high pass filtered to remove the DC (direct current) using open source speech software (WaveSurfer, version 1.8.8) (Sjolander & Beskow, 2010). Each acoustic measure (mean F_0, F_0 range, mean intensity, intensity range, and duration) was analyzed, using Praat software (Boersma & Weenink, 2016). Mean F_0 was measured using the autocorrelation method described in Praat. F_0 range was measured as the ratio of the maximum to the minimum F_0 reached within each sentence, and intensity was range measured as the difference between the highest and the lowest intensity reached within each sentence. To reduce F_0 variability resulting from gender differences in vocal tract length, we used a logarithmic semitone-scale to transfer female/male pitch, following this formula: semitone = 12× log2(Hz) (Liu, et al., 2012; Thompson, Marin, & Stewart, 2012). Therefore, all statistical analyses were performed on the female/male F_0 values in semitones.

First, a two-way (Emotion × Talker group) Multivariate ANOVA (MANOVA) was conducted. There were five dependent variables (mean $F_{0,}$ F_0 range, mean intensity, intensity range and duration) and two independent variables: emotion ("angry", "happy", and "sad") and talker group (TH and CI talkers).

Next, we analyzed the data with a similar approach to that described by Chatterjee et al. (2015, 2016), using each acoustic measure (mean $F_{0,}$ F_0 range, mean intensity and intensity range) of "happy" compared with "sad", in order to explore whether there are differences between these two talker groups. Mean F_0 contrast of "happy"/ "sad" (H/S) was measured using the ratio of each acoustic measure of the "happy" sentences, and the same measure obtained with the “sad” sentences. F_0 range contrast of H/S was measured with the difference in the F_0 range for “happy” and “sad” sentences. Mean intensity contrast of H/S was measured using the intensity difference between these two emotion sentences. Intensity range contrast of H/S was measured with the difference in ratio of maximum intensity to minimum intensity for “happy” and “sad” sentences. Independent *t*-tests were performed to compare the four acoustic measures of H/S contrasts (Mean $F_{0,}$ F_0 range, mean intensity, and intensity range) between TH and CI talker groups.

Emotion classification: To further determine which acoustic measures are most predictive of the three emotions (“angry”, “happy” and “sad”) produced by the Mandarin-speaking adults with CIs and their TH counterparts, a decision tree algorithm was used in WEKA (Waikato Environment for Knowledge Analysis) software (Witten, Frank, Hall, & Pal, 2016). Decision trees are considered to be a widely-used machine learning method, employing a tree-like, decision-modeling graph that shows the classification after-effects in an easily understandable manner (Bhargava, Sharma, Bhargava, & Mathuria, 2013; Selby & Porter, 1988). The “J48” decision tree-inducing algorithm (WEKA’s implementation of C4.5) was used to represent classifiers and implement a top-down induction (Rajesh, Maiti, & Reena, 2018). In Weka, attributes are considered as instances and features. The original dataset with seven attributes (mean F_0, F_0 range, mean intensity, intensity range, duration, gender and emotion) for the emotional sentences produced by individuals with TH and CI were preprocessed in Weka and stored in a database in a ARFF (Attribute Relation File Format) file. Cross validation is a standard method for testing the accuracy of each algorithm in machine learning experiments. Four-fold cross validation was conducted for the seven attributes.

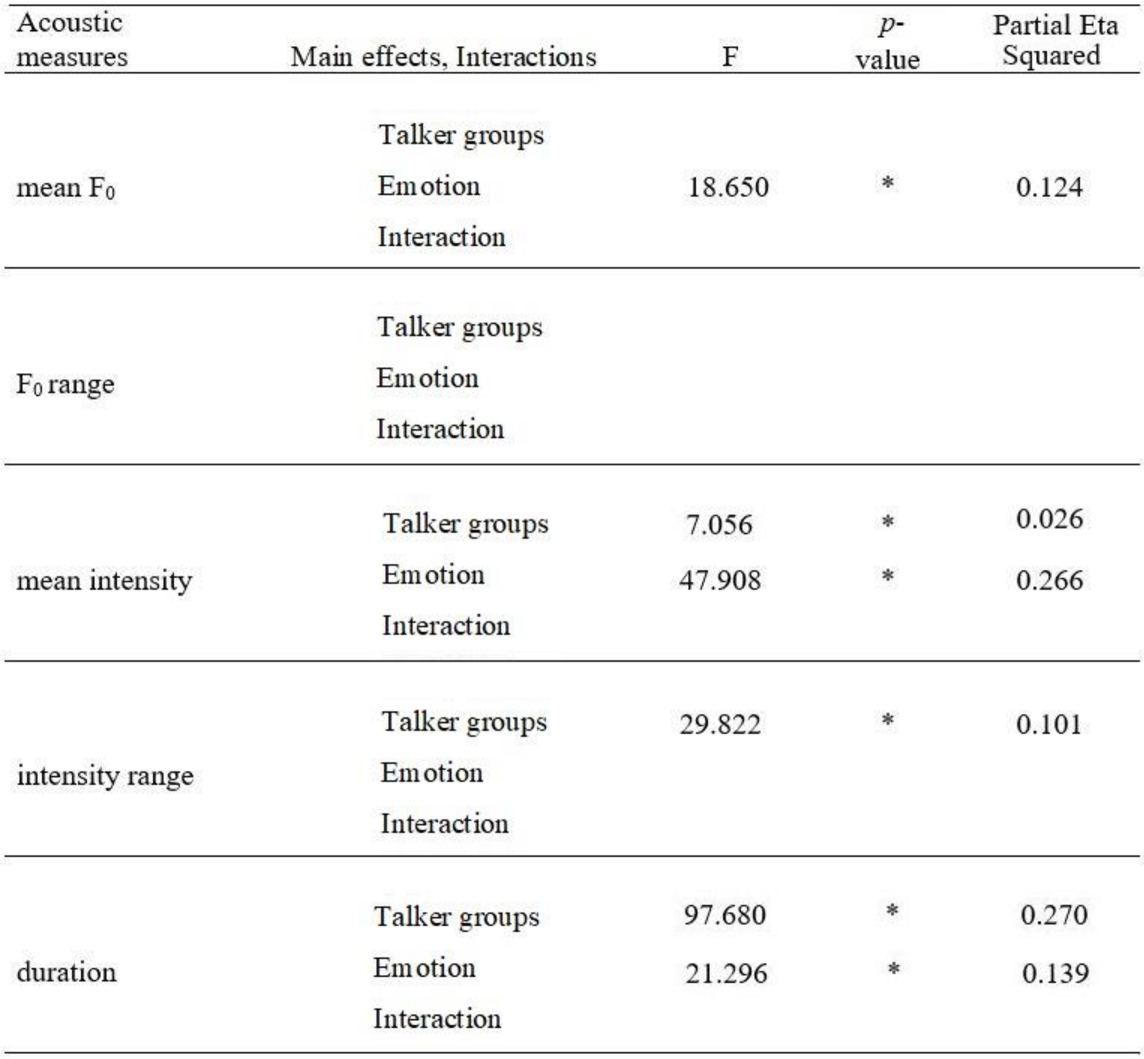
The purpose of emotion classification is to determine which attributes are most predictive of three emotions by each talker group. Therefore, noninformative attributes should be removed prior to applying the J48 algorithms in WEKA. The receiver operating characteristic (ROC) analysis choose optimal models and dispose suboptimal ones prior to the classification (Guimaraes et al., 2010; Huang, Hung, & Chen, 2010). The ROC curve refers to the ratio of true positive rate versus false positives rate (Fawcett, 2006), and the area underneath the ROC curve represents the total judgement of a test (Beshah & Hill, 2010). That is, the perfect prediction method representing 100% specificity (no false positives) has a ROC area equal to 1.0. A random guess (no prediction value) will tend toward 0.5 (Beshah & Hill, 2010; Fawcett, 2006). Therefore, any area value for classifiers between 0.5 and 1 has prediction value, and the higher the number, the better the prediction value. Any classifier showing an area value less or equal to 0.5 will be removed to obtain higher quality optimal models.

Results

Acoustics analysis results:

A summary of MANOVA result is displayed in Table 5.2.

Table 5.2. *Summary of MANOVA, including five acoustic measures (mean F_0, F_0 range, mean intensity, intensity range, and duration) for TH talkers and CI talkers produced "angry", "happy", and "sad" emotions. Note: *= $p < 0.05$.*



Acoustic measures	Main effects, Interactions	F	*p*-value	Partial Eta Squared
mean F_0	Talker groups			
	Emotion	18.650	*	0.124
	Interaction			
F_0 range	Talker groups			
	Emotion			
	Interaction			
mean intensity	Talker groups	7.056	*	0.026
	Emotion	47.908	*	0.266
	Interaction			
intensity range	Talker groups	29.822	*	0.101
	Emotion			
	Interaction			
duration	Talker groups	97.680	*	0.270
	Emotion	21.296	*	0.139
	Interaction			

The main findings are also shown graphically in Figures 5.1 to 5.5. These graphs summarize the five acoustic measures (mean F_0, F_0 range, mean intensity, intensity range, and duration) for TH talkers (n=15) and CI talkers (n=15) in producing of "angry", "happy", and "sad" emotions. In these figures, TH talkers are shown in solid bars with different acoustic measures in "angry", "happy" and "sad" emotions. CI talkers are shown in dashed bars under the same conditions.

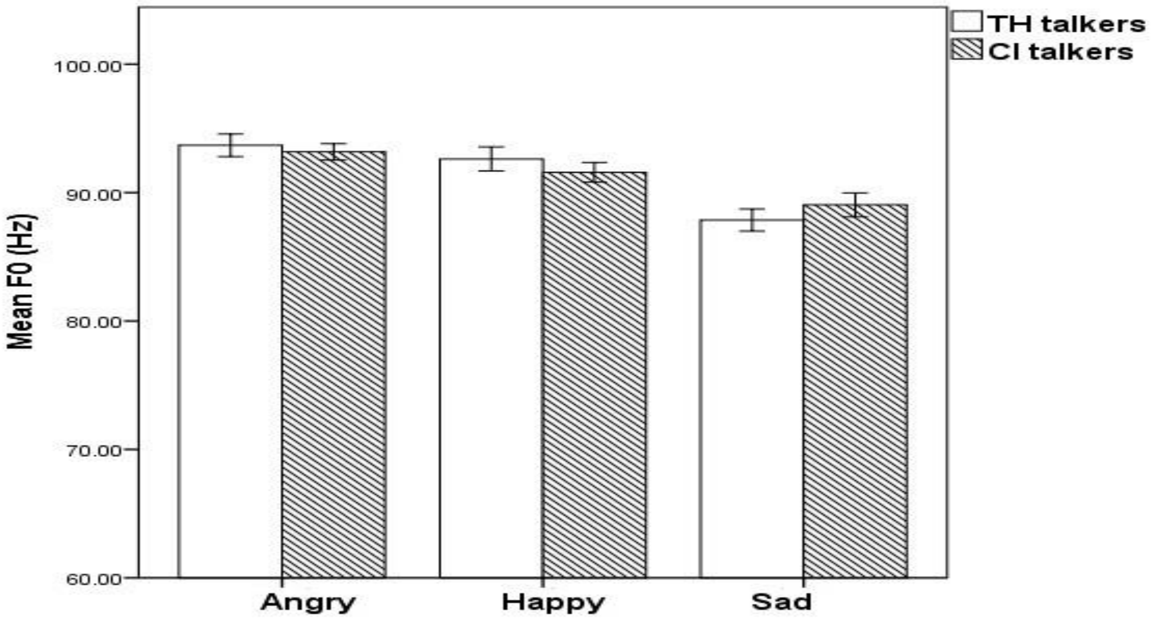


Figure 5.1. The mean F_0 produced across three sentences by individuals with TH (*n=15*) (solid bars) and with CIs (*n=15*) (dashed bars) in “angry”, “happy” and “sad” emotions.

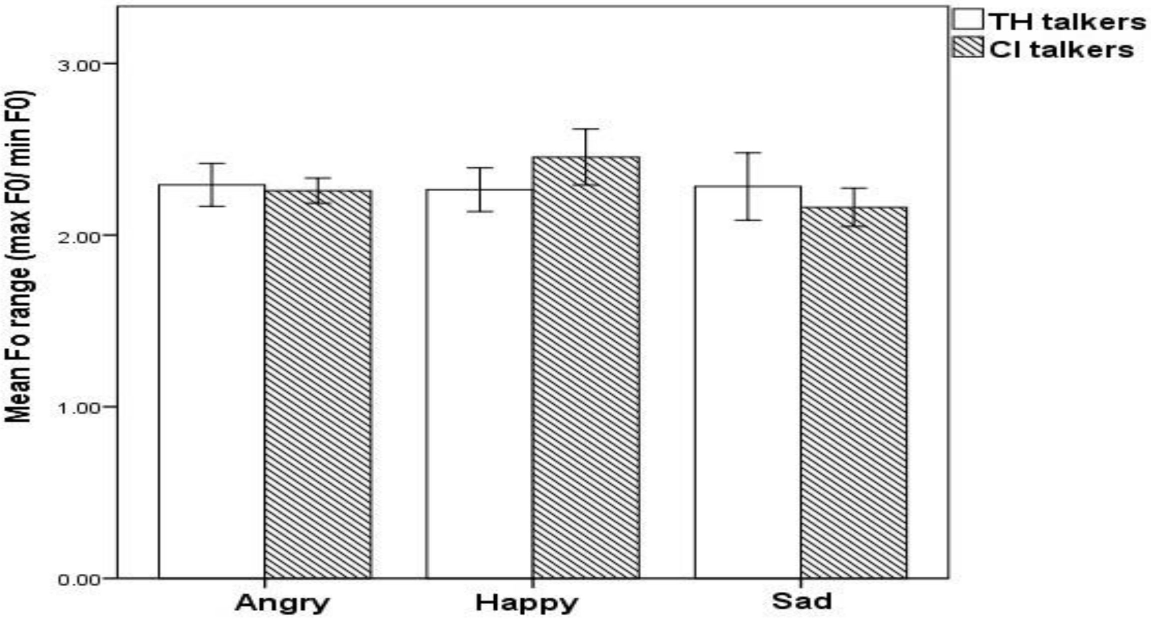


Figure 5.2. The mean F_0 range produced across three sentences by individuals with TH (*n=15*) (solid bar) and with CIs (*n=15*) (dashed bar) in “angry”, “happy” and “sad” emotions.

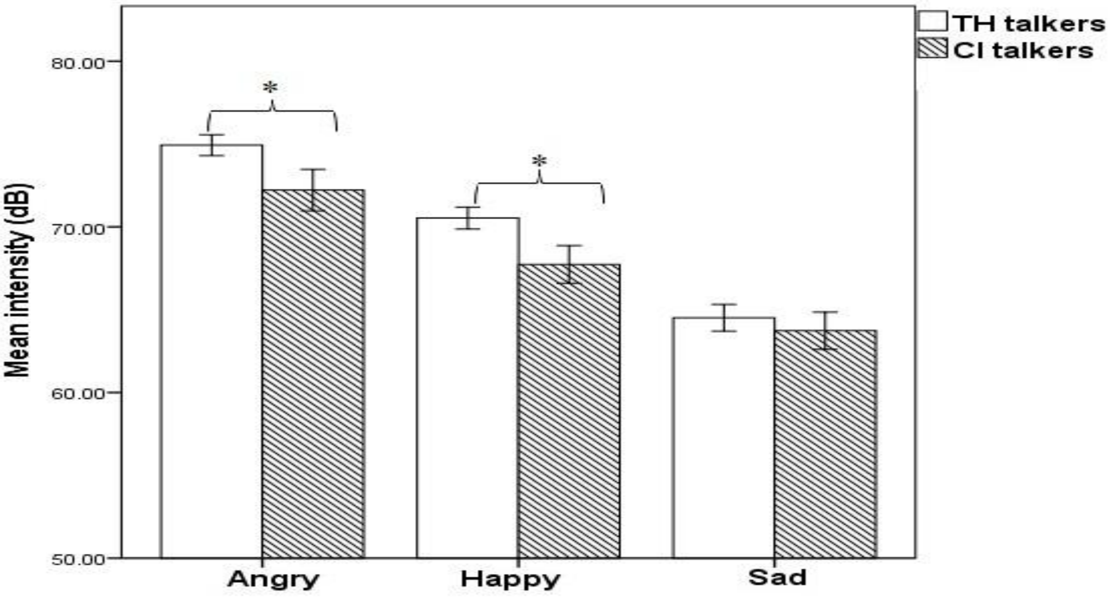


Figure 5.3. The mean intensity produced across three sentences by individuals with TH (*n=15*) (solid bar) and with CIs (*n=15*) (dashed bar) in “angry”, “happy” and “sad” emotions.

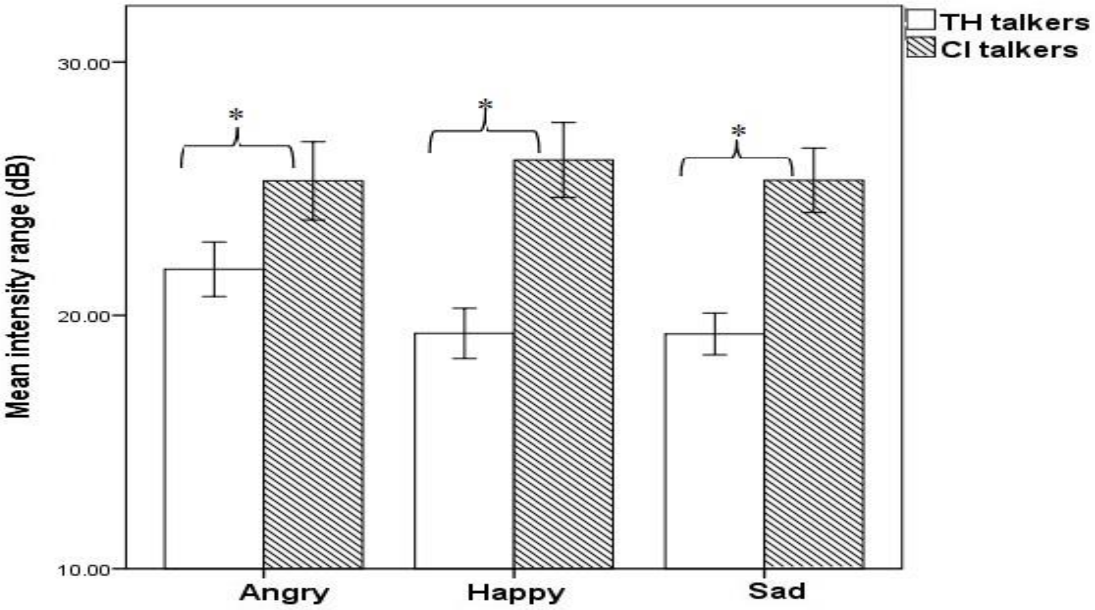


Figure 5.4. The mean intensity range produced across three sentences by individuals with TH (*n=15*) (solid bar) and with CIs (*n=15*) (dashed bar) in “angry”, “happy” and “sad” emotions.

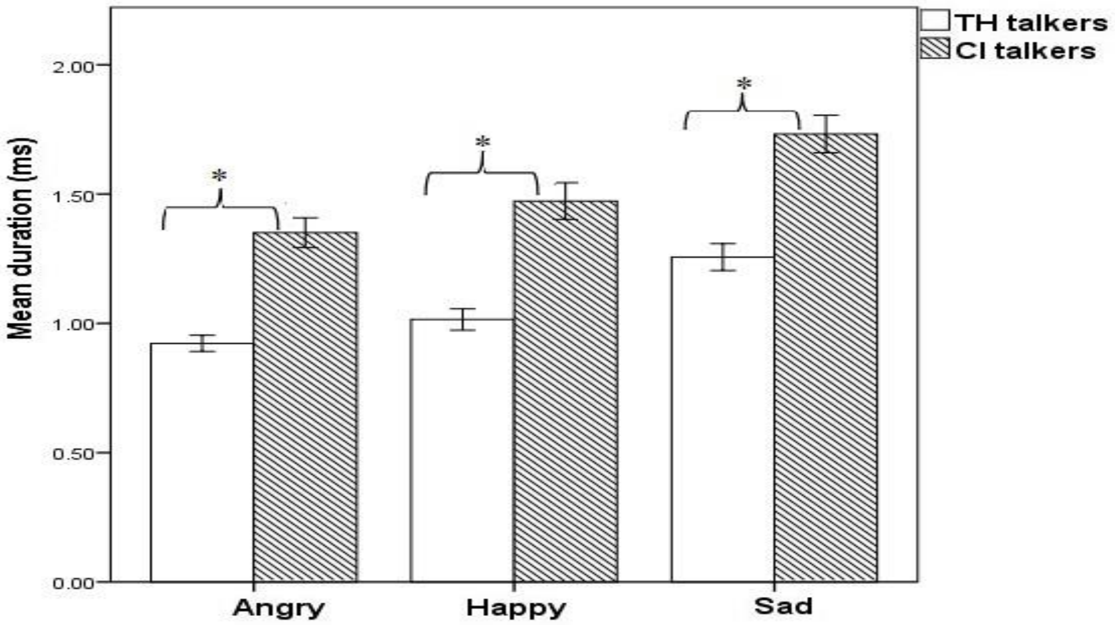


Figure 5.5. The mean sentence duration produced across three sentences by individuals with TH (*n=15*) (solid bar) and with CIs (*n=15*) (dashed bar) in “angry”, “happy” and “sad” emotions.

Five patterns are observed from Figures 5.1 to 5.5. The first pattern suggests that TH and CI talkers produced similar patterns of mean F_0 in “angry”, “happy”, and “sad” sentences. The second pattern shows that CI talkers exhibited similar patterns of F_0 range in “angry”, “happy”, and “sad” sentences. The third pattern shows that CI talkers demonstrated decreased mean intensity values in “angry”, “happy”, and “sad” sentences, compared to TH talkers. The fourth pattern shows that CI talkers produced an increased intensity variability compared to TH talkers in three emotions. The fifth pattern shows that female and male CI talkers produced longer sentence duration in “angry”, “happy”, and “sad” sentences, compared to their TH counterparts.

A two-way (Emotion × Talker group) MANOVA for TH and CI talkers revealed a significant main effect for Talker group (TH and CI) for intensity [F (1, 264) = 7.056, p <

0.05, $\eta p^2 = 0.026$], intensity range [F (1, 264) = 29.822, $p < 0.05$, $\eta p^2 = 0.101$] and duration [F (1, 264) = 97.680, $p < 0.05$, $\eta p^2 = 0.270$]. Also, there was a significant main effect for Emotion [F (10, 522) = 13.299, $p < 0.05$, $\eta p^2 = 0.203$]. However, there is no significant Emotion × Talker group interaction.

Following the approach described by Chatterjee et al. (2015, 2016), we further analyzed the acoustic measures (mean $F_{0,}$ F_0 range, mean intensity and intensity range) of CI and TH talkers using H/S contrasts. The results of four separate independent *t*-tests showed that there is a significant difference in H/S contrasts on the ratio of mean F_0 ($t = 2.707$, $p < 0.05$), as compared with TH. No other significant differences in H/S contrasts.

Emotion classification results:

Preliminary tests of each attribute showed that intensity range and gender of TH talkers were removed as suboptimal classifiers due to the results of ROC areas (less or close to 0.5). Therefore, five attributes (mean F_0, F_0 range, mean intensity, duration and emotion) were preprocessed in WEKA for TH. The J48 algorithm took all samples as input (3 sentences × 3 emotions × 4 acoustic measures × 15 talkers for TH group). Followed the same test procedure, gender, intensity range and F_0 range attributes were removed for CI talkers. Therefore, four attributes (mean F_0, mean intensity, duration and emotion) were preprocessed in WEKA for CI. The J48 algorithm took all samples as input (3 sentences × 3 emotions × 3 acoustic measures × 15 talkers for CI group). The results of J48 algorithm for talkers with TH (n=15) are presented in panel A of Figure 5.7, and the results for talkers with CIs (n=15) are presented in panel B.

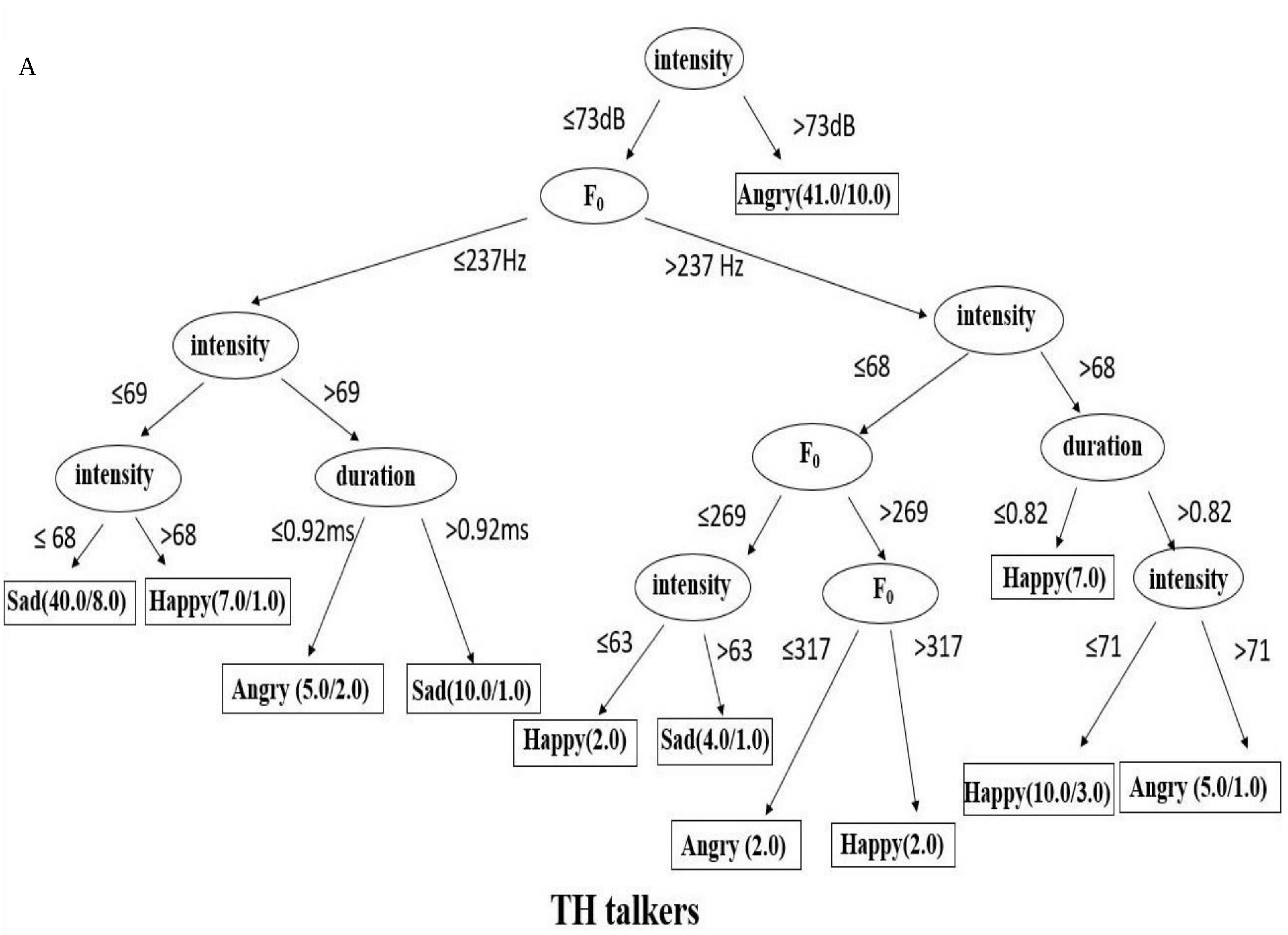
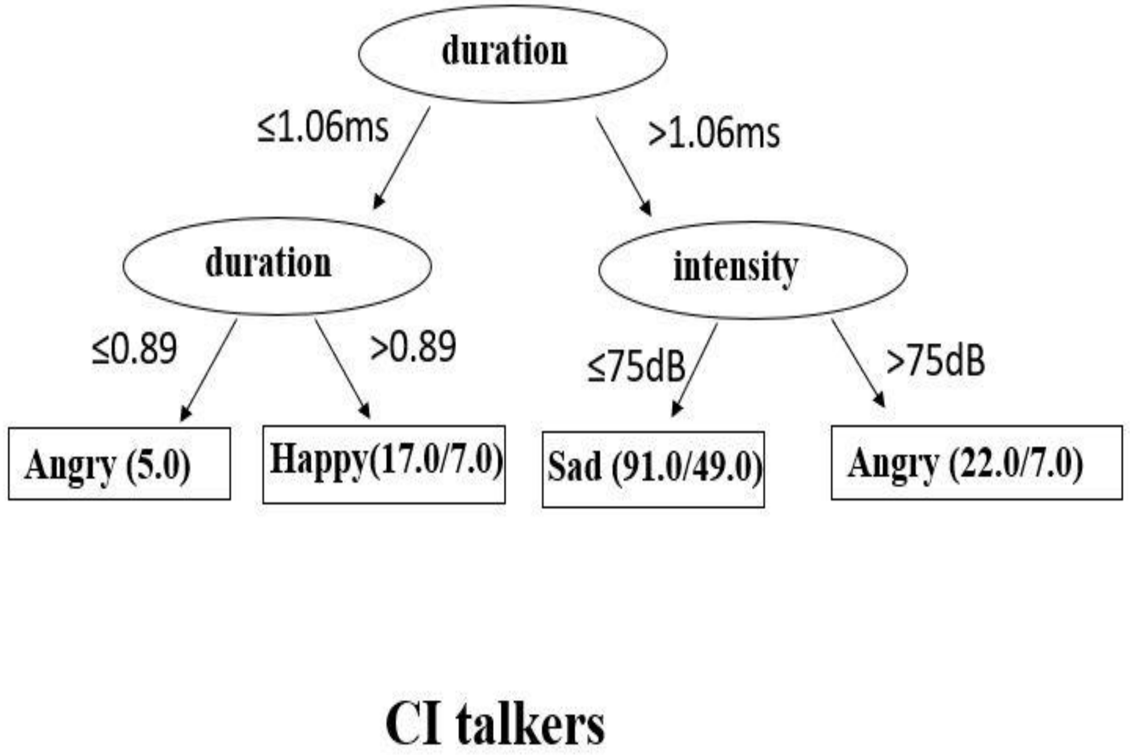


Figure 5.7. Classification of acoustic measures in production using J48 algorithm in WEKA. CI talkers(n=15) and TH talkers (n=15) are presented in Panel A and B separately.

B.



CI talkers

Figure 5.7. (cont.) Classification of acoustic measures in production using J48 algorithm in WEKA. CI talkers(n=15) and TH talkers (n=15) are presented in Panel A and B separately.

Figure 5.7. shows multiple paths from root to leaf, representing the procedure used to predict each different emotion. Visual inspection of two different decision trees showed that different talker samples presented different paths, suggesting that TH talkers and CI talkers produced different proportions of acoustic measures to predict different emotions. The results of algorithm showed that TH talkers (panel A) presented a more interactive complex decision tree than CI talkers (panel B). For example, there are 11 leaf nodes representing a classification and 11 branching factors representing a choice among the three emotions for TH talkers, whereas there are only four leaf nodes and three branch nodes for CI talkers. Taking the “angry” data as an example, the algorithm model showed that TH talkers correctly distinguished the most “angry” emotions using mean intensity in the decision tree. The rest of correctly-distinguished “angry” responses were predicted by F_0, then followed by mean intensity. There were two different paths for mean intensity to classify “angry”. One path used duration to distinguish “angry”. The other further divided into two branch nodes to predict “angry”, one using F_0 and the other using duration, followed by intensity. In contrast, according to the algorithm results, duration is the most predictive classifier used by CI talkers to distinguish “angry” and “happy” from “sad”. In addition, the algorithm results showed that both duration and intensity are most important classifier to predict “sad” emotion. The findings of emotion classification indicated that secondary cues (amplitude and duration) are the most predictive in classifying three emotions produced by CI talkers.

Next, confusion matrices containing information about actual and predicted classifications are presented in Table 5.3.

Table 5.3. *Confusion matrices of classifications by adult Mandarin-speaking TH talkers (panel A) and CI talkers (panel B). Talkers’ predicted classifications are presented vertically, and algorithm classifications are organized horizontally. Numbers on the main diagonal (shown in bold and shaded) indicate the correct instances.*

A

	Algorithm classification for TH talkers		
Predicted	Angry	Happy	Sad
Angry	**33**	10	2
Happy	19	**16**	10
Sad	1	7	**37**

B

	Algorithm classification for CI talkers		
Predicted	Angry	Happy	Sad
Angry	**23**	16	6
Happy	17	**20**	8
Sad	9	20	**16**

Table 5.3. Panels A and B present error matrices (numbers of correctly-classified emotions) for the TH and CI talkers. Values along the diagonals represent correct numbers of instances, with row totals adding to 45. The results of decision tree algorithm showed that correctly-classified emotions for TH and CI are 63.7% and 43.7%, respectively. Visual inspection suggests that the correctly-classified emotions by algorithm classification for the TH talkers (total numbers of correctly-classified emotions = 86) are higher than that of the CI talkers (total numbers of correctly-classified emotions = 59). In addition, the “sad” emotion produced by the CI talkers received poorer classification and a higher degree of misclassifying “happy” as “angry”, as compared to that of the TH talkers. The findings suggest that CI talkers present different acoustic patterns that were used to classify the three emotions, as compared to the TH talkers.

In summary, the acoustic analysis results of Experiment 3 suggest that Mandarin-speaking TH and CI talkers differ in using acoustic cues to produce three emotions. CI talkers demonstrate decreased mean intensity values with increased intensity variability, and increased sentence duration values in their sentence production, as compared with TH talkers. Also, CI talkers produced smaller H/S contrasts on mean F_0 than TH talkers. Furthermore, emotion classification results suggest that different acoustic cues were used by the TH and CI talkers to classify the three emotions in their speech. According to the algorithm results, CI talkers used duration as the most important classifier, followed by intensity, to predict the three emotions. On the contrary, algorithm results of the TH talkers show that intensity is the most important classifier, followed by F_0 to predict the three emotions.

Discussion

In this emotional prosody production experiment, it was hypothesized that Mandarin-speaking CI talkers demonstrate decreased mean values of F_0, and intensity, and increased sentence duration values in their sentence productions, compared to Mandarin-speaking TH individuals. However, the results of the acoustic analysis varied across the different acoustic measures. That is, these findings suggest significant decreased mean intensity values with increased intensity variability, and increased sentence duration values in Mandarin-speaking CI talkers' sentence productions, while no significant mean F_0 and F_0 range differences in three emotions, compared to their TH counterparts. However, mean F_0 in H/S contrasts produced by CI talkers is significant difference with TH talkers. That is, CI talkers produced smaller H/S contrasts in mean F_0 than TH talkers, which is consistent with previous studies in English (Chatterjee et al., 2015; Chatterjee et al., 2016).

First, as expected, both the female and male CI talkers demonstrated increased sentence duration values in their sentence production in the three emotions, as compared with their TH counterparts. These findings are consistent with previous findings that CI talkers demonstrate significantly longer sentence duration, compared to their TH counterparts (Chuang, Yang, Chi, Weismer, & Wang, 2012; Leder et al., 1987; Perrin, Berger-Vachon, Topouzkhanian, Truy, & Morgon, 1999; Shin, 2018).

Also, CI talkers produced significant different H/S contrasts in mean F_0 compared with TH talkers, which is consistent with previous studies of English, which show that CI talkers produce smaller contrasts in mean pitch in comparison of "happy" and "sad" speech, as compared to their TH counterparts (Chatterjee et al., 2015; Chatterjee et al., 2016; Wang, et al., 2013).

Surprisingly, contrary to our hypotheses, no significant differences were obtained between the CI talkers and TH talkers for the mean F_0 and F_0 range of the three emotion sentences in the current study. A possible explanation is that in the current study, Mandarin-speaking CI users may have compensated for impaired auditory input by utilizing a relatively long duration of electrical hearing experiences and social interactions with other TH talkers in daily living. Recall that all the CI adults were recruited from a major city (Beijing, capital of China), and that these participants had a long history of device use and a high education level (at least college). In future research, examining more severe symptoms of CI patients' (i.e., with less CI experience and social interaction history) data might be helpful to further test this hypothesis.

In addition, CI talkers produced an increased intensity variability compared to TH talkers. One possible explanation is that using a directional microphone in the current study may have caused problems with apparent changing intensity when talkers moved their heads during the recording section. For this reason, a condenser lapel microphone may be useful in future studies to minimize these potential problems.

Lastly, the results of emotion classification show that both TH and CI talkers produced different pathways of acoustic measures to encode different emotions in their speech. According to the decision-tree graph, TH talkers presented a more interactive complex case than CI talkers in terms of the way the algorithm predicted the three emotions. For example, the algorithm suggested that in TH talkers' emotional prosody production, intensity was most useful to distinguish "angry" from the other emotions, and F_0 and intensity were most predictive to distinguish "sad" in most cases. In contrast, very simple classification patterns were obtained for the productions of CI talkers. For example, the algorithm showed that duration was the most predictive in distinguishing "angry" and "happy" from "sad", and both duration and intensity were the most correct classifiers to predict "sad". These findings of the algorithm indicated that the acoustic measures of intensity and duration are most predictive in classifying the three emotions in the productions of CI users. These findings are consistent with the perceptual results of Experiment 2 using modified acoustic cues, that is, enhancement of secondary cues can benefit CI listeners in perceiving emotional prosody.

CHAPTER 6

EXPERIMENT 4: EMOTIONAL PROSODY PRODUCED BY TH AND CI TALKER GROUPS: PERCEPTUAL ANALYSIS

Much available literature concerning the perceptibility of Mandarin speech produced by talkers with CIs concerns tone production. Xu et al. (2004) conducted a tone-pattern intelligibility test using speech samples recorded from seven TH children and four CI children (age range from 4-9 years old). Four native Mandarin-speaking adults with TH listened to the speech samples and judged the intelligibility of tone production. The results showed that children with CIs received lower score in tone production, as compared to their counterparts. The findings suggested that the children with CI demonstrate tone production with flatter F_0 patterns and with degraded intelligibility. Consistent with these findings, other studies have shown that Mandarin-speaking children with CI have deficits in consistently mastering the tone contrasts in speech production, even if they can produce a normal pitch range (Peng, Tomblin, Cheung, Lin, & Wang, 2004; Zhou, Huang, Chen, & Xu, 2013).

To date, there has been less research on the perceptibility of emotional prosody in Mandarin-speaking adults with CIs. This chapter presents an experiment comparing the production of emotional prosody by TH and CI talkers. In addition to the acoustic approach taken in Experiment 3, a perceptual analysis is conducted to evaluate and interpret potential differences between the two talker groups. TH Mandarin-speaking individuals rated the emotional sentences produced by TH talkers and CI talkers. The purpose was to determine whether the patterns analyzed in the acoustical analysis (Experiment 3) can be perceived by TH listeners.

Participants

Ten Mandarin-speaking TH listeners (Group 4) participated in an emotion discrimination task and a rating task. Of these listeners, 6 were male and 4 were female. Following the procedure described by Nakata el al. (2012) for testing interrater reliability, two additional Mandarin-speaking adult raters were chosen as a second group for rating tasks. The TH listeners were recruited from the campus of the University of Texas at Dallas. All participants were native speakers of Mandarin Chinese with no self-reported history of speech, hearing, emotional or language impairment. The age range of the participants was from 22 to 35 years (mean≈ 26.2 years, SD=3.67).

Procedure

Stimuli: The same three Mandarin sentences described in Experiment 1(see Appendix B) were used as stimuli. A total of 270 stimuli (3 sentences × 3 emotions × 15 participants) were generated. These were the recordings produced by both TH talkers and CI talkers (from Experiment 3).

Data collection: The stimuli were presented at an average level of 65 dB SPL (A-weighting) via a single Altec-Lancing Bluetooth speaker (IMW475), located approximately two feet from the listeners. Presentation of audio stimuli was randomized with six blocks, with 45 sentences randomized within a block. The stimuli were presented on a computer monitor. Although the productions of the TH and CI talkers did not include sentences made with a “neutral” emotion, TH listeners were offered a “neutral” emotion choice in the perception test (i.e., in addition to “angry”, “happy”, and “sad”). A “neutral” choice was included because, based on literature showing that CI talkers may have deficits in F_0 control and tend to produce flatter F_0 patterns (Selleck & Sataloff, 2014; Xu et al., 2004), it was hypothesized that Mandarin-speaking CI talkers would receive a significantly higher proportion of "neutral" judgements than TH talkers. Listeners' accuracy was recorded using DirectRT experiment monitoring software (Jarvis, 2014).

After participants finished the perceptual test, they were asked to listen to all utterances from CI and TH speakers again, and one by one, and rate how closely each sentence matched the target emotion, using a 10-point scale. A 10-point scale was chosen for its greater reliability and precision for the response, as compared with 5- or 7-point formats (Dawes, 2008). All stimuli were randomized with six blocks, with 45 sentences randomized within a block, and with 5 seconds for an inter-block-interval (IBI) between each block. Prior to the beginning of the rating test, listeners were instructed to "listen carefully to each sentence and rate how closely each sentence will match the correctly expressed emotion" and to "please rate the stimuli you just heard and choose from 1 to 10 using the keyboard with number pad." During the rating test, listeners first heard the sentence stimuli, followed by the display of the10-point scale ranging from 1 (extremely poor) to 10 (extremely good) on a computer screen. Rating scores were obtained and recorded from the input of the keyboard number pad using DirectRT experiment monitoring software (Jarvis, 2014).

Method of analysis

Data Analysis: Interrater reliability (IRR) was calculated for scoring test validity by examining data for two TH listeners in a second group who rated 50% of the stimuli. Next, the perceptual accuracy of the TH listeners in identifying the emotional prosody produced by TH and CI talkers was analyzed in a two-way Emotion ("angry", "happy", "sad" and "neutral") by Talker group (TH and CI) repeated measured ANOVA. In addition, rating scores for the two different talker groups were analyzed using the ANOVA method.

Results

Interrater reliability (IRR) results showed an interclass correlation coefficient of 0.707 ($p <$ 0.001), suggesting an acceptable degree of reliability on the perceptual judgments.

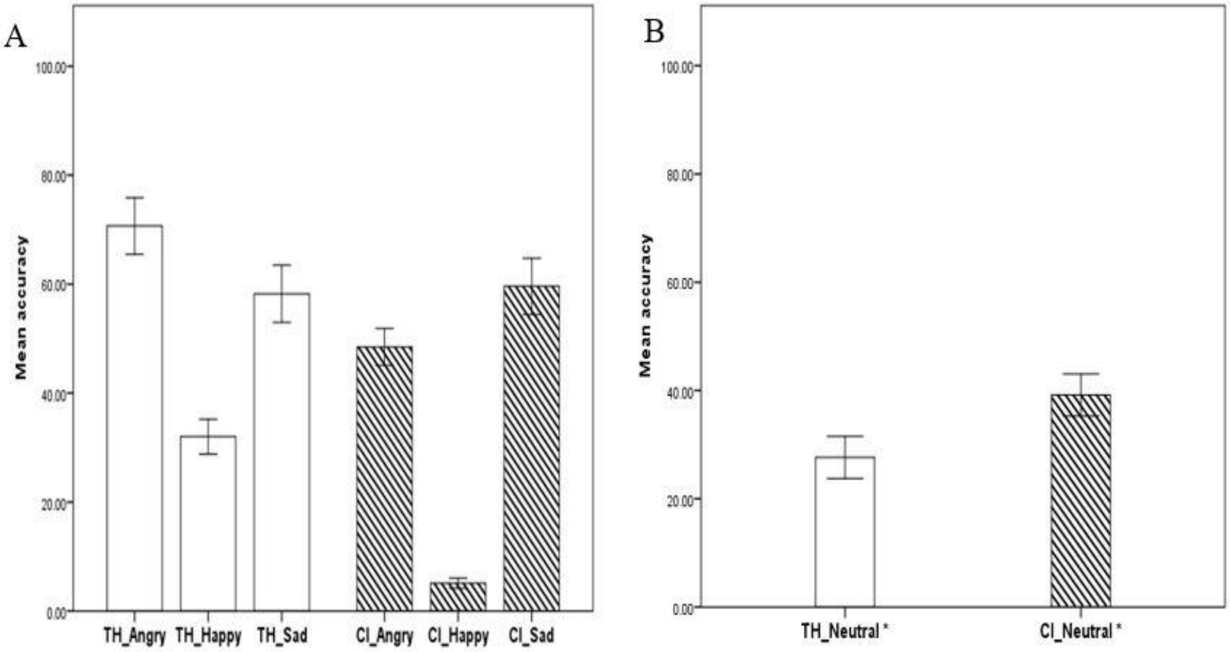


Figure 6.1. TH listeners (*n=10*) accuracy for "angry", "happy", and "sad" sentences produced by talkers with TH (*n=15*) (solid bar) and with CIs (*n=15*) (dashed bar) in panel A. TH listeners (*n=10*) perceptual accuracy for "neutral" emotion in panel B. Note: * indicates judgment only.

The mean accuracy of the judgements of the "angry", "happy" and "sad" sentences by TH listeners are shown in Panel A, with "neutral" emotion judgements in Panel B (Figure 6.1).

Visual inspection of these figures suggests that TH talkers received higher judgements than CI talkers from the TH listeners for “angry”, “happy” and “sad” stimuli. In addition, correct judgements of “happy” sentences were low for both talker groups, particularly for CI talkers’ productions (4%). Although the productions of the TH and CI talkers did not include the “neutral” emotion, TH listeners nevertheless registered a “neutral” choice for many of the sentences, and there was an even higher degree of “neutral” judgements for the CI talkers. The results of repeated measures ANOVA revealed a significant main effect for Talker group [*F* (1, 27) = 103.409, $p < 0.05$, ηp^2= 0.920] and Emotion [*F* (3, 27) = 23.440, $p < 0.05$, ηp^2= 0.723]. In addition, there was a significant two-way (Emotion × Talker group) interaction, indicating that listeners responded differently across the three different emotions for both talker groups [*F* (3, 27) = 28.607, $p < 0.05$, ηp^2= 0.761].

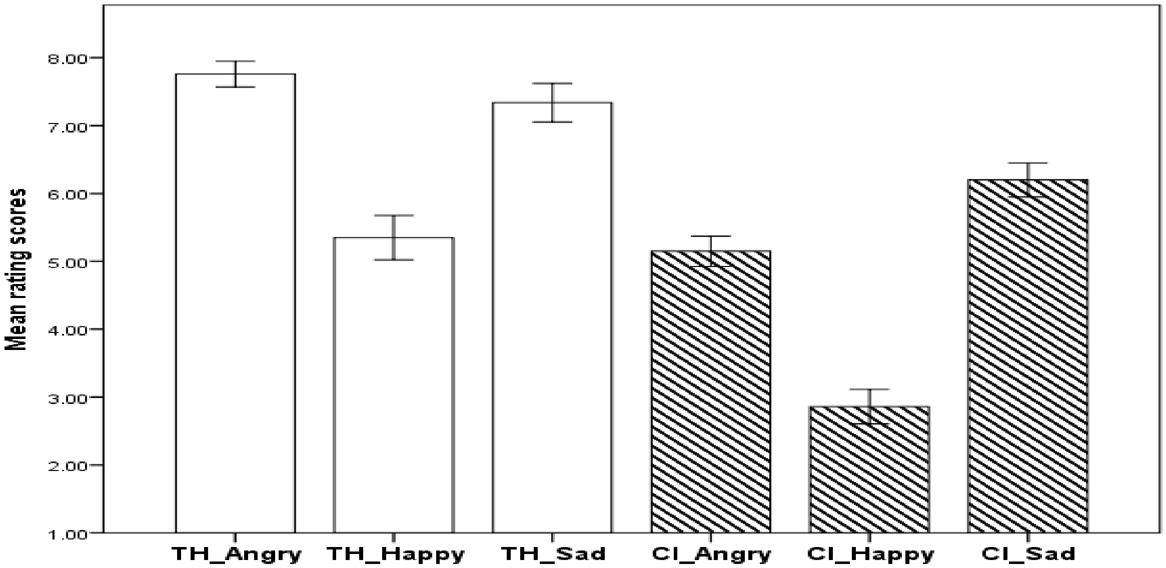


*Figure 6.2.*TH listeners’ (*n=10*) rating scores for emotional prosody production in “angry”, “happy”, and “sad” sentences produced by talkers with TH (*n=15*) (solid bar) and with CIs (*n=15*) (dashed bar). Note: no “neutral” stimuli were included.

The mean scores for the two talker groups rated by TH listeners, collapsed across the different emotions, are presented in Figure 6.2. Visual inspection suggest that patterns of rating scores closely matched the accuracy data presented in Figure 6.1 (panel A). That is, rating scores for TH talkers were higher than those for CI talkers, and "happy" were received low rating scores for both talker groups. The results of repeated measures ANOVA revealed a significant main effect for Talker groups [F (1, 18) =149.497, p < 0.05, ηp^2= 0.943] and Emotion [F (2, 18) = 44.671, p < 0.05, ηp^2= 0.832]. In addition, there was a significant interaction between emotion and talker group, indicating that TH listeners rated the sentences differently across the three emotions for both talker groups [F (2, 18) = 14.578, p < 0.05, ηp^2= 0.618].

In summary, as hypothesized, TH listeners showed decreased accuracy in detecting emotion ("angry", "happy" and "sad") in sentences produced by Mandarin-speaking CI talkers, as compared with TH talkers. In particular, CI talkers demonstrate difficulty in producing the "happy" emotion. In addition, in a four-alternative, forced-choice emotion recognition task including a "neutral" choice, CI talkers received more judgements for the "neutral" emotion than did TH talkers, even though these produced sentences were not intended to express a "neutral" emotion.

Discussion

As expected, TH listeners showed decreased accuracy in detecting emotion ("angry", "happy" and "sad") in sentences produced by Mandarin-speaking CI talkers, as compared with TH talkers. The findings are consistent with previous studies of vocal emotion production in

Japanese (Nakata et al., 2012), and as wells with studies of tone production in Mandarin (Peng et al., 2004; Zhou et al., 2013). The present findings suggest that CI users demonstrate impaired emotional prosody production, and that the impairment can be perceived by TH listeners. In particular, CI users appears to demonstrate difficulty in producing the “happy” emotion, which also is the most difficulty emotion to perceive by these CI users in Experiment 1. Interestingly, in a four-alternative, forced-choice emotion recognition task including a “neutral” choice, CI talkers received more judgements for the “neutral” emotion than did TH talkers, even though these produced sentences were not intended to express a “neutral” emotion. This pattern of results suggest that CI users spoke with impaired $F_{0,}$ resulting in a less perceptible emotional production. This is consistent with Mandarin literature showing CI users demonstrate tone production with flatter F_0 patterns, and also with English literature showing English-speaking CI users failed to control pitch compared to their TH counterpart (Campisi et al., 2005; Evans & Deliyski, 2007; Hamzavi, Deutsche, Baumgartner, Bigenzahn, & Gstoettner, 2000; Hassan et al., 2011; Jones & Munhall, 2002; Ubrig et al., 2011).

The results of the rating data closely matched the accuracy data, further confirming that CI users demonstrate impaired emotional prosody production, and that the impairment can be perceived by TH listeners.

CHAPTER 7

RELATIONSHIP BETWEEN PERCEPTION AND PRODUCTION OF EMOTIONAL PROSODY IN CI USERS

Much available literature has focused on the correlation between tone perception and tone production by Mandarin-speaking CI users (Peng et al., 2004; Xu et al., 2011; Zhou et al., 2013). Peng et al. (2004) examined 30 Mandarin-speaking children with CIs (mean age = 9 years old) performing tone production and tone recognition tasks using 48 words with targeted tones. The results showed a significant but weak correlation ($r = 0.44$, $p < 0.05$) between tone recognition and tone production accuracy. These findings suggest that good tone perception may be an essential condition to obtain good tone production. In line with these findings, Xu et al. (2011) investigated tone perception using a computerized tone contrast test, and tone production using a picture-naming procedure, in a group of 25 Mandarin-speaking children with CIs (mean age = 9.5 years old). The results showed a significant, strong correlation between tone perception and tone production ($r = 0.805$, $p < 0.001$). Consistent with these findings, other studies have also suggested that good tone production in Mandarin-speaking CI users depends on accurate tone perception (Peng et al., 2004; Zhou et al., 2013).

To date, there has been less research on the correlation between emotional prosody perception and production in CI users. As for non-tonal language, Nakata et al. (2012) examined emotional sentence perception of 18 Japanese-speaking children with CIs (age range from 5 to 13 years old) for “happy”, “sad”, and “angry” productions. In addition, these children with CIs were asked to imitate Japanese sentences expressing “surprise” and “disappointment”. The results showed a moderate ($r = 0.56$, $p < 0.05$) correlation between sentence prosody perception and production by children with CIs (although children with CIs were tested in different emotions in the perception and production tasks). These findings suggest that sentence emotional perception plays an important role in obtaining sentence-based emotional production.

Little is known about whether there is a correlation between the perception and production of emotional prosody by tonal language-speaking adults with CIs. In this chapter, we hypothesize a significant relationship between the production and perception of emotional prosody by CI users. That is, poor perception of emotional prosody contributes to imprecise emotional prosody production by Mandarin-speaking adults with CIs. This is studied by exploring the relationship between the accuracy of perception of emotional prosody in Mandarin-speaking adults with CIs and scores rated by TH listeners for CI adults' emotional prosody productions.

Participants

The same CI participants from Experiment 1 took part in this experiment: 15 Mandarin-speaking adults with CIs from the China cohort (group 2).

Procedure

Stimuli: The stimuli used were the same three Mandarin sentences described in Experiment 1(see Appendix B) that convey three emotions ("angry", "happy", and "sad").

Data collection: The accuracy data for emotional prosody perception by CI adults were obtained from Experiment 1, and included "angry", "happy", and "sad" in the natural speech condition. The data for the TH listeners' rating scores for CI adults' emotional prosody productions in "angry", "happy", and "sad" sentences were obtained from Experiment 4.

Method of analysis

Data Analysis: General linear regression models using SPSS 22.0 (IBM corporation, 2013) to analyze the data of emotional prosody perception by CI listeners and the scores rated by TH listeners for CI adults' emotional prosody productions. The regression model was fitted as follows: First, the collinearity among the three independent variables (perceptual accuracy of "angry", "happy" and "sad") was analyzed before applying the regression model, to confirm that there were no multicollinearity issues. Second, residuals were checked after running initial the initial regression to show that the residuals were uncorrelated and, normally distributed with homoscedasticity. Third, the results were checked to show that there were no outliers.

We then generalized a simple regression model to analyze the relationship between the mean production scores of CI talkers (rated by the TH listeners) and the mean perception accuracy of CI listeners'. Next, we created multiple regression models to predict the dependent variable (the TH rating scores of the CI talkers' productions) from the aforementioned three predictor variables. Finally, the data were further analyzed, using a backward stepwise regression, to determine the contribution of each predictor. The backward elimination regression plotted all three predictors in the first model, to calculate the contribution of each emotion against a removal criterion. If a predictor (emotion) did not make a significant contribution to the dependent variable (where the removal criterion was the probability of F-to-remove >=

0.100), then it was removed from the model, and the model was reassessed for the contribution of the remaining independent variables.

Results

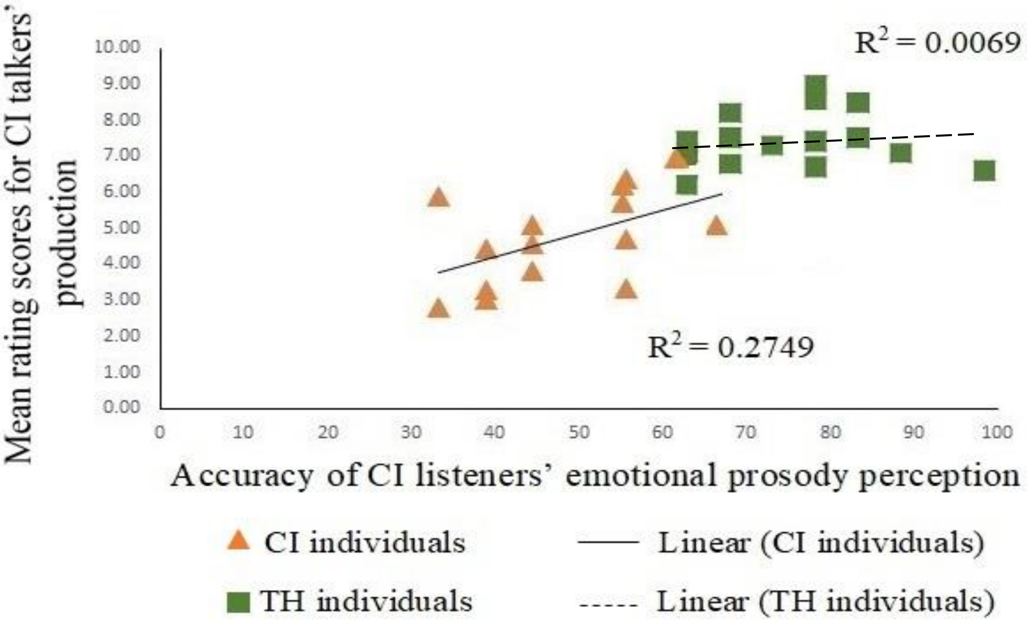
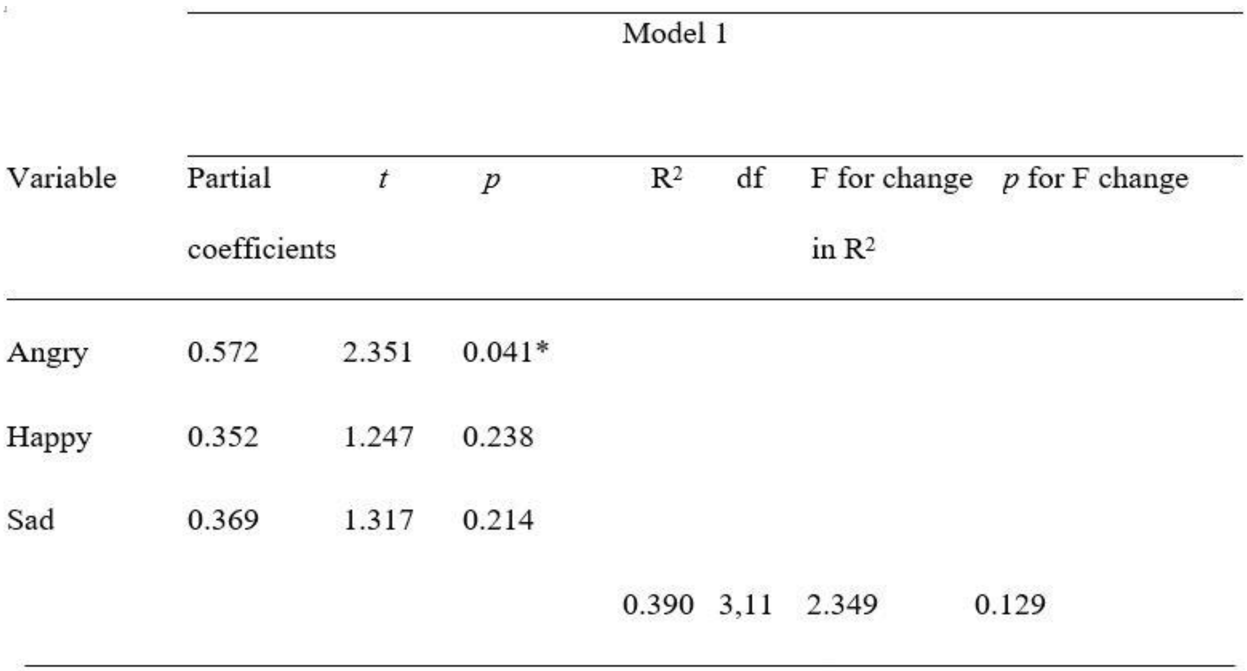


Figure 7.1. Scatter plot of accuracy percentage and rating score of production for CI individuals *(n=15)* (triangle orange) with regression fit for the strongest correlation. Scatter plot for TH individuals *(n=15)* (green square) is added for comparison.

Figure 7.1. displays a scatterplot depicting the production scores for CI talkers rated by the TH listeners (y-axis) and the perception accuracy of CI listeners' in the emotion identification task (x-axis). Each triangle represents one CI individual and each square represents one TH individual. The production scores are averages across 90 judgements (=10 TH listeners × 3 emotions × 3 sentences). The perception accuracy values are means across 18 judgements (=3 emotions × 3 sentences × 2 professional actor talkers).

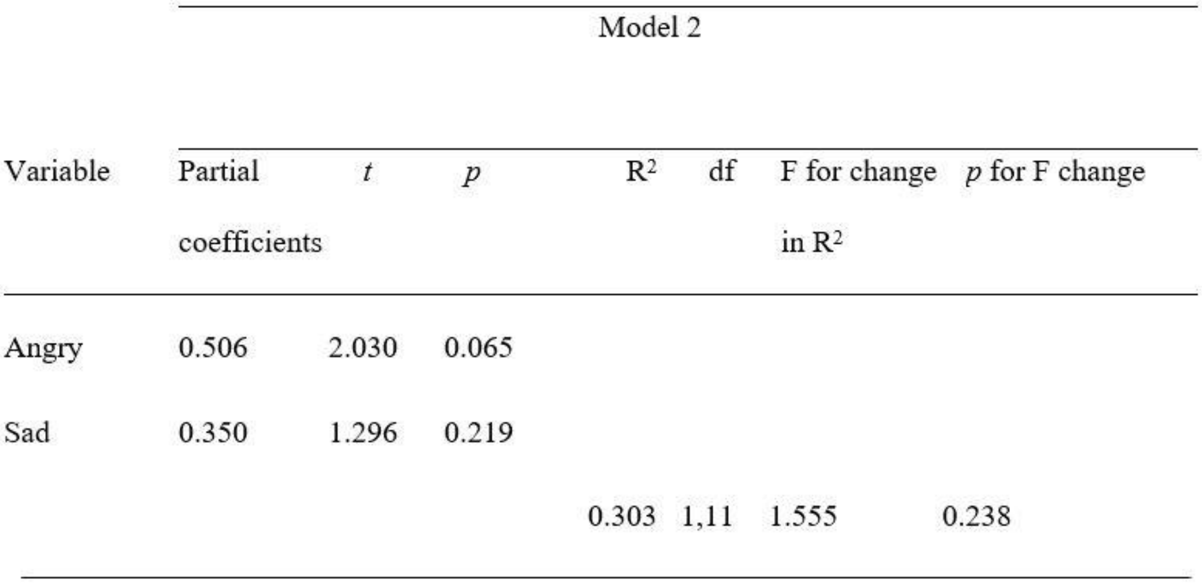
Visual inspection shows that lower perception accuracy in CI individuals is associated with lower scores received for their emotional prosody productions. Results of a simple linear regression showed a significant positive correlation (r = 0.524, p < 0.05) between the accuracy of CI users' perceptions and the TH listeners' rating scores of the CI users' productions. Scores for the TH individuals are added in Figure 7.1 for comparison. For the TH individuals, there is not significant relation between perception and production (r = 0.083, p = 0.769, ns), although they achieved higher rating scores and higher perceptual accuracy, compared to the CI individuals.

Table 7.2. *Summary of multiple regression models in which the accuracy of "angry", "happy", and "sad" emotions are related to rating scores for CI talkers' productions.*



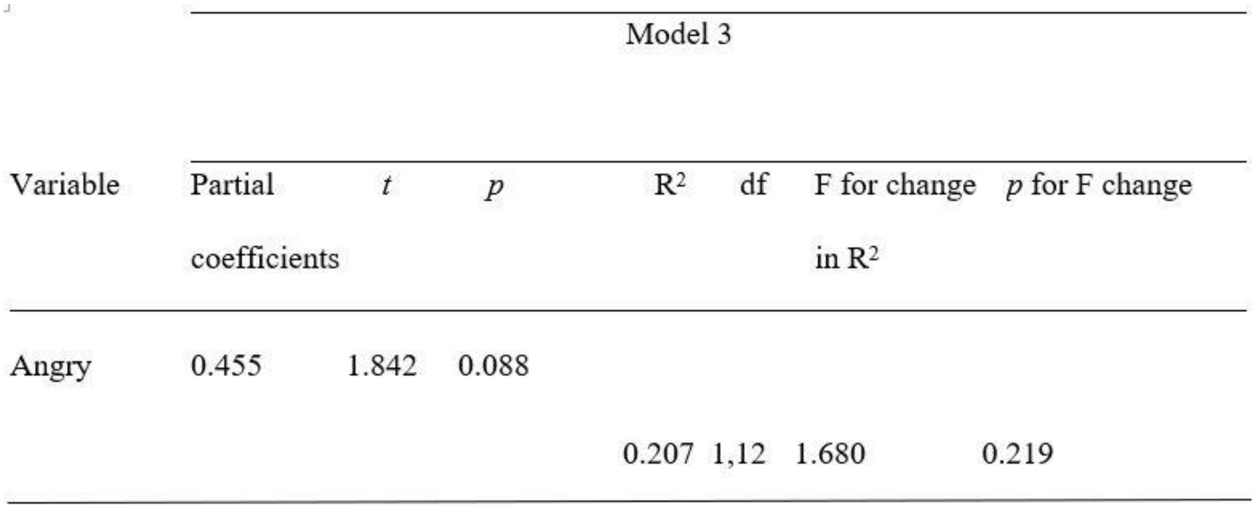
	Model 1						
Variable	Partial coefficients	t	p	R^2	df	F for change in R^2	p for F change
Angry	0.572	2.351	0.041*				
Happy	0.352	1.247	0.238				
Sad	0.369	1.317	0.214				
				0.390	3,11	2.349	0.129

Table 7.2. (cont.) *Summary of multiple regression models in which the accuracy of "angry", "happy", and "sad" emotions are related to rating scores for CI talkers' productions.*



Model 2

Variable	Partial coefficients	*t*	*p*	R^2	df	F for change in R^2	*p* for F change
Angry	0.506	2.030	0.065				
Sad	0.350	1.296	0.219				
				0.303	1,11	1.555	0.238



Model 3

Variable	Partial coefficients	*t*	*p*	R^2	df	F for change in R^2	*p* for F change
Angry	0.455	1.842	0.088				
				0.207	1,12	1.680	0.219

Note: * indicates $p < 0.05$.

Table 7.2. displays the results of the multiple regression model associated with the scores of CI talkers' productions (rated by the TH listeners) as predicted by the three variables (perceptual accuracy of "angry", "happy", and "sad" emotions, respectively). The tests for collinearity suggested a low level of multicollinearity (*variance inflation factor (VIF)* = 1.076 for accuracy of “angry”, 1.062 for accuracy of “happy”, and 1.014 for accuracy of “sad”). Using the backward stepwise regression method, the three predictor variables were all entered to Model 1, the predictor of “happy” emotion was removed in Model 2, and the predictor of “sad” emotion was removed in Model 3.

Three patterns of results are observed in Table 7.2. First, the results of the regression analysis in Model 1 provided partial confirmation that the scores for CI talkers’ productions could be predicted. The partial coefficients for the three predictors were accuracy of “angry”, $r = 0.572$, $t = 2.351$, $p < 0.05$; accuracy of “happy”, r = 0.352, $t = 1.247$, $p = 0.238$, ns; and accuracy for “sad”, r = 0.369, $t = 1.317$, $p = 0.214$, ns. These findings suggest that the rating scores for CI talkers’ production is significantly predicted by accuracy of “angry” but not predicted by the accuracy of either “sad” or “happy”. Second, after removal of the “happy” variable in Model 2, results show that, compared to Model 1, the R-squared value has decreased (from 0.390 to 0.304), suggesting that the removal of the “happy” variable decreased the ability of the model to explain the variation in CI talkers’ production scores (8.4%). Third, after further removal of the “sad” variable in Model 3, the R-squared values showed a larger decreased (from 0.390 in Model 1 to →0.207 in Model 3) and lacked significance. The ability of Model 3 to explain the variance in CI talkers’’ production scores had therefore further reduced to 20.7%. These findings suggest that removal of the “happy” and “sad” variables did not significantly improve prediction.

In summary, these findings suggest that the perception of emotional prosody is significantly correlated ($r = 0.524$, $p < 0.05$) with the production of emotional prosody by CI recipients. In addition, the best fitting model for predicting the production of emotional prosody by CI recipients is a linear combination of the three emotions, with the accuracy of the "angry" variable (followed by the "sad" and "happy" variables) playing a major role in explaining to the production scores of CI talkers as rated by the TH listeners.

Discussion

In the current study, we observed a significant and moderate correlation (r = 0.524, $p <$ 0.05) between the emotional prosody perception ability of Mandarin-speaking CI listeners and TH listeners' rating scores of the productions made by these same CI individuals. These findings are broadly consistent with previous studies on tone production and tone perception for Mandarin-speaking CI users (Peng et al., 2004; Xu et al., 2011; Zhou et al., 2013) and on sentence prosody perception and production by Japanese-speaking children with CIs (Nakata et al., 2012).

The results show that the low perceptual accuracy of CI users is associated with the low ratings received for their emotional prosody productions. These findings suggest that poor perception of emotional prosody contributes to their imprecise speech prosody production.
A first potential explanation is that CI users demonstrate a reduced ability to form correct speech internal models. Current speech production models propose at least two important stages in realizing speech output: 1) the formation of the internal model, and 2) the maintenance of that model through ongoing feedback. According to Perkell et al. (2000), the speech internal model uses both sensory and auditory feedback to map desired sounds into speech by shaping the vocal tract. As the natural central auditory system matures, this internal

model increases its accuracy in reproducing more precise sequences of speech sounds (Perkell et al., 2000). Long-term uses of CIs therefore result in the eventual improvement of prosodic quality, suggesting that CIs can strengthen the internal representation of speech.

A second potential explanation of why poor perception relates to poor production is that CI users have an impaired auditory system, which results in problems with monitoring auditory feedback and difficulty detecting prosodic cues in emotional prosody perception. Auditory feedback uses sensory information acquired during ongoing speech to modulate pitch, and plays an essential role in maintaining normal speech production (Jones & Munhall, 2002; Jones & Keough, 2008; Selleck & Sataloff, 2014; Tourville, Reilly, & Guenther, 2008). Problems monitoring auditory feedback can affect speech production for individuals with HI (Lane & Webster, 1991; Monsen, Engebretson, & Vemula, 1979; Tobey, Geers, Brenner, Altuna, & Gabbert, 2003; Ubrig et al., 2011). Therefore, those individuals with CI may produce a monotone or may use excessive variations in pitch, such as abrupt changes in pitch and pitch breaks while talking (Hamzavi et al., 2000; Hocevar-Boltezar et al., 2006; ller Kirk & Hill-Brown, 1985; Tobey, et al., 2003; Ubrig et al., 2011; Xu & Zhou, 2011).

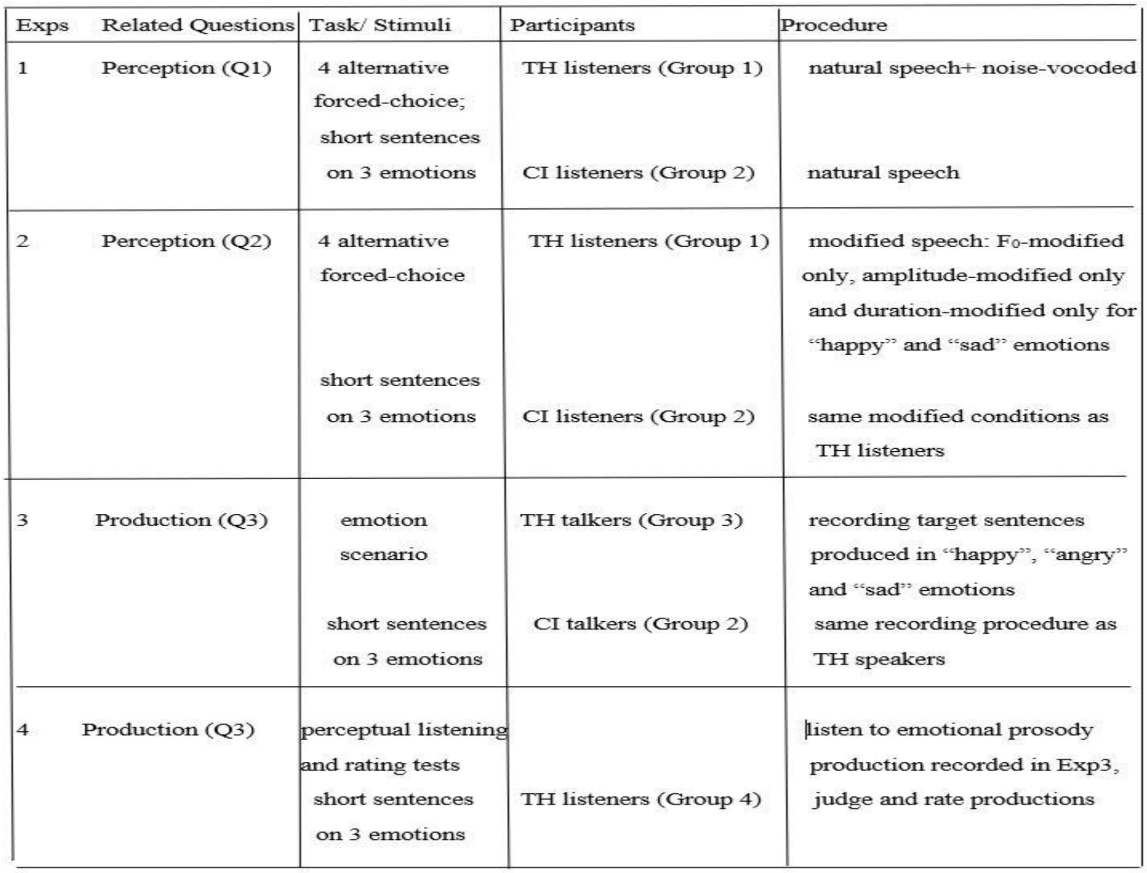
The results of a stepwise regression model suggest that the perceptual accuracy of the “angry” sentences (followed by the “sad” and “happy” sentences) plays a major role in contributing to the TH listeners’ rating scores of CI talkers’ productions. These findings are consistent with the patterns of emotion identification in CI users’ in Experiment 1. Recall that Experiment 1 showed that the “angry” sentences were the most highly identified, followed by the “sad” sentences, while the “happy” sentences were the least identifiable. In all, these findings further support the hypothesis of poor perception of emotional prosody in Mandarin-speaking individuals contributing to their imprecise speech prosody production.

CHAPTER 8

GENERAL DISCUSSION

The overall goal of this dissertation is to explore the perception and production of emotional prosody by Mandarin-speaking TH and CI listeners and to compare these data with findings for non-tonal languages. We further examine the differences between Mandarin-speaking TH and CI listeners in perceiving and producing acoustic cues for emotional prosody. Lastly, we investigate the relationship between the production and perception of emotional prosody by these tonal language-speaking CI users. Four experiments were conducted to achieve these goals. A review of the four experiments is provided in Table 8.1.

Table 8.1. *A review of the four experiments in this dissertation.*



Exps	Related Questions	Task/ Stimuli	Participants	Procedure
1	Perception (Q1)	4 alternative forced-choice;	TH listeners (Group 1)	natural speech+ noise-vocoded
		short sentences on 3 emotions	CI listeners (Group 2)	natural speech
2	Perception (Q2)	4 alternative forced-choice	TH listeners (Group 1)	modified speech: F_0-modified only, amplitude-modified only and duration-modified only for “happy” and “sad” emotions
		short sentences on 3 emotions	CI listeners (Group 2)	same modified conditions as TH listeners
3	Production (Q3)	emotion scenario	TH talkers (Group 3)	recording target sentences produced in “happy”, “angry” and “sad” emotions
		short sentences on 3 emotions	CI talkers (Group 2)	same recording procedure as TH speakers
4	Production (Q3)	perceptual listening and rating tests		listen to emotional prosody production recorded in Exp3,
		short sentences on 3 emotions	TH listeners (Group 4)	judge and rate productions

All four experiments converged on showing the difference in perception and production of emotion prosody between Mandarin-speaking TH and CI adults. The findings suggest that the CI adults demonstrate deficits in both emotion prosody perception and production, compared to their TH counterparts. Detailed information of four experiments are as follows.

Experiment 1 explored the differences in emotional prosody perception between Mandarin-speaking listeners with TH and with CIs. The results showed that Mandarin-speaking CI users demonstrated deficits in perceiving emotional prosody, compared to their TH counterparts. The TH listeners performed better in the natural speech than in the noise-vocoded speech, and the noise-vocoded speech produced lower intelligibility when spectral resolution was reduced. In addition, the results showed that the performance of Mandarin-speaking CI listeners in natural speech is similar to that of TH adults at a lower channel setting (i.e., at 4-channel), in contrast to 8-channel shown for CI listeners in previous comparable studies of non-tonal languages (e.g., English). These findings are consistent with the FV hypothesis (Zhu, 2013), which claims that Mandarin (a tonal language) uses pitch for purposes of linguistic tone and therefore has less prosodic space in which to signal the pitch of paralinguistic information (e.g., emotional states). However, since the data rely critically on participants with CIs, we cannot rule out that the implants themselves caused problems in processing emotional prosody, and more research is warranted. In terms of comparing TH and CI listeners perceiving different emotions in natural speech, CI listeners showed more difficulty identifying the “sad” and “happy” emotions compared to TH listeners, although CI listeners’ accuracy order of “sad” and “happy” was similar to that of the TH listeners (“sad” > “happy”). In addition, CI listeners showed a high degree of confusing “happy” with “angry” and confusing “sad” with “neutral”, whereas TH listeners did not show these patterns.

Experiment 2 explored whether enhancement of secondary cues (duration and amplitude) can benefit CI listeners to perceive emotions. The results suggest that enhancement of duration cues provide minimal gain in perceiving the “sad” emotion for CI listeners, while enhancement of amplitude and F_0 provide more appreciably increased accuracy for the “happy” emotion. These findings are consistent with previous tone perception studies suggesting that due to spectro-temporal cues (especially pitch cues) are limited in current CI speech-processing strategies, and therefore, secondary cues (duration and amplitude) might play a more important role than expected for CI users (Fu & Zeng, 2000; Luo & Fu, 2004; Wei et al., 2004). These findings suggest that individual cues for different emotions vary for CI listeners. In addition, these findings have implications for emotional prosody conversion speech synthesis programs designed for Mandarin which presently focus on F_0 patterns (i.e., enhanced duration cues for “sad” and enhanced amplitude cues for “happy” for Mandarin-speaking CI listeners) (Jiang et al., 2005; Tao et al., 2006; Wen et al., 2011).

Experiment 3 investigated whether TH talkers and CI talkers differ in the production of acoustic cues for emotions expressed at the sentence level. Analysis of five acoustic properties (mean F_0, F_0 range, mean intensity, intensity range, and duration) supported this hypothesis. Across the three emotions, the values of the acoustic measures differed for mean intensity, which was lower for CI users, and intensity range, which was higher for the CI users. When the “happy” and “sad” emotions were considered alone (following previous studies conducted by Chatterjee et al. in 2015 and 2016), CI users were found to show smaller mean F_0 contrasts than their TH counterparts. Altogether, these patterns for Mandarin appear similar to those of non-tonal languages.

The data were further examined using a machine classification (decision tree) algorithm to determine which acoustic measures are most predictive in the three aforementioned emotions produced by talkers with TH and CIs, and to ascertain whether CI talkers use the same acoustics measures as those employed by TH talkers to predict emotions. According to the decision-tree graph, TH talkers presented a more interactive complex case than CI talkers in terms of the way the algorithm predicted the three emotions. The decision-tree for the TH talkers reflects a major role of F_0, especially in distinguishing the “happy” and “sad” emotions, whereas for the CI talkers F_0 played no role. These findings are consistent with previous acoustic analyses showed that, for instance, CI talkers demonstrated deficits in mean F_0 for producing “happy” and “sad” contrasts. In addition, the findings of the algorithm indicated that the acoustic measures of amplitude and duration are most predictive in classifying the three emotions in the productions of CI users. These findings are consistent with the perceptual results of Experiment 2 using modified acoustic cues, that is, enhancement of secondary cues can benefit CI listeners to perceive emotional prosody.

Experiment 4 analyzed the production data in a perceptual manner to determine whether CI talkers demonstrate deficits in producing emotional prosody production compared to TH talkers, and whether these deficits can be perceived by TH listeners. Results showed that the TH listeners showed decreased accuracy in detecting emotions (“angry”, “happy” and “sad”) in sentences produced by CI talkers. Listeners’ rating scores closely reflected the patterns in the talkers’ accuracy scores. These findings confirm that CI users demonstrate impaired emotional prosody production, and that such impairment can be perceived by TH listeners. In addition, the CI talkers received more judgments for the "neutral" emotion than did TH talkers, even though the produced sentences were not intended to express a "neutral" emotion. This pattern of results suggest that CI talkers produce speech with impaired $F_{0,}$ thus generating emotions that are less perceptible (and therefore monotone or "neutral") productions.

Lastly, the relationship between the production and perception of emotional prosody by CI users was explored. The accuracy data for emotional prosody perception by CI adults were obtained from Experiment 1, and included "angry", "happy", and "sad" in the natural speech condition. The data for the TH listeners' rating scores for CI adults' emotional prosody productions in "angry", "happy", and "sad" sentences were obtained from Experiment 4. The results of simple regression indicate that the lower perceptual accuracy of CI individuals is significantly correlated with lower rating scores received for their emotional prosody production. These outcomes are consistent with previous studies on tone production and tone perception for Mandarin-speaking CI users (Peng et al., 2004; Xu et al., 2011; Zhou et al., 2013) and on sentence prosody perception and production by Japanese-speaking children with CIs (Nakata et al., 2012). These findings also suggest that poor perception of emotional prosody contributes to imprecise speech prosody production through a reduced ability to form correct speech internal models and by problems in monitoring auditory feedback with difficulty detecting prosodic cues in the speech.

In summary, the findings of the current study show similarities and differences with previous studies of CI users who speak non-tonal languages (e.g., English and Japanese) (Chatterjee et al., 2015; Chatterjee et al., 2016; Gilbers et al., 2015; House, 1994; Luo, Fu, &

Galvin III, 2007; Nakata et al., 2012; Pereira, 2000; Pereira, 2000b; Zhu et al., 2016). At least three similarities can be noted: First, the findings of the current perception experiments show that Mandarin-speaking CI users demonstrate deficits in perceiving emotional prosody. In addition, Mandarin-speaking CI users showed higher identification of the “sad” emotion than the “happy” emotion, and often confuse “happy” with “angry” and “sad” with “neutral”. Second, the findings of the current production experiments suggest that Mandarin-speaking CI users produced the same patterns of acoustic cues for the basic emotions (“angry”, “happy”, and “sad”). That is, “angry” is produced with high F_0 value, large amplitude variation, and short duration; “happy” is produced with relatively high F_0 value, relatively large amplitude variation, and relatively short duration; while “sad” is produced with low F_0 values, small amplitude variation, and long duration. In addition, Mandarin-speaking CI users produced universal patterns of increased sentence duration values in their emotional prosody production, and smaller contrasts in mean F_0 in comparison of “happy” and “sad” productions, as compared to their TH counterparts. This same pattern was noted in a comparison of CI and TH talkers in non-tonal languages (Chatterjee et al., 2015; Chatterjee et al., 2016). Third, the findings of the current study suggest that poor perception of emotional prosody contributes to CI users’ imprecise emotional prosody productions, which is consistent with studies conducted by Nataka et al. (2012) for Japanese.

In contrast, differences of the current study from previous literature are as follows: First, the perceptual performance of Mandarin-speaking CI listeners with natural speech is similar to that of Mandarin-speaking TH adults at a lower channel setting (i.e., at 4-channel), compared to 8-channel reported in comparable non-tonal language studies (e.g., English/Japanese) (Chatterjee et al., 2015; Chatterjee et al., 2015; Friesen et al., 2001; Zhu, et al., 2016). These findings suggest that Mandarin-speaking CI users demonstrate lower accuracy than non-tonal language-speaking CI users with respect to hearing emotional prosody. This pattern is consistent with the FV hypothesis (Zhu, 2013), although, as noted earlier, more data are needed. Second, the findings of a perceptual experiment using modified acoustic cues suggest that individual cues for different emotions vary for Mandarin-speaking CI users (e.g., slightly improved “sad” accuracy with modified duration, and improved “happy” accuracy with modified amplitude). These findings have implications for emotional prosody conversion speech synthesis programs designed for Mandarin that presently focus solely on F_0 patterns (Jiang et al., 2005; Tao et al., 2006; Wen et al., 2011).

In general, the data obtained in this dissertation can improve our understanding of the nature of emotional prosody perception and production by Mandarin-speaking individuals with TH and with hearing impairments. If replicated in future experiments, these findings may also play a role in designing rehabilitation benefits and sound processing strategies for individuals with hearing impairments.

Limitations and Future Research

Although the present findings suggest new insights into the processing of emotional prosody in CI users who speak tonal languages, a number of caveats should be considered. In both the perceptual and production experiments, several experimental issues may have limited the generalizability of results. The use of a relatively small stimulus set (three declarative sentences spoken by two actors in four emotions) could have posed problems in a number of ways. Although declarative sentences are a logical syntactic structure to start with (as they can be easily spoken on different emotions), these have different sentence-level intonation patterns than questions in Mandarin (Yuan, Shih, & Kochanski, 2002; Yuan, 2006) and could therefore show different results. Thus, for TH listeners, a neutral tone declarative sentence might be perceived as "happy" due to the final rising sentence-level intonation (Tao, et al., 2006).

Increasing the number of sentences used (including sentences with different patterns of lexical tones) may be especially important for this type of research. For instance, Gu et al. (2018) examined the perceptual accuracy of Mandarin-speaking children with CIs identifying declarative and interrogative utterances with different sentence-final tones, as compared with that of their TH peers. The results show that children with CIs demonstrated significantly lower accuracy than their TH peers, were in identifying declarative and interrogative utterances with different sentence-final tones. The tone patterns in the current study were: 3-4-2-1-1-3; 3-4-2-1-4-1; and 2-3-1-4-2-1. That is, two of the sentences ended with a high and flat tone (1), while one ended with a fall-rising tone (3). Conducting future experiments with added tonal variability would be valuable.

In addition, in order to more ideally explore the perception capabilities of Mandarin-speaking individuals with CIs, it would be useful to have more spoken samples to work with. Although examining the four "basic" emotions ("angry", "happy", "sad", and "neutral") provides a good basis for exploring emotional prosody perception, ongoing studies of prosody are now beginning to characterize patterns for other emotions, including "fear", "disappointment", and "surprise". Future studies might profitably explore these patterns.

Other potential caveats concern participants and procedures. Experiment 2 separately modified the TH productions of “sad” and “happy” sentences in terms of F_0, amplitude and duration. The magnitude of these modifications (F_0 was increased approximately 75 % and duration increased by 25% for “happy”; F_0 was decreased 20% and duration decreased by 10% for “sad”) was based on studies of English (Assmann & Nearey, 2008) and included subjective concerns for naturalness in synthesized speech quality. The small perceptual effects observed in the present studies may have resulted, in part, as a response to these small-magnitude manipulations. Revisiting this experiment with larger modifications might result in greater observed effects.

Recently, researchers have questioned whether the use of noise-excited, channel-vocoded stimulators with TH listeners provide reliable simulation of CI listeners. Current work by Winn and Litovsky (Winn & Litovsky, 2015) suggests that if the source of vocoded speech changes, the perception of both indexical and nonindexical information can alter drastically. More work using synthetic speech in this area will be needed.

Furthermore, because only one tonal language (i.e., Mandarin) was investigated in this dissertation, it is difficult to determine the general of the FV hypothesis (Zhu, 2013). Future studies can further test the FV hypothesis to investigate the extent to which prosodic processing limitations recruit capacities such as auditory memory. Studies might also examine whether an interaction occurs between the extent of emotional prosody processing and the complexity of the tonal systems involved (e.g., three-level register tonal languages, such as Thai, versus seven-level contour tonal languages, such as Cantonese).

Some directions for future extensions of the production experiment are as follows: First, there are different ways to elicit emotions for production analyses, other than the currently-used method. For example, Hubbard (2016) asked talkers to speak each sentence three times in order to increase expressive repetitions (e.g., “happy”, “happier”, and “happiest”). That is, each talker could produce a greater range of productions for each type of emotion. This method would provide an additional benefit of increasing the overall amount of production data available for analysis. Second, other classification approaches in the field of machine learning, such as artificial neutral networks, could be useful in examining both the production and perception data.

Clinical Implications

This dissertation suggests that Mandarin-speaking individuals with CIs demonstrate deficits in the perception and production of emotional prosody as compared with their TH counterparts. These data contribute to a growing body of studies describing how emotional prosody may be particularly vulnerable in individuals with a hearing impairment, particularly those with CIs. In addition, the information gained from this research can help improve the spoken communication skills and benefit rehabilitative efforts for Mandarin-speaking CI recipients. Fully understanding the communicative intent and emotional state of speakers is critical in order to contribute to positive spoken communication (Chatterjee et al., 2015). CI users who experience inaccurate perception of emotional expressions may receive insufficient and incorrect feedback from their peers, leading to low self-esteem, isolation, and rejection affecting their life quality (Schorr, 2005; Schorr, et al., 2009). Few studies have investigated the rehabilitation of emotion perception in adult CI users, and most of those studies have focused on using emotion cues of facial expressions, rather than on vocal expressions (Dyck & Denver, 2003; Jiam et al., 2017). Although facial expressions may play an important role in navigating difficult listening conditions, the vocal emotional state of speakers provides critical information in social communication when facial expressions cannot be detected (e.g., listening to the radio) (Chatterjee et al., 2015; Luo et al., 2007).

The data from Experiment 2 using separate conditions of F_0-, amplitude-, and duration-modified speech suggest that CI listeners showed benefit from different types of acoustic information for different emotions. One clinical implication is the potential use of different cue enhancement strategies for particular emotions designed to be heard by CI users. For instance, since the present data found a small but positive improvement in “sad” sentence perception with increased duration, and appreciable gain in accuracy for “happy” sentence recognition with increases in F_0 and amplitude, it may make sense to create training stimuli that increase the duration of the “sad” emotion, while enhancing amplitude and F_0 for the “happy” emotion. This focus may shorten the amount of time needed for intense auditory rehabilitative training, which is currently noted to be remarkably long (Zhang et al., 2012).

Approaches to the training of emotional prosody for individuals with CIs that rely on phonetic contrasts of the vowels and consonants, did not report improvement in emotion identification (Zhang, Dorman, Fu, & Spahr, 2012). In contrast, other approaches using repeated emotional sentences or music training had reported some success in training emotional prosody to adults with CIs (Krull, Luo, & Iler Kirk, 2012; Petersen, Mortensen, Hansen, & Vuust, 2012) and children with CIs (Good, et al., 2017). The current findings of improved emotional prosody recognition with modified acoustic cues in synthesized speech may add a new direction for this type of speech training.

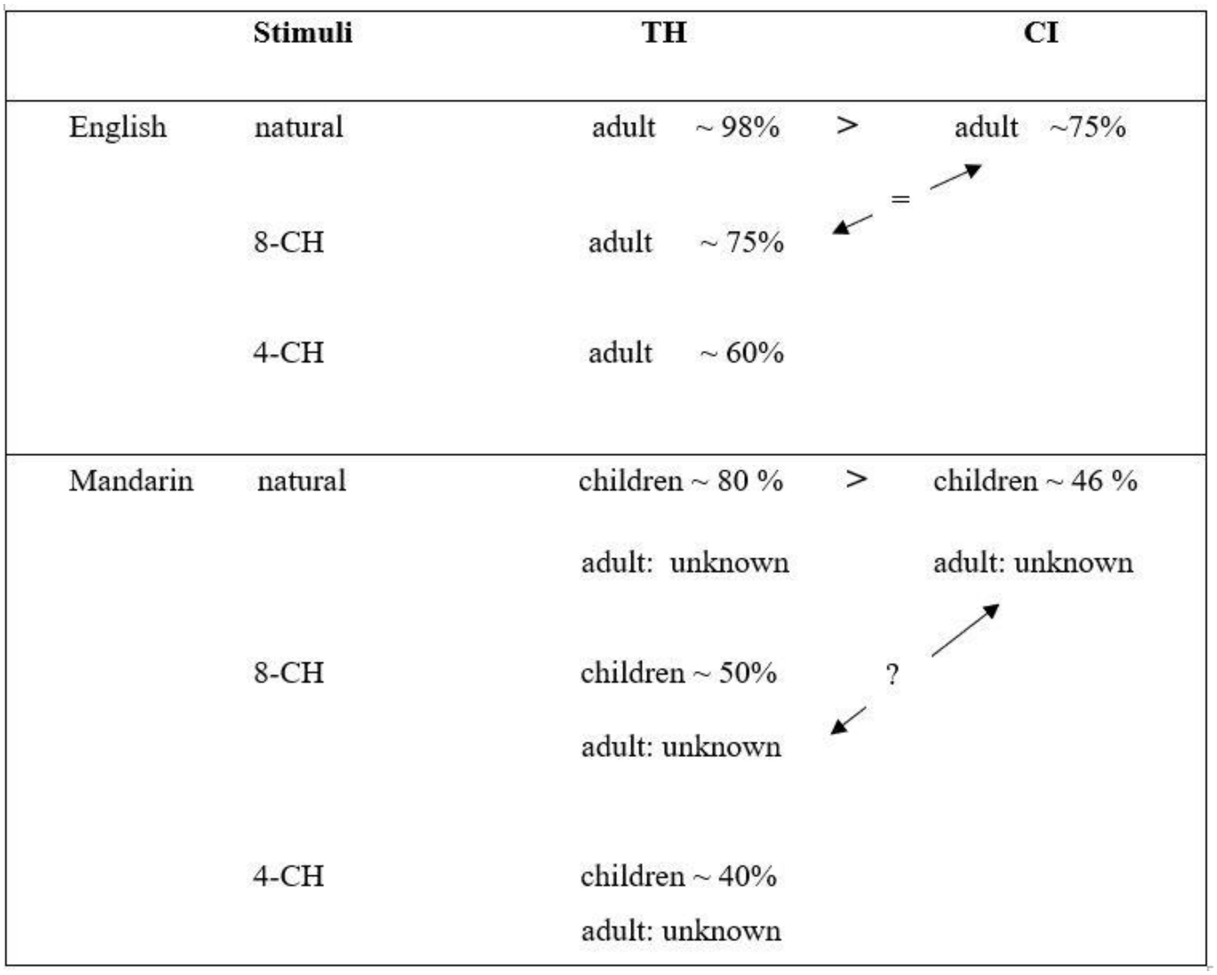
Lastly, Mandarin-speaking CI users demonstrated a slower rate of emotional sentence production compared to their TH counterparts, which is in line with previous studies showing slowed speech in general for talkers with CIs (Chuang et al., 2012; Leder et al., 1987; Perrin et al., 1999; Shin, 2018). Many reasons have been offered for why the speech of CI users are notably slow (e.g., Shin, 2018). The present findings suggest that an additional factor may be that CI users may tend to slow their speech as a strategy to express emotion more clearly, due to their poor perception of emotional prosody and correlated problems in monitoring auditory feedback.

In summary, Mandarin-speaking individuals with CIs show deficits in emotional prosody perception and production. Speech science experimentation can offer new means of understanding these individuals' limitations and this can lead to improvement in the communication skills and quality of life for individuals with HI and CIs.

APPENDIX A

SUMMARY OF VOCAL EMOTION PERCEPTION STUDIES CONDUCTED BY CHATTERJEE ET AL. (2015, 2017)

Table A.1. *Overall emotional prosody recognition in five emotions (“angry”, “happy”, “sad”, “neutral”, and “scared”) across children and adults with TH and with CIs in English and Mandarin, summarizing based on two studies conducted by Chatterjee and her colleagues (2015, 2017). It is noted that these two studies used different testing materials and populations.*



	Stimuli	TH		CI
English	natural	adult ~ 98%	>	adult ~75%
	8-CH	adult ~ 75%	↙ = ↗	
	4-CH	adult ~ 60%		
Mandarin	natural	children ~ 80 % adult: unknown	>	children ~ 46 % adult: unknown
	8-CH	children ~ 50% adult: unknown	↙ ? ↗	
	4-CH	children ~ 40% adult: unknown		

APPENDIX B

LIST OF STIMULI FOR ALL EXPERIMENTS

Sentence 1

I have to say: ah.
Wo$_3$Bu$_4$You$_2$De$_1$Shuo$_1$: Ah$_{3.}$
我不由得说： 啊.

Sentence 2

You come and fix it.
Ni$_3$Guo$_4$Lai$_2$Xiu$_1$Yi$_4$Xiu$_{1.}$
你过来修一修.

Sentence 3

Go do your own stuff.
Mang$_2$Ni$_3$De$_1$Shi$_4$ Qing$_2$Ba$_{1.}$
忙你的事情吧.

REFERENCES

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology, 37*(5), 715.

Arlinger, S. (2003). Negative consequences of uncorrected hearing loss-a review. *International Journal of Audiology, 42*, 2S17-2S20.

Assmann, P. F., & Nearey, T. M. (2008). Identification of frequency-shifted vowels. *The Journal of the Acoustical Society of America, 124*(5), 3203-3212.

Bachorowski, J., & Owren, M. J. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science, 6*(4), 219-224.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614.

Bänziger, T., & Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech Communication, 46*(3), 252-267.

Beshah, T., & Hill, S. (2010). Mining road traffic accident data to improve safety: Role of road-related factors on accident severity in ethiopia. Paper presented at the *AAAI Spring Symposium: Artificial Intelligence for Development,*

Bhargava, N., Sharma, G., Bhargava, R., & Mathuria, M. (2013). Decision tree analysis on j48 algorithm for data mining. *Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering, 3*(6).

Boersma, P., & Weenink, D. (2016). *Praat: Doing Phonetics by Computer.[Computer Program].Version 6.0.19,* retrieved 13 June 2016 from http://www.praat.org/.

Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., & Jäncke, L. (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Cognitive Brain Research, 9*(3), 227-238.

Busso, C., Parthasarathy, S., Burmania, A., AbdelWahab, M., Sadoughi, N., & Provost, E. M. (2017). MSP-IMPROV: An acted corpus of dyadic interactions to study emotion perception. *IEEE Transactions on Affective Computing, 8*(1), 67-80.

Campisi, P., Low, A., Papsin, B., Mount, R., Cohen-Kerem, R., & Harrison, R. (2005). Acoustic analysis of the voice in pediatric cochlear implant recipients: A longitudinal study. *The Laryngoscope, 115*(6), 1046-1050.

Chang, J. E., Bai, J. Y., & Zeng, F. (2006). Unintelligible low-frequency sound enhances simulated cochlear-implant speech recognition in noise. *IEEE Transactions on Biomedical Engineering, 53*(12), 2598-2601.

Chatterjee, M., Christensen, J. A., Kulkarni, A. M., Deroche, M. L., Damm, S. A., Bosen, A. K., Limb, C. J. (2016, February). Voice emotion communication by listeners with cochlear implants. *Poster Presented at: Association for Research in Otolaryngology 39th Annual Midwinter Meeting;San Diego, California.*

Chatterjee, M., Kulkarni, A. M., Christensen, J. A., Deroche, M. L., & Limb, C. J. (2015). Voice emotion recognition and production by individuals with normal hearing and with cochlear implants. *The Journal of the Acoustical Society of America, 137*(4), 2205-2205.

Chatterjee, M., Zion, D. J., Deroche, M. L., Burianek, B. A., Limb, C. J., Goren, A. P., Christensen, J. A. (2015). Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hearing Research, 322*, 151-162.

Chen, F., Wong, L. L., & Hu, Y. (2014). Effects of lexical tone contour on mandarin sentence intelligibility. *Journal of Speech, Language, and Hearing Research, 57*(1), 338-345.

Chen, G. T. (1974). The pitch range of English and Chinese speakers. *Journal of Chinese Linguistics*, 159-171.

Chen, Y., Wong, L. L., Chen, F., & Xi, X. (2014). Tone and sentence perception in young mandarin-speaking children with cochlear implants. *International Journal of Pediatric Otorhinolaryngology, 78*(11), 1923-1930.

Chinese Corpus Consortium, (2006). Retrieved from http://www.d-ear.com/CCC/corpora.htm

Chuang, H., Yang, C., Chi, L., Weismer, G., & Wang, Y. (2012). Speech intelligibility, speaking rate, and vowel formant characteristics in mandarin-speaking children with cochlear implant. *International Journal of Speech-Language Pathology, 14*(2), 119-129.

Chun, D. M. (2002). *Discourse intonation in L2: From theory and research to practice* John Benjamins Publishing.

Cowie, R., & Cornelius, R. R. (2003). Describing the emotional states that are expressed in speech. *Speech Communication, 40*(1), 5-32.

Dawes, J. (2008). Do data characteristics change according to the number of scale points used. *International Journal of Market Research, 50*(1), 61-77.

Dellaert, F., Polzin, T., & Waibel, A. (1996). Recognizing emotion in speech. Paper presented at the *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference On, 3,* 1970-1973.

Duanmu, S. (2007). *The phonology of standard chinese* Oxford University Press.

Epstein, M. A. (2002). Voice quality and prosody in english. *Doctoral Dissertation, University of California, Los Angeles.*

Evans, M. K., & Deliyski, D. D. (2007). Acoustic voice analysis of prelingually deaf adults before and after cochlear implantation. *Journal of Voice: Official Journal of the Voice Foundation, 21*(6), 669-682. doi:S0892-1997(06)00089-0 [pii]

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861-874.

Fishman, K. E., Shannon, R. V., & Slattery, W. H. (1997). Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. *Journal of Speech, Language, and Hearing Research, 40*(5), 1201-1215.

Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America, 110*(2), 1150-1163.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech, 1*(2), 126-152.

Fu, Q., & Zeng, F. (2000). Identification of temporal envelope cues in chinese tone recognition. *Asia Pacific Journal of Speech, Language and Hearing, 5*(1), 45-57.

Fu, Q., & Zeng, F. (2013). Identification of temporal envelope cues in chinese tone recognition. *Asia Pacific Journal of Speech, Language and Hearing,*

Gandour, J., Wong, D., Dzemidzic, M., Lowe, M., Tong, Y., & Li, X. (2003). A cross-linguistic fMRI study of perception of intonation and emotion in chinese. *Human Brain Mapping, 18*(3), 149-157.

Garnham, C., O'driscoll, M., Ramsden, R., & Saeed, S. (2002). Speech understanding in noise with a med-el COMBI 40 cochlear implant using reduced channel sets. *Ear and Hearing, 23*(6), 540-552.

Gay, T. (1978). Physiological and acoustic correlates of perceived stress. *Language and Speech, 21*(4), 347-353.

Gilbers, S., Fuller, C., Gilbers, D., Broersma, M., Goudbeek, M., Free, R., & Başkent, D. (2015). Normal-hearing listeners' and cochlear implant users' perception of pitch cues in emotional speech. *I-Perception, 6*(5), 0301006615599139.

Good, A., Gordon, K. A., Papsin, B. C., Nespoli, G., Hopyan, T., Peretz, I., & Russo, F. A. (2017). Benefits of music training for perception of emotional speech prosody in deaf children with cochlear implants. *Ear and Hearing, 38*(4), 455.

Gray, J. A. (1994). Three fundamental emotion systems. *The Nature of Emotion: Fundamental Questions, 14,* 243-247.

Greasley, P., Sherrard, C., & Waterman, M. (2000). Emotion in language and speech: Methodological issues in naturalistic approaches. *Language and Speech, 43*(4), 355-375.

Gu, WT., & Zhu, Y. (2018, March). *Perceptual Identification of Declarative and Interrogative Utterances in Mandarin-Speaking Cochlear Implanted Children.* Poster session presented at the CI2018 Emerging Issues in Cochlear Implantation, Washington, DC.

Guimaraes, A. J., Pizzini, C. V., De Abreu Almeida, M., Peralta, J. M., Nosanchuk, J. D., & Zancope-Oliveira, R. M. (2010). Evaluation of an enzyme-linked immunosorbent assay using purified, deglycosylated histoplasmin for different clinical manifestations of histoplasmosis. *Microbiology Research, 1*(1), 10.4081/mr.2010.e2. Epub 2010 Mar 17. doi:e2 [pii]

Hamzavi, J., Deutsche, W., Baumgartner, W. D., Bigenzahn, W., & Gstoettner, W. (2000). Short-term effect of auditory feedback on fundamental frequency after cochlear implantation. *Audiology, 39*(2), 102-105.

Han, D., Liu, B., Zhou, N., Chen, X., Kong, Y., Liu, H., Xu, L. (2009). Lexical tone perception with HiResolution and HiResolution 120 sound-processing strategies in pediatric mandarin-speaking cochlear implant users. *Ear and Hearing, 30*(2), 169-177. doi:10.1097/AUD.0b013e31819342cf [doi]

Hassan, S. M., Malki, K. H., Mesallam, T. A., Farahat, M., Bukhari, M., & Murry, T. (2011). The effect of cochlear implantation and post-operative rehabilitation on acoustic voice analysis in post-lingual hearing impaired adults. *European Archives of Oto-Rhino-Laryngology, 268*(10), 1437-1442.

He, A., Deroche, M. L., Doong, J., Jiradejvong, P., & Limb, C. J. (2016). Mandarin tone identification in cochlear implant users using exaggerated pitch contours. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology, 37*(4), 324-331. doi:10.1097/MAO.0000000000000980 [doi]

Henry, B. A., & Turner, C. W. (2003). The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners. *The Journal of the Acoustical Society of America, 113*(5), 2861-2873.

Hocevar-Boltezar, I., Radsel, Z., Vatovec, J., Geczy, B., Cernelc, S., Gros, A., Zargi, M. (2006). Change of phonation control after cochlear implantation. *Otology & Neurotology : Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology, 27*(4), 499-503. doi:10.1097/01.mao.0000224083.70225.b7 [doi]

House, D. (1994). Perception and production of mood in speech by cochlear implant users. Paper presented at the *Icslp.*

Howie, J. M. (1976). *Acoustical studies of mandarin vowels and tones* Cambridge University Press.

Hsiao, F. (2008). Mandarin melody recognition by pediatric cochlear implant recipients. *Journal of Music Therapy, 45*(4), 390-404.

Huang, M., Hung, Y., & Chen, W. (2010). Neural network classifier with entropy based feature selection on breast cancer diagnosis. *Journal of Medical Systems, 34*(5), 865-873.

Hubbard, D. J., Faso, D. J., Assmann, P. F., & Sasson, N. J. (2017). Production and perception of emotional prosody by adults with autism spectrum disorder. *Autism Research,*

Hubbard, D. J. (2016). *Production and Perception of Affective Prosody by Adults with Autism Spectrum Disorder.*

Ryant, N., Yuan, J., & Liberman, M. (2014, May). Mandarin tone classification without pitch tracking. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on* (pp. 4868-4872). IEEE.

IBM corporation. (2013). IBM SPSS statistics for windows, version 22.0.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations.

Jarvis, B. (2014). DirectRT (version 2014.1. 127). *New York: Empirisoft Corporation.*

Jiam, N., Caldwell, M., Deroche, M., Chatterjee, M., & Limb, C. (2017). Voice emotion perception and production in cochlear implant users. *Hearing Research,*

Jiang, D., Zhang, W., Shen, L., & Cai, L. (2005). Prosody analysis and modeling for emotional speech synthesis. Paper presented at the *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference On, 1* I/281-I/284 Vol. 1.

Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. *Handbook of Emotions, 2,* 220-235.

Jones, J. A., & Keough, D. (2008). Auditory-motor mapping for pitch control in singers and nonsingers. *Experimental Brain Research, 190*(3), 279-287.

Jones, J. A., & Munhall, K. G. (2002). The role of auditory feedback during phonation: Studies of mandarin tone production. *Journal of Phonetics, 30*(3), 303-320.

Jongman, A., Wang, Y., Moore, C. B., & Sereno, J. A. (2006). *Perception and production of mandarin chinese tones* na.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770.

Juslin, P. N., & Scherer, K. R. (2005). Vocal expression of affect. *The New Handbook of Methods in Nonverbal Behavior Research,* 65-135.

Kawahara, H., Masuda-Katsuse, I., & De Cheveigne, A. (1999). Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication, 27*(3), 187-207.

Keppel, G. (1991). Design and analysis: A researcher handbook prentice hall. *Englewood Cliffs, New Jersey,*

Kong, Y., Lee, T., Yuan, M., & Yu, W. (2012). Relative contributions of temporal and spectral cues for mandarin and cantonese tone recognition. *The Journal of the Acoustical Society of America, 131*(4), 3478-3478.

Kong, Y., Winn, M. B., Poellmann, K., & Donaldson, G. S. (2016). Discriminability and perceptual saliency of temporal and spectral cues for final fricative consonant voicing in simulated cochlear-implant and bimodal hearing. *Trends in Hearing, 20,* 2331216516652145.

Krull, V., Luo, X., & Iler Kirk, K. (2012). Talker-identification training using simulations of binaurally combined electric and acoustic hearing: Generalization to speech and emotion recognition. *The Journal of the Acoustical Society of America, 131*(4), 3069-3078.

Lambrecht, K., & Polinsky, M. (1997). Typological variation in sentence-focus constructions. *Cls, 33,* 189-206.

Lane, H., & Webster, J. W. (1991). Speech deterioration in postlingually deafened adults. *The Journal of the Acoustical Society of America, 89*(2), 859-866.

Leder, S. B., Spitzer, J. B., Kirchner, J. C., Flevaris-Phillips, C., Milner, P., & Richardson, F. (1987). Speaking rate of adventitiously deaf male cochlear implant candidates. *The Journal of the Acoustical Society of America, 82*(3), 843-846.

Lehiste, I. (1970). Suprasegmentals (cambridge, MA.). *LehisteSuprasegmentals1970,*

Lehiste, I. (1979). Perception of sentence and paragraph boundaries. *Frontiers of Speech Communication Research,* 191-201.

Lehiste, I., & Wang, W. S. (1976). *Perception of sentence boundaries with and without semantic information* na.

Li, A., Fang, Q., & Dang, J. (2011). Emotional intonation in a tone language: Experimental evidence from chinese. *ICPhS XVII, Hong Kong, 1*, 1198-1201.

Liang, Z. (1963). Tonal discrimination of mandarin chinese. *Acta Physiologica Sinica, 26*, 85-91.

Liu, C., Azimi, B., Tahmina, Q., & Hu, Y. (2012). Effects of low harmonics on tone identification in natural and vocoded speech. *The Journal of the Acoustical Society of America, 132*(5), EL378-EL384.

Liu, F., Jiang, C., Thompson, W. F., Xu, Y., Yang, Y., & Stewart, L. (2012). The mechanism of speech processing in congenital amusia: Evidence from mandarin speakers. *PLos One, 7*(2), e30374.

Liu, F. (1924). Szu sheng shih yen lu [experimental studies of tone]. *Ch'un Yi Shanghai.*

Liu, P., & Pell, M. D. (2012). Recognizing vocal emotions in mandarin chinese: A validated database of chinese vocal emotional stimuli. *Behavior Research Methods, 44*(4), 1042-1051.

Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and speech, 47*(2), 109-138.

Liu, T., Hsu, C., & Horng, M. (2000). Tone detection in mandarin-speaking hearing-impaired subjects. *Audiology, 39*(2), 106-109.

Lu, H., Peng, SC., Deroche, M. L., Chatterjee, M., Lin, YS. (2017, June). *Voice emotion recognition by Mandarin-speaking pediatric cochlear implant users.* Poster session presented at the Conference on Implantable Auditory Prostheses, Lake Tahoe, CA.

ller Kirk, K., & Hill-Brown, C. (1985). Speech and language results in children with a cochlear implant. *Ear and Hearing, 6*(3), 36S-47S.

Loizou, P. C. (2006). Speech processing in vocoder-centric cochlear implants. *Advances in Oto-Rhino-Laryngology, 64*, 109-143. doi:94648 [pii]

Luo, X., & Fu, Q. (2004). Enhancing chinese tone recognition by manipulating amplitude envelope: Implications for cochlear implants. *The Journal of the Acoustical Society of America, 116*(6), 3659-3667.

Luo, X., Fu, Q., & Galvin III, J. J. (2007). Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends in Amplification*, *11*(4), 301-315. doi:11/4/301 [pii]

Massaro, D. W., Cohen, M. M., & Tseng, C. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in mandarin chinese. *Journal of Chinese Linguistics*, 267-289.

Meister, H., Landwehr, M., Pyschny, V., Walger, M., & Wedel, H. v. (2009). The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. *International Journal of Audiology, 48*(1), 38-48.

Monsen, R. B., Engebretson, A. M., & Vemula, N. R. (1979). Some effects of deafness on the generation of voice. *The Journal of the Acoustical Society of America, 66*(6), 1680-1690.

Moore, B. C. (2003). Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants. *Otology & Neurotology, 24*(2), 243-254.

Moore, B. C., & Carlyon, R. P. (2005). Perception of pitch by people with cochlear hearing loss and by cochlear implant users. *Pitch* (pp. 234-277) Springer.

Moore, B. C., & Moore, G. A. (2003). Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects. *Hearing Research, 182*(1-2), 153-163.

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of mandarin chinese tones. *The Journal of the Acoustical Society of America, 102*(3), 1864-1877.

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.

Mozziconacci, S. J., & Hermes, D. J. (1999). Role of intonation patterns in conveying emotion in speech. *ICPhS 1999,* 2001-2004.

Nakata, T., Trehub, S. E., & Kanda, Y. (2012). Effect of cochlear implants on children’s perception and production of speech prosody. *The Journal of the Acoustical Society of America, 131*(2), 1307-1314.

Nie, K., Barco, A., & Zeng, F. G. (2006). Spectral and temporal cues in cochlear implant speech perception. *Ear and Hearing, 27*(2), 208-217. doi:10.1097/01.aud.0000202312.31837.25 [doi]

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review, 97*(3), 315.

Pak, C. L., & Katz, W. F. (2017). Recognition of emotional prosody in mandarin: Evidence from a synthetic speech paradigm. *The Journal of the Acoustical Society of America, 141*(5), 3701-3701.

Pell, M. D. (2000). Intonation and emotion. *The Journal of the Acoustical Society of America, 108*(5), 2533-2533.

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior, 33*(2), 107-120.

Peng, S., Lu, H., Lu, N., Lin, Y., Deroche, M. L., & Chatterjee, M. (2017). Processing of acoustic cues in lexical-tone identification by pediatric cochlear-implant recipients. *Journal of Speech, Language, and Hearing Research, 60*(5), 1223-1235.

Peng, S., Tomblin, J. B., Cheung, H., Lin, Y., & Wang, L. (2004). Perception and production of mandarin tones in prelingually deaf children with cochlear implants. *Ear and Hearing, 25*(3), 251-264.

Pereira, C. (2000a). Dimensions of emotional meaning in speech. Paper presented at the *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion.*

Pereira, C. (2000b). The perception of vocal affect by cochlear implantees. *Cochlear Implants,* 343-345.

Pereira, C. M. (2000). *Perception and expression of emotion in speech* Macquarie University Sydney.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., Zandipour, M. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics, 28*(3), 233-272.

Perrin, E., Berger-Vachon, C., Topouzkhanian, A., Truy, E., & Morgon, A. (1999). Evaluation of cochlear implanted children's voices. *International Journal of Pediatric Otorhinolaryngology, 47*(2), 181-186.

Peters, K. P. (2006). Emotion perception in speech: Discrimination, identification, and the effects of talker and sentence variability.

Petersen, B., Mortensen, M. V., Hansen, M., & Vuust, P. (2012). Singing in the key of life: A study on effects of musical ear training after cochlear implantation. *Psychomusicology: Music, Mind, and Brain, 22*(2), 134.

Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation.*

Plutchik, R. (1984). Emotions: A general psychoevolutionary theory. *Approaches to Emotion, 1984*, 197-219.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology, 17*(03), 715-734.

Raithel, V., & Hielscher-Fastabend, M. (2004). Emotional and linguistic perception of prosody. reception of prosody. *Folia Phoniatrica Et Logopaedica: Official Organ of the International Association of Logopedics and Phoniatrics (IALP), 56*(1), 7-13. doi:10.1159/000075324 [doi]

Rajesh, R., Maiti, J., & Reena, M. (2018). Decision tree for manual material handling tasks using WEKA. *Ergonomic design of products and worksystems-21st century perspectives of asia* (pp. 13-24) Springer.

Rodero, E. (2011). Intonation and emotion: Influence of pitch levels and contour type on creating emotions. *Journal of Voice, 25*(1), e25-e34.

Ross, E. D., Edmondson, J. A., & Seibert, G. B. (1986). The effect of affect on various acoustic measures of prosody in tone and non-tone languages-a comparison based on computer-analysis of voice. *Journal of Phonetics, 14*(2), 283-302.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161.

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality, 11*(3), 273-294.

Ryant, N., Yuan, J., & Liberman, M. (2014). Mandarin tone classification without pitch tracking. Paper presented at the *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference On,* 4868-4872.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*(1), 227-256.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32*(1), 76-92.

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion, 1*(4), 331-346.

Schorr, E. (2005). Social and emotional functioning of children with cochlear implants.

Schorr, E. A., Roth, F. P., & Fox, N. A. (2009). Quality of life for children with cochlear implants: Perceived benefits and problems and the perception of single words and emotional sounds. *Journal of Speech, Language, and Hearing Research, 52*(1), 141-152.

Schröder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M., & Gielen, S. C. (2001). (2001). Acoustic correlates of emotion dimensions in view of speech synthesis. Paper presented at the *Interspeech,* 87-90.

Schuller, B., Rigoll, G., & Lang, M. (2004). Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture. Paper presented at the *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference On, 1,* I-577.

Selby, R. W., & Porter, A. A. (1988). Learning from examples: Generation and evaluation of decision trees for software resource analysis. *IEEE Transactions on Software Engineering, 14*(12), 1743-1757.

Selleck, M. A., & Sataloff, R. T. (2014). The impact of the auditory system on phonation: A review. *Journal of Voice: Official Journal of the Voice Foundation, 28*(6), 688-693. doi:10.1016/j.jvoice.2014.03.018 [doi]

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.

Shin, S. (2018). The effect of auditory input on the rate of speech production by cochlear implant users. *(Doctoral Dissertation).*

Sidorov, M., Brester, C., Ultes, S., & Schmitt, A. (2017). Salient cross-lingual acoustic and prosodic features for english and german emotion recognition. *Dialogues with social robots* (pp. 159-169) Springer.

Sjolander, K., & Beskow, J. (2010). WaveSurfer (1.8.8). *Retrieved from Http://Www.Speech.Kth.Se/Wavesurfer/ on December 30, 2011.*

Sobin, C., & Alpert, M. (1999). Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. *Journal of Psycholinguistic Research, 28*(4), 347-365.

Soto, J. A., & Levenson, R. W. (2009). Emotion recognition across cultures: The influence of ethnicity on empathic accuracy and physiological linkage. *Emotion, 9*(6), 874.

Stein, N. L., & Oatley, K. (1992). Basic emotions: Theory and measurement. *Cognition & Emotion, 6*(3-4), 161-168.

Su, Q., Galvin, J. J., Zhang, G., Li, Y., & Fu, Q. J. (2016). Effects of within-talker variability on speech intelligibility in mandarin-speaking adult and pediatric cochlear implant patients. *Trends in Hearing, 20,* 10.1177/2331216516654022. doi:10.1177/2331216516654022 [doi]

Svirsky, M. A., Lane, H., Perkell, J. S., & Wozniak, J. (1992). Effects of short-term auditory deprivation on speech production in adult cochlear implant users. *The Journal of the Acoustical Society of America, 92*(3), 1284-1300.

Tan, J., Dowell, R., & Vogel, A. (2016). Mandarin lexical tone acquisition in cochlear implant users with prelingual deafness: A review. *American Journal of Audiology, 25*(3), 246-256.

Tao, J., Kang, Y., & Li, A. (2006). Prosody conversion from neutral speech to emotional speech. *IEEE Transactions on Audio, Speech, and Language Processing, 14*(4), 1145-1154.

Tao, D., Deng, R., Jiang, Y., Galvin, J. J.,3rd, Fu, Q. J., & Chen, B. (2015). Melodic pitch perception and lexical tone perception in mandarin-speaking cochlear implant users. *Ear and Hearing, 36*(1), 102-110. doi:10.1097/AUD.0000000000000086 [doi]

Thompson, W. F. (2014). *Music in the social and behavioral sciences: An encyclopedia* SAGE Publications.

Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences of the United States of America, 109*(46), 19027-19032. doi:10.1073/pnas.1210344109 [doi]

Titze, I. R. (2000). *Principles of voice production* National Center for Voice and Speech.

Tobey, E. A., Angelette, S., Murchison, C., Nicosia, J., Sprague, S., Staller, S. J., Beiter, A. L. (1991). Speech production performance in children with multichannel cochlear implants. *Otology & Neurotology, 12*, 165-173.

Tobey, E. A., Geers, A. E., Brenner, C., Altuna, D., & Gabbert, G. (2003). Factors associated with development of speech production skills in children implanted by age five. *Ear and Hearing, 24*(1 Suppl), 36S-45S. doi:10.1097/01.AUD.0000051688.48224.A6 [doi]

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage, 39*(3), 1429-1443.

Tye-Murray, N., Spencer, L., Bedia, E. G., & Woodworth, G. (1996). Differences in children's sound production when speaking with a cochlear implant turned on and turned off. *Journal of Speech, Language, and Hearing Research, 39*(3), 604-610.

Ubrig, M. T., Goffi-Gomez, M. V., Weber, R., Menezes, M. H., Nemr, N. K., Tsuji, D. H., & Tsuji, R. K. (2011). Voice analysis of postlingually deaf adults pre- and postcochlear implantation. *Journal of Voice : Official Journal of the Voice Foundation, 25*(6), 692-699. doi:10.1016/j.jvoice.2010.07.001 [doi]

Uldall, E. (1960). Attitudinal meanings conveyed by intonation contours. *Language and Speech, 3*(4), 223-234.

Uskul, A. K., Paulmann, S., & Weick, M. (2016). Social power and recognition of emotional prosody: High power is associated with lower recognition accuracy than low power. *Emotion, 16*(1), 11.

Van Zyl, M. (2014). *Perception of Prosody by Cochlear Implant Recipients,*

Wang, D. J., Trehub, S. E., Volkova, A., & van Lieshout, P. (2013). Child implant users' imitation of happy-and sad-sounding speech. *Frontiers in Psychology, 4*, 351.

Wang, S., Liu, B., Dong, R., Zhou, Y., Li, J., Qi, B., Zhang, L. (2012). Music and lexical tone perception in chinese adult cochlear implant users. *The Laryngoscope, 122*(6), 1353-1360.

Wang, S., Liu, B., Zhang, H., Dong, R., Mannell, R., Newall, P., Han, D. (2012). Mandarin lexical tone recognition in sensorineural hearing-impaired listeners and cochlear implant users. *Acta Oto-Laryngologica, 133*(1), 47-54.

Wang, S., Xu, L., & Mannell, R. (2011). Relative contributions of temporal envelope and fine structure cues to lexical tone recognition in hearing-impaired listeners. *Journal of the Association for Research in Otolaryngology, 12*(6), 783-794.

Wang, W., Zhou, N., & Xu, L. (2011). Musical pitch and lexical tone perception with cochlear implants. *International Journal of Audiology, 50*(4), 270-278.

Wei, C., Cao, K., & Zeng, F. (2004). Mandarin tone recognition in cochlear-implant subjects. *Hearing Research, 197*(1), 87-95.

Wen, M., Wang, M., Hirose, K., & Minematsu, N. (2011). Prosody conversion for emotional mandarin speech synthesis using the tone nucleus model. Paper presented at the *Twelfth Annual Conference of the International Speech Communication Association,*

Whalen, D. H., & Xu, Y. (1992). Information for mandarin tones in the amplitude contour and in brief segments. *Phonetica, 49*(1), 25-47.

Whalley, K., & Hansen, J. (2006). The role of prosodic sensitivity in children's reading development. *Journal of Research in Reading, 29*(3), 288-303.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *The Journal of the Acoustical Society of America, 52*(4B), 1238-1250.

Winn, M. B., & Litovsky, R. Y. (2015). Using speech sounds to test functional spectral resolution in listeners with cochlear implants. *The Journal of the Acoustical Society of America, 137*(3), 1430-1442.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: Practical machine learning tools and techniques* Morgan Kaufmann.

Wong, P. (2012). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic mandarin lexical tone productions. *Journal of Phonetics, 40*(1), 141-151.

Xu, L., Chen, X., Lu, H., Zhou, N., Wang, S., Liu, Q., Han, D. (2011). Tone perception and production in pediatric cochlear implants users. *Acta Oto-Laryngologica, 131*(4), 395-398.

Xu, L., Li, Y., Hao, J., Chen, X., Xue, S. A., & Han, D. (2004). Tone production in mandarin-speaking children with cochlear implants: A preliminary study. *Acta Oto-Laryngologica, 124*(4), 363-367.

Xu, L., & Pfingst, B. E. (2003). Relative importance of temporal envelope and fine structure in lexical-tone perception (L). *The Journal of the Acoustical Society of America, 114*(6), 3024-3027.

Xu, L., & Zhou, N. (2011). Tonal languages and cochlear implants. *Auditory prostheses* (pp. 341-364) Springer.

Xu, Y. Chapter in Routledge Handbook of Phonetics, Katz, W and Assmann, P. (under contract).

Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Deng, Z., Lee, S., Busso, C. (2004). An acoustic study of emotions expressed in speech. Paper presented at the *Interspeech,*

You, M., Chen, C., & Bu, J. (2005). CHAD: A chinese affective database. Paper presented at the *International Conference on Affective Computing and Intelligent Interaction,* 542-549.

Yu, F., Chang, E., Xu, Y., & Shum, H. (2001). Emotion detection from speech to enrich multimedia content. Paper presented at the *Pacific-Rim Conference on Multimedia,* 550-557.

Yuan, J. (2006). Mechanisms of question intonation in mandarin. *Chinese spoken language processing* (pp. 19-30) Springer.

Yuan, J., Shen, L., & Chen, F. (2002). The acoustic realization of anger, fear, joy and sadness in chinese. Paper presented at the *Interspeech.*

Yuan, J., Shih, C., & Kochanski, G. P. (2002). Comparison of declarative and interrogative intonation in chinese. Paper presented at the *Speech Prosody 2002, International Conference,*

Zhang, S., Ching, P., & Kong, F. (2006). Acoustic analysis of emotional speech in mandarin chinese. Paper presented at the *International Symposium on Chinese Spoken Language Processing,* 57-66.

Zhang, S. (2008). Emotion recognition in chinese natural speech by combining prosody and voice quality features. Paper presented at the *International Symposium on Neural Networks,* 457-464.

Zhang, T., Dorman, M. F., Fu, Q. J., & Spahr, A. J. (2012). Auditory training in patients with unilateral cochlear implant and contralateral acoustic stimulation. *Ear and Hearing, 33*(6), e70-9. doi:10.1097/AUD.0b013e318259e5dd [doi]

Zhou, N., Huang, J., Chen, X., & Xu, L. (2013). Relationship between tone perception and production in prelingually deafened children with cochlear implants. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology, 34*(3), 499-506. doi:10.1097/MAO.0b013e318287ca86 [doi]

Zhu, Y. (2013). *Expression and recognition of emotion in native and foreign speech: The case of mandarin and dutch* Netherlands Graduate School of Linguistics.

Zhu, Z., Miyauchi, R., Araki, Y., & Unoki, M. (2016). Recognition of vocal emotion in noise-vocoded speech by normal hearing and cochlear implant listeners. *The Journal of the Acoustical Society of America*, *140*(4), 3271-3271.

BIOGRAPHICAL SKETCH

Cecilia L. Pak, formerly known as Sishi Liu, is a PhD candidate in the Communication Sciences and Disorders program at The University of Texas at Dallas. After graduating from of Capital Medical University (Beijing, China), she trained in several clinical settings. During her internship in the Department of Otolaryngology (Beijing Tongren Hospital, China), she developed a passion for studying audiology and hearing science in people with hearing impairments. In China, she worked on various projects such as sound localization program of individuals with cochlear implants (2009) and central auditory functional development after cochlear implant surgery (2008). At The University of Texas at Dallas, after earning a master's degree in Applied Cognition and Neuroscience, she worked with Dr. William Katz in the Speech Production Lab, and in the Dallas Cochlear Implant Lab, under the guidance of Dr. Tobey and Dr. Warner-Czyz. She also enjoys teaching and is a teaching assistant at UT Dallas (Neuroanatomy, Functional Neuroanatomy, Exceptional Children, Anatomy and Physiology of Speech and Hearing, Neural Basis of Communication, and Audiology)

CURRICULUM VITAE

Pak, Cecilia Liu
cecilia.liu@utdallas.edu
The University of Texas at Dallas, 800 W. Campbell Rd.
Department of Behavioral and Brain Sciences, Richardson, TX
75080

Education

The University of Texas at Dallas, Dallas, U.S.A.
Ph.D. in Communication Sciences and Disorders, 2018

The University of Texas at Dallas, Dallas, U.S.A.
M.S. in Applied Cognition and Neuroscience,2011

Capital Medical University, Beijing, China
M.D. in Clinical Medicine, 2008

Teaching and research experience

Teaching Assistant, several neuroscience and communication disorders classes, including Neuroanatomy, Functional Neuroanatomy, Exceptional Children, Anatomy and Physiology of Speech and Hearing, Neural Basis of Communication, and Audiology. The University of Texas at Dallas. 2009-current

Research Assistant, Beijing Institute of Otorhinolaryngology (Beijing, China). 2008-2009.

Clinical experience

Intern, Department of Otolaryngology - Head and Neck Surgery, Beijing Tongren Hospital (Beijing, China). 2008-2009.

Department of Internal Medicine, Beijing Tongren Hospital (Beijing, China). 2007-2008.

Department of Pediatrics, Beijing Tongren Hospital (Beijing, China). 2007-2008.

Department of OB/GYN, Beijing Tongren Hospital (Beijing, China). 2007-2008.

Department of Surgery, Beijing Tongren Hospital (Beijing, China). 2007-2008.

Department of Orthopedics, Mindong Hospital (Fujian, China). 2006.

Publications

Katz, W. F., **Pak, C. L.**, & Shin, S. (2018). Acoustics and emotion in tonal and non-tonal languages: Findings from individuals with typical hearing and with cochlear implants. *The Journal of the Acoustical Society of America*, *144*(3), 1840-1840.

Pak, C. L., & Katz, W. F. (2017). Recognition of emotional prosody in mandarin: Evidence from a synthetic speech paradigm. *The Journal of the Acoustical Society of America*, *141*(5), 3701-3701.

Liang, S., **Liu, S.,** Li, Y., & Han, D. (2009) "The Research Approach of Audiology and Speech Rehabilitation after Cochlear Implantation", *Journal of Audiology and Speech Pathology*, 17(1), 61-63. In Chinese.

Oral and poster presentation

Pak, C.L., Sun, W., Katz, W. F., & Zhang, H. (2018, March). *Recognition of vocal emotion in Mandarin speaking adults with cochlear implants.* Poster session presented at the CI2018 Emerging Issues in Cochlear Implantation, Washington, DC.

Pak, C.L., & Katz, W. F. (2017, June). *Recognition of emotional prosody in mandarin: Evidence from a synthetic speech paradigm.* Poster session presented at the 173rd Meeting of the Acoustical Society of America (ASA**)**, Boston, MA.

Pak, C.L. (2016, November). *Linguistic and Emotional Prosody in the Speech of Cochlear Implanted Mandarin-Chinese speakers*. Oral presentation at the Friday Seminars in Speech, Language, and Hearing (FLASH), Dallas, TX.

Liu, S., Tobey, E.A., Geer, A., & Sundarrajan, M. (2013, November). *Comparison of sounds produced in children with cochlear implants*. Poster session presented at American Speech-Language-Hearing Association (ASHA) Convention, Chicago, IL.

Liu, S., Tobey, E.A., Sundarrajan, M., Nicholas, J. (2011, November). *Comparison of sounds produced in young children with cochlear implants.* Poster session presented at Annual Convention of ASHA, San Diego, CA.

Liu, S. (2011, April). *Comparison of sounds produced in young children with cochlear implants.* Poster session presented at the Callier Promotion of Academic and Clinical Excellence (PACE) Research Forum, Dallas, TX.

Liu, S. (2010, April). *Speech production after cochlear implantation at 12 and 24 months of age*. Poster session presented at the Callier PACE forum, Dallas, TX.

Awards and grants

Jim Jerger Research in Audiology Fellowship (2018)
UTD small PhD research grant (2017)
Applied American Otological Society (AOS) Fellowship (2018)
Applied Plural Research Scholarship (2018)
Applied the Council of Academic Programs in Communication Science and Disorders (CAPCSD) Ph.D. Scholarship (2018)