Melodic contour training and its effect on speech perception for cochlear implant recipients

Chi Yhun Lo
BSpHSc, Macquarie University, 2013

Centre for Language Sciences
Music, Sound and Performance Lab

Department of Linguistics
Faculty of Human Sciences
Macquarie University, Sydney, Australia

This thesis is submitted in fulfilment of the requirements for the degree of
Master of Research (Linguistics)

October 2014

Table of contents

Abstract v

Declaration vi

Acknowledgement vii

Acronyms and abbreviations viii

Figures ix

Tables xi

1 Overview 1

1.1 Background and motivation 1

1.2 Objectives 2

1.3 Outline 2

1.4 Research questions 3

2 Introduction 4

2.1 Cochlear implants 4

2.2 Sound processing strategies 5

2.3 Pitch perception 6

2.4 Intensity Perception 9

2.5 Speech perception 9

2.5.1 Vowels and consonants 11

2.5.2 Prosody 16

2.5.3 Speech in noise 18

2.5.4 Cognitive factors 20

2.6 Variation in auditory performance 21
2.7 Music 23
2.8 Music perception 23
2.9 Training and transfer..26
2.9.1 Brain plasticity 26
2.9.2 The musician effect..28
2.9.3 OPERA hypothesis...29
2.9.4 Musical Training..30
2.9.5 Melodic contour training: interval size and note durations...............................32
3 Materials and Methods 34
3.1.1 Participants 34
3.1.2 Melodic Contour Training Program (MCTP) 36
3.1.3 Materials 40
3.1.4 Procedures..46
3.1.5 Statistical methods 47
4 Results 48
5 Discussion and conclusion 63
6 References 72
7 Appendix A 86
8 Appendix B 88
9 Appendix C 94

Abstract

Cochlear implant (CI) recipients have generally good performance for speech in quiet, but have difficulty in adverse conditions such as noise, and more complex tasks such as music listening. Auditory training has been proposed as a means of improving speech perception for CI recipients, and most recent efforts have been focussed on the potential benefits that music-based training may have.

This study evaluated two melodic contour training programs that varied in musical mechanism, and evaluated their relative efficacy as measured on speech perception tasks. These melodic contours were simple 5-note sequences formed into 9 patterns such as “Rising” or “Falling”. One training program controlled difficulty by manipulating interval sizes, the other by note durations. Sixteen CI recipients and twelve normal hearing listeners were tested on a speech perception battery for a baseline measure, and then commenced melodic contour training for 6 weeks, after which they were retested.

Results indicated there were some benefits for speech perception tasks after melodic contour training. Specifically, consonant perception in quiet was improved, as was question/statement prosody. There was no significant difference between either musical mechanism, suggesting that both conferred benefits for training CI recipients to better perceive speech.

Declaration

I hereby declare that this thesis has not been submitted for any degree to any other university or institution. The sources of information used and the extent to which the work of others has been utilised have been indicated in this thesis in the manner conventionally approved in the research field in which the thesis fits.

The research in this thesis has been approved by the Macquarie University Faculty of Human Sciences Research Ethics Sub-Committee (reference: 5201400348).

Chi Yhun Lo

10/10/14

Acknowledgement

I am ever so grateful to my principal supervisor, Associate Professor Catherine McMahon, for being a guiding force and inspiration in my very early research career from day one. Thank you Catherine, for affording me so many opportunities.

I have also been an exceptionally fortunate student, and find myself supported by two researchers who are influential in the field of music research; my associate supervisor, Professor William Forde Thompson; and my adjunct supervisor, Dr Valerie Looi. I am ever-thankful for their knowledge, guidance, and support.

Many thanks to Ian Deng, for programming the Melodic Contour Training Program which was such a key component of this study; Brett Swanson for his technical assistance with the AuSTIN materials and the pitch ranking task; Felicity Cox for helping with the interpretation of the confusion matrices; and Robert Mannell for having supplied me with both encouragement and speech materials.

The clinical staff at SCIC have been incredibly generous with their time, and the recruitment of cochlear implant recipients would not have been possible without the assistance of Jane Brew and Samantha Barrett - many thanks. I am also grateful to all the people who have volunteered to be a part of this study.

I also acknowledge the financial support for this study provided by the Research Training Pathway Scholarship from Macquarie University.

Finally, thanks to my family, friends, and in particular my partner Arani - looking forward to spending time with you all… at least until the next project!

Acronyms and abbreviations

4AFC	Four-alternative forced choice paradigm
4TB	Four-talker babble
A + E	Acoustic and electric hearing
ACE	Advanced Combination Encoder
AuSTIN	Australian Sentences Test in Noise
CI	Cochlear implant
CNC	Consonant-nucleus-consonant
ELU	Ease of Language Understanding (model)
F0	Fundamental frequency
F1	First formant
F2	Second formant
JND	Just-noticeable-difference
MBEP	Macquarie Battery of Emotional Prosody
MCI	Melodic contour identification
MCTP	Melodic Contour Training Program
NH	Normal hearing
OPERA	Overlap, Precision, Emotion, Repetition, Attention (hypothesis)
PEBL	Psychology Experiment Building Language
PEPS-C	Profiling Elements of Prosody in Speech-Communication
PPS	Pulses per second
SIN	Speech in noise
SNR	Signal to noise ratio
SRT	Speech reception threshold
SSQ	Speech, Spatial and Qualities of Hearing Scale

Figures

Figure 2.1. Envelope extraction 6

Figure 2.2. Spectrographs of 8 vowels.. 12

Figure 2.3 Consonant and nasal stops.. 15

Figure 3.1. The 9 melodic contours used in the Melodic Contour Training Program 36

Figure 3.2. Melodic Contour Training Program in Practice Mode 39

Figure 3.3. Melodic Contour Training Program in Training mode. 39

Figure 3.4. Spectrograms for voiced stop consonants /ba, da, ga/ 42

Figure 3.5. Intonation curves for PEPS-C stimuli 45

Figure 4.1. Week-to-week interval threshold scores for the Melodic Contour Training Program (Interval group) 50

Figure 4.2. Week-to-week duration threshold scores for the Melodic Contour Training Program (Duration group). 51

Figure 4.3. Baseline and post-training efficiency scores for the Victoria Stroop Test. 52

Figure 4.4. Baseline and post-training performance for the pitch ranking test 53

Figure 4.5.Baseline and post-training SRTs for AuSTIN. 54

Figure 4.6. Baseline and post-training performance for consonant discrimination in quiet. ..55

Figure 4.7. Confusion matrix for Interval group 56

Figure 4.8. Confusion matrix for Duration group. 57

Figure 4.9.Baseline and post-training performance for the consonant discrimination with 4TB. 58

Figure 4.10. Baseline and post-training performance for the MBEP 60

Figure 4.11. Baseline and post-training performance for the PEPS-C....61

Figure 4.12. Baseline and post-training performance for the SSQ12....62

Figure 8.1. Individual results from training across weeks. 88

Figure 8.2. Individual results for VST 88

Figure 8.3. Individual results for Pitch Discrimination (131 Hz) 89

Figure 8.4. Individual results for Pitch Discrimination (440 Hz) 89

Figure 8.5. Individual results for AuSTIN 90

Figure 8.6. Individual results for consonant discrimination in quiet. 90

Figure 8.7. Individual results for consonant discrimination with 4TB. 91

Figure 8.8. Individual results for MBEP(Happy/Sad) 91

Figure 8.9. Individual results for MBEP(Fear/Irritation). 92

Figure 8.10. Individual results for prosody discrimination (question/statement) 92

Figure 8.11. Individual results for SSQ12 scores 93

Tables

Table 2.1. The 12 most common Australian English consonants. 14

Table 3.1. Demographic information for Cochlear implant (CI) recipients 35

Table 3.2. Acoustic features for stop consonants 42

Table 3.3. Acoustic features of the Macquarie Battery of Emotional Prosody 44

Table 3.4. Acoustic features for statement intonations...44

Table 3.5. Acoustic features for question intonations 45

Table 4.1. Main effects of Session, Program, and Interactions for all tests 48

Table 7.1. MCTP (Interval) frequency and duration information 86

Table 7.2. MCTP (Duration) frequency and duration information 87

1 Overview

1.1 Background and motivation

Cochlear implants (CIs) are remarkable bionic devices that enable the perception of sound for most persons diagnosed with severe to profound deafness. Designed primarily for the purpose of speech perception, they are generally satisfactory in quiet environments. CIs are not without limitations however, and difficulty with speech in noisy environments is a commonly cited problem (Stickney et al., 2004). As oral communication occurs in the presence of complex and noisy acoustic environments, this limitation can have a direct impact on social interactions and quality of life outcomes.

Advances in our understanding of neuroplasticity and learning capacity, has led to interest in formal auditory training with many investigators proposing that it may form a component of comprehensive (re)habilitation (Boothroyd, 2007; Fu & Galvin 2007; Olson & Canada, 2013). As some studies have demonstrated that normal hearing (NH) musicians are particularly adept listeners under a range of acoustic environments and tasks (Fuller et al., 2014; Kraus, 2010), the incorporation of music as a tool for improving language based tasks is a focus for many studies (e.g. Besson, Chobert, & Marie, 2011; Kraus & Chandrasekaran, 2010; Patel, 2014). Music and speech are perceptually unique, but also feature many similarities and overlapping features. The growing body of evidence for a “musician effect” (Fuller et al., 2014); the potential for overtraining skills with the complexity of music perception relative to speech perception (Patel, 2011); and the connection with tonal languages, has generated great interest in music training and its generalizability to speech perception, particularly when the auditory signal is poor.

1.2 Objectives

- To develop, evaluate, and compare the outcomes of two melodic contour training programs for cochlear implant (CI) recipients.
- To explore the transference of basic non-linguistic musical abilities to speech perception and attention.
- To investigate the relative efficacy of training between two musical mechanisms: pitch intervals (frequency), and note durations (time) for CI recipients.

1.3 Outline

The thesis consists of five chapters.

Chapter 1 provides an overview of the research objectives and background for the study.

Chapter 2 provides a broad introduction to the cochlear implant system and associated auditory perception with cochlear implants. It also reviews music training, neuroplasticity, and the theoretical frameworks exploring the connection between music skills, training, and speech perception.

Chapter 3 consists of the materials and methods (participants, test stimuli, the Melodic Contour Training Program, procedures, and statistical methods) used in the study.

Chapter 4 provides the results of the current study.

Chapter 5 provides a general discussion, limitations of the study, areas for further research, and conclusions.

1.4 Research questions

The key research questions the experimental work in this thesis sought to answer were whether individuals with a CI improve on a trained music task, and whether these skills were generalizable to specific elements involved in speech perception. In particular, they sought to address:

- Does a melodic contour training program using different musical mechanisms, such as pitch and note duration, improve specific areas of speech perception, such as speech-in-noise, consonant discrimination, question/statement identification, emotional prosody, or provide subjective benefit of hearing or communicative ability?
- Can a melodic contour training program improve the tracking of the fundamental frequency, important for intonation cues, and improve discrimination of formant trajectories, important for consonant discrimination, such as /ba, da, ga/?
- Does melodic training confer benefits to selective attention, needed for speech perception in more complex listening environments, such as in noise?

2 Introduction

2.1 Cochlear implants

The cochlear implant (CI) is a biomedical device that extracts acoustic information, and electrically encodes this into a signal that can be interpreted by the central nervous system as the sensation of sound. The key components required in a CI system are: 1. a microphone to receive sound; 2. A speech processor that converts the microphone output into a stimuli for an array of electrodes implanted in the cochlea; 3. A connection across the skin that transmits power and stimuli; 4. An implanted receiver/stimulator that will receive, decode, and transfer signal to individual electrodes within an electrode array (Wilson & Dorman, 2008). As a CI system can bypass missing and damaged cochlear hair cells, they are a suitable form of intervention for people with severe to profound sensorineural hearing loss.

A Cochlear Ltd. Nucleus implant has 22 intra-cochlear electrodes, while the more recent Nucleus 24 and Freedom implants have an additional pair. Electrodes have a specific configuration known as the stimulation mode that describes the current signal flow between at least two electrodes known as the *active* and *reference* electrodes. These modes of stimulation include: 1. *bipolar*, the flow of current between a pair of electrodes; 2. *common ground*, one electrode is active and current flows through all other electrodes acting as a connected reference; 3. *monopolar*, the most commonly used mode in which current flows through one intra-cochlear electrode, and at least one extra-cochlear electrode (Grayden & Clark, 2008).

It is the electrical current pulses delivered through the electrodes of the implant that initiate the firing of auditory nerve fibres, which are transmitted to the brain, providing the sensation of sound. As the electrode array consists of multiple electrodes (typically 22 for a

Cochlear Ltd. device), the tonotopic distribution of frequencies as found in a normally functioning cochlea can be crudely replicated; that is, electrodes located in the basal position activate in response to high frequency sounds, while those located in the apical position activate in response to low frequency sounds.

2.2 Sound processing strategies

The function of the sound processor is to filter and extract acoustic information, and encode them with a radio frequency that is transmitted to the electrode array. Currently, the main sound processing strategies that are employed in CIs are: Continuous Interleaved Sampling (CIS), Spectral Peak (SPEAK), HiResolution (HiRes), Fine Structure Processing (FSP), and Advanced Combination Encoder (ACE). As all participants in the experiment component of this thesis used Cochlear Ltd. CIs from the Nucleus system range, the focus of this thesis will be on the default strategy - the advanced combination encoder (ACE).

ACE is a multichannel temporal-envelope based system. Before the intra-cochlear electrode receives stimulation, envelopes across multiple channels are “scanned” to identify signals with *n*-highest amplitudes that correspond with *m*-channels. In effect, this *n*-of-*m* approach employed in the ACE strategy will only encode the most salient acoustic features rather than stimulating the full set of electrodes, thus reducing the density of stimulation. In addition, ACE uses an interleaved stimulation such that only one channel is activated at any one time. Typical stimulation rates are between 250 and 2500 pulses per second (pps). The process of interleaving and the *n*-of-*m* approach avoids envelope cue smearing that may result when electrode interactions occur simultaneously (Zeng, 2004). This leads to the potential reduction of interference across both the electrodes and the regions of excitation across the cochlea (Wilson & Dorman, 2009). A schematic of an envelope extraction is provided in Figure 2.1.

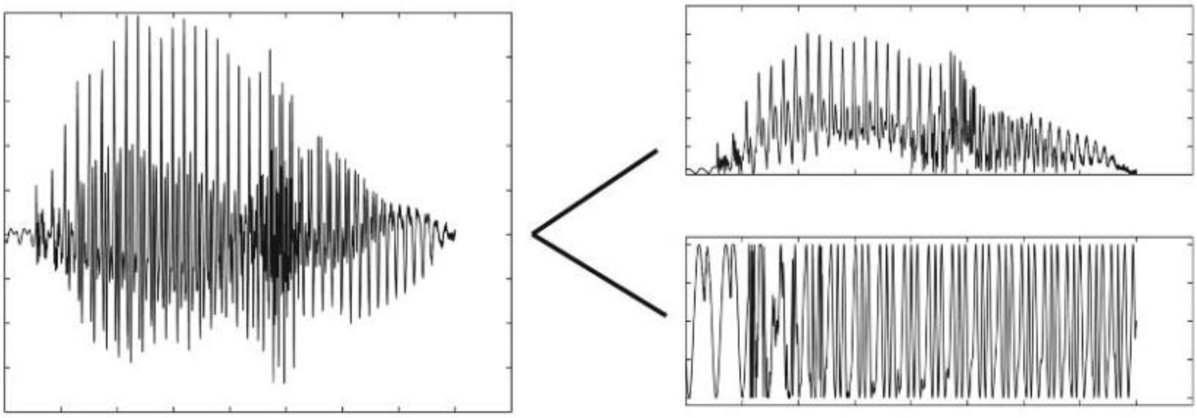


Figure 2.1. Envelope extraction. The left waveform represents a word with both temporal envelope and fine structure. The top right represents an extracted envelope, while the bottom right represents the fine structure. From Rubinstein (2004)

The temporal envelope system used in the ACE strategy is sufficient for speech perception in quiet, but poor performance in speech-in-noise and music perception are thought to be due to the lack of fine-structure processing available to the CI recipient (Lorenzi et al., 2006; Smith, Delgutte, & Oxenham, 2002). Additionally, beneficial acoustic cues may also be lost with the ACE strategy, as it makes the assumption that the maximal points across the envelope are the most relevant; but this is not always the case, for example, noise will consistently exceed speech in a noisy environment.

2.3 Pitch perception

Pitch perception is one of the most important auditory functions for both speech and music. In speech, the fundamental frequency (F0 - the lowest frequency of an acoustic waveform) plays a key role in the segregation and grouping of sounds necessary for auditory scene analysis (Bregman, 1994). In a tonal language such as Mandarin, pitch inflections form contrasts, such as in the well-known example of ‘ma’, whereby a steady inflection, a rising inflection, a falling-rising inflection, and a falling inflection, each confer different meanings. Pitch is also important in music, as melody is defined primarily by how pitch varies across time. Plack and Oxenham (2005) define pitch specifically in the context of music as “that attribute of sensation whose variation is associated with musical melodies”.

Pitch is the perceptual correlate of an acoustic waveform that is based (and varies) primarily with its frequency (or periodicity). With pure tone stimuli, pitch corresponds to the frequency of the fundamental, with low frequencies corresponding to low pitches, and high frequencies with high pitches. However, pure tones are an uncommon occurrence in real world acoustic stimuli. Complex harmonic tones are far more common, and consist of multiple pure tones (called harmonics) which have a mathematical relationship as integer multiples of the F0. The relative amplitudes of these harmonics have a great impact on the timbre, or sound quality that is perceived. It is also possible for the fundamental to be extracted from the relationship of the complex harmonics, when F0 is not present. This phenomena is known as the pitch of the missing fundamental (Schouten, 1962), and provides the ability to resolve F0, even when it is masked.

To date, mechanisms of pitch in NH individuals are still debated (see Yost, 2009 for a review), but the CI model of encoding pitch offers a unique paradigm to gain further insights, due to the possible isolation of various mechanisms (McKay, McDermott, & Carlyon, 2000). Two primary mechanisms of pitch are place coding, due to the tonotopic arrangement of the cochlea, and temporal coding, resulting from the timing of neural information arriving at the auditory cortex. It is likely that a combination of both are utilised by NH individuals.

However, CI strategies, particularly for Cochlear Ltd implant systems, tend to focus on place coding.

As the ACE strategy has a fixed, constant pulse rate, it relies on its multichannel, envelope based system and the natural tonotopic arrangement of the cochlea to provide place cues for pitch perception. This approach is generally successful with basal electrodes providing higher pitches than their apical counterparts (Busby et al., 1994). It should be noted that the electrode array cannot utilise the entirety of the cochlea, as the electrode cannot be inserted as far as the apical tip without increasing the risk of cochlear damage. This results in the apical portion of the cochlea receiving minimal low-frequency stimulation (Limb and Roy, 2013). Additionally, it is theoretically possible for periodicity cues to provide F0 information either by amplitude modulation of the pulse train, or by the rate of stimulation (Limb & Roy, 2014). But the fixed pulse rate of ACE means these are typically poorly accessed, and results by Zeng (2002) have found that CI recipients show saturation above 300 Hz. Even when this rate is increased, there is a tendency for channel interactions to occur, negating any potential perceptual pitch gains (Tang, Benitez, & Zeng, 2011). In addition, these periodicity cues are also particularly susceptible to distortion from noise and reverberation (Qin & Oxenham, 2005).

Generally, pitch ranking scores of CI recipients have been found to be significantly poorer than NH listeners (Sucher & McDermott, 2007). This poorer pitch perception is, at least in part, due to limitations with the CI device being unable to provide reliable F0 information, and a lack of fine structure features. The critical pitch perception ability to resolve the fundamental from its corresponding harmonic complex is simply not available with the ACE strategy, and direct stimulation of the apical portions of the cochlea are often not provided.

2.4 Intensity Perception

Intensity (and its perceptual correlate of loudness) is the auditory sensation that can be categorised on a scale from quiet to loud (ANSI, 1973). A person with NH has a dynamic range of approximately 120 dB perceivable in 200 discriminable steps, which far exceeds the upper range for CI recipients at 20 dB in about 20 steps (Zeng, 2004). The compressive function of the cochlea is not available for CI recipients, and because this dynamic range is so relatively reduced, loudness will increase or decrease at an exponential rate, resulting in much steeper/quicker changes for CI recipients (Zeng & Shannon, 1994). Normal functioning cochleae will provide exponential response at low frequencies, and a linear response for high frequencies, thus providing compressive loudness.

When multiple channels are stimulated sequentially, each channel contributes to the loudness that is perceived. Loudness perception for CI recipients is related to current, and pulse rate. As pulse rate has a fixed rate in the ACE strategy, current modulation creates different sensations of loudness. Intensity cues are an important feature for distinguishing between more sonorous vowels and consonants, various stop consonants and fricatives, and play a role in extracting prosodic features, and will be explained in further detail in the following sections.

2.5 Speech perception

Speech is a remarkably complex waveform that consists of multiple, constantly changing cues such as spectrum, pitch, and amplitude that can operate with a high degree of independence (Greenberg & Ainsworth, 2004). It is this ability to multitask, inhibiting non-relevant acoustic stimuli, and integrating speech cues, which provides meaning when signal information is transferred from the peripheral hearing system to the higher cognitive centres. Speech perception requires the mapping of (seemingly) continuous acoustic changes, and categorising them into discrete perceptual units (Bidelman, Moreno, & Alain, 2013).

Speech consists of multiple cues across spectral and temporal domains (Stevens, 1980). The primary cause for speech perception deficits in CI recipients is due to signal degradation (Moberly et al., 2005), as both spectral and temporal cues in CI devices provide a crude percept relative to that in NH listeners (Vandali et al., 2000). Despite limitations with the CI system, speech perception in quiet is typified by a generally high performance, due primarily to the robust nature of speech that features an inherently high level of redundancy (Rubinstein, 2004). Assmann and Summerfield (2004) found that redundancy contributed to understanding, by limiting confusions due to errors in speech production, bridging gaps due to loss and distortion in the speech signal, and compensating memory or attentional deficits. These are present at the acoustic level, phonetic level, linguistic level, and within semantic context (Assmann & Summerfield, 2004).

A study by Shannon et al. (1995) presented 8 NH listeners with spectrally-degraded speech tasks (consonants, vowels, and simple sentences). Despite the lack of spectral cues, performance was at 90% correct identification with primarily temporal cues. While this was not a perfect model of CI speech processing, it nonetheless showed that amplitude structure derived from gross temporal cues (such as in the ACE strategy) was sufficient for accurate speech perception - for situations when those temporal cues can be accessed.

2.5.1 *Vowels and consonants*

Vowels and consonants are universal to all languages of the world. In Australian English, they can be broadly defined by their distribution within words, with vowels being conceptualised as the nucleus, while consonants form encompassing syllable edges (Cox, 2012). In addition, as the centre of the syllable is associated with an open vocal tract, vowels are also regarded as more sonorous (Stevens, 1980). This relative change in amplitude is an important cue for distinguishing vowels from consonants (Stevens 1980). This distinction is relatively simple for CI recipients using ACE, as their instantaneous amplitude is well preserved for each channel (Moberly et al., 2014).

Traditionally, acoustic speech sounds are analysed from the perspective of speech production (Stevens, 1980). As vowels are formed with a greater oral cavity than consonants, a distinguishing feature of all vowels are their resonant characteristics known as formants. These resonant characteristics can be manipulated by tongue position, with the first formant (F1) being associated with tongue height, and the second formant (F2) being associated with the horizontal position of the tongue. F1 and F2 cues provide sufficient information to discriminate most vowels for NH listeners. Similarly for CI recipients, these two cues are generally sufficient for activating unique patterns of electrode stimulation (Marozeau, Simon, & Innes-Brown, 2014).

Performance between CI recipients and NH listeners is near parity when F1 and F2 have a clear and distinct separation, as found in the formant pattern of /i/ as shown in Figure 2.2 (Dorman, 1993). But there is greater difficulty for identification of vowels such as /ɑ/ and /u/ as the formant spacing is similar, and the remaining discriminating cue is high frequency intensity differences which require both fine-structure and intensity cues (Dorman, 1993). In a study examining vowel recognition in embedded sentences, Iverson et al. (2006) found that CI recipients used the F2 cue similarly to NH listeners, but the F1 cue was less correlated, possibly due to poorer spectral representation at the F1 frequency range. A study by Nie, Barco, and Zeng (2006) also found that vowel perception was more reliant on spectral cues, critically dependent on the number of electrodes, and minimally affected by stimulation rate; while consonant perception was more reliant on temporal cues, and independent of the electrodes between 4 and 12.

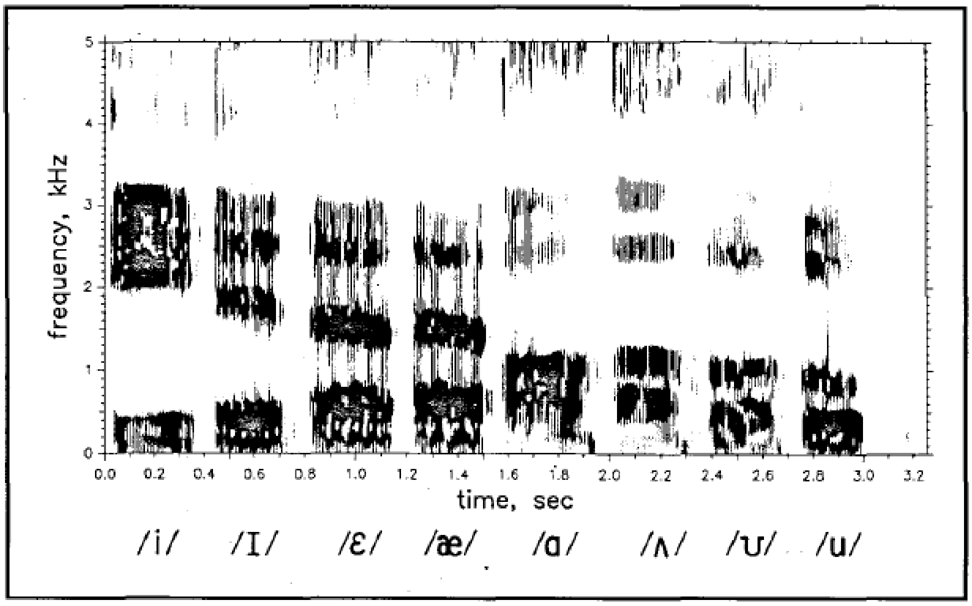


Figure 2.2. Spectrographs of 8 vowels. From Dorman (1993).

Duration cues are particularly important in Australian English, providing the distinction between the words “hid” and “heed”. Recent findings by Donaldson et al. (2013) have indicated that some CI recipients rely more heavily on duration cues to distinguish vowels, particularly when spectral cues are less accessible. This cue-trading, or cue-weighting is not unique to CI listening. NH individuals will weight relevant cues to more accurately perceive a phoneme. For example in a noisy environment, a NH listener may place more weight on spectral cues that are more salient, but as CI recipients have less available cues, the options for cue-weighting are more constrained (Moberly et al., 2014). As postlingually implanted CI recipients would have previously employed cue-weighting decisions based on their listening experiences prior to implantation, training programs may assist in helping recipients shift their cue-weighting strategies which may provide perceptual benefits (Moberly et al., 2014).

In Australian English, consonants are pulmonic sounds (created with outward airflow) that consist of multiple acoustic cues such as place of articulation, manner of articulation, and voicing (Cox, 2012). Place of articulation refers to the specific location at which articulation occurs (e.g. bilabial); manner of articulation describes the nature of the articulatory constriction (e.g. stops); and voicing refers to the movement or absence of movement in the vocal folds. Identification of consonants requires accurate discrimination of one, two, or all of these cues, depending on the context. Confusions occur when these cues are not correctly identified, for example if voicing is not heard when /ba/ is presented, the listener will most likely perceive /pa/ instead. Multiple studies have indicated that CI recipients find place of articulation distinctions the most difficult, followed by manner of articulation, and voicing as the best perceived (McDermott, 1993; Vandali, 2001). Place of articulation is the most difficult as it is reliant on both fine-structure temporal and spectral cues (Välimaa et al., 2002).

As with vowels, formants signal the identity of both intended consonant and speaker, but the additional cues of place of articulation, manner of articulation (associated with temporal gaps and durations), and voicing contrasts (derived primarily from waveform periodicity) (Nie, Stickney, & Zeng, 2005), provides additional redundancy, as well as complexity. In general, CI recipients perform reasonably well on consonant perception tasks with current generation CI systems, with multiple studies indicating a broad performance between 55 and 70% accuracy (Incerti, Ching, & Hill, 2011; Skinner et al., 2002).

There are 24 consonants in Australian English, but many studies use a subset of the 12 most common consonants (/p, t, k, b, d, g, f, v, s, z, m, n/) that provide a suitable selection of place of articulation, manner of articulation, and voicing contrasts (Plant, 1997). These 12 consonants and their contrastive features can be seen in Table 2.1.

Table 2.1. The 12 most common Australian English consonants.

Voiceless consonants are placed on the left of each cell, and voiced consonants on the right.

Manner of articulation		**Place of articulation**			
		Bilabial	Labiodental	Alveolar	Velar
	Stop	p b		t d	k g
	Nasal	m		n	
	Fricative		f v	s z	

Stop consonants are perceived when a subsequent pressure build-up is released (Cox, 2012); Stevens and Keyser (1989) suggest this sudden increase of intensity is a key distinguishing feature of stops in general. For the stop consonants /b, d, g, p, t, k/ and nasal stop /m/, the cues for place include the spectral properties in the burst of noise that follows the release, voice onset time (VOT), and subsequent F2 transitions into the adjacent vowel (Dorman, 1993; Liberman, Delattre, Cooper, & Gerstman, 1954); all of which are very short in duration, typically around 20 ms for VOT and up to 50 ms in total (Vandali, 2001). A study by Roman et al. (2003) investigated CI recipients' stop-discrimination ability using auditory evoked potentials. Their results indicated that CI recipients were able to represent the very short duration times in VOT at the cortical level. Distinction between voiced /b, d, g/, and unvoiced /p, t, k/ consonants can be made by distinguishing between small differences in the VOT and burst aspiration, but distinctions within classes such as /b/ from /d/, or /p/ from /t/ are reliant on tracking the trajectory of F2 (Liberman et al., 1954). These F2 trajectories are acoustically unique for each stop, as shown in Figure 2.3.

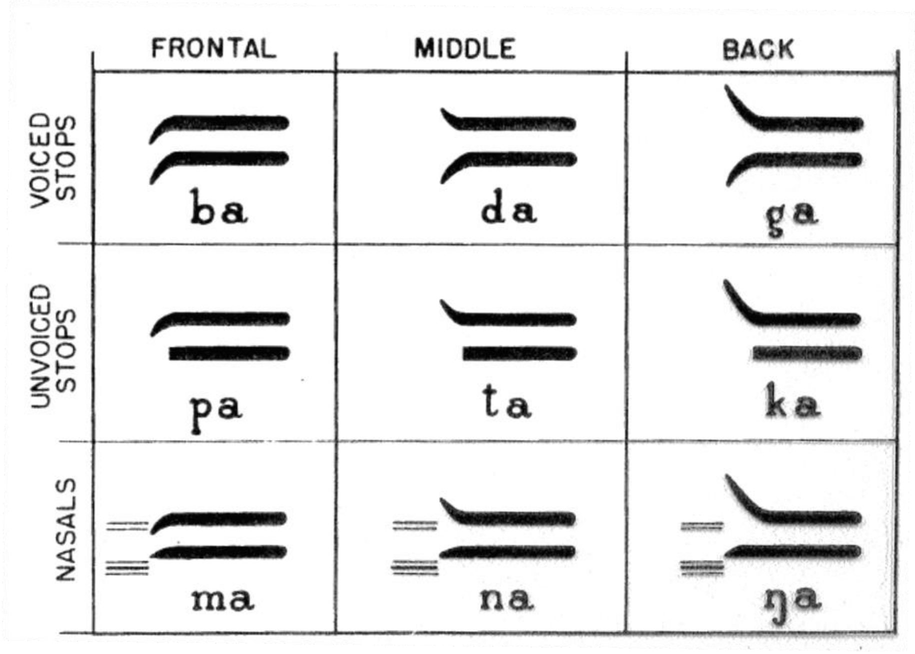


Figure 2.3 Consonant and nasal stops. From Lieberman and Blumstein (1988).

Fricatives are characterised by constriction to the airway passages that allows some measure of turbulent airflow to pass (as opposed to stops that create complete occlusions). Many place cues are embedded in the spectral content, with /s/ being characterised by energy above 4 kHz, in comparison to a relatively flat spectra found in /f/. But the main cues afforded by the ACE system for fricative distinction are intensity, rise time, duration, and silent interval duration, all of which can be accommodated by their temporal envelope. In the case of contrasting /f/ and /s/, intensity differences play a key role (Dorman, 1993).

2.5.2 *Prosody*

Prosody consists of patterns of intonation, rhythm, stress, and speech rate. While definitions vary across disciplines and their specific area of interest, broadly speaking - prosodic features provide the form of a spoken utterance (Cutler, Oahan, & van Donselaar, 1997). It is the aspects of speech which are modified to alter meaning at a suprasegmental level (i.e. above the phoneme level). While far from a complete and exhaustive list, Mozziconacci (2002) states prosodic cues provide information such as the speaker's gender, age and physical condition, emotion, attitudes, the dialogue partner, or the situation, all of which are referred to as indexical information by Luo, Fu, and Galvin (2007). Despite its importance in everyday speech communication, prosodic tests are not typically part of a test battery in clinical audiology practice.

Our two areas of focus can be split broadly into linguistic prosody, that confers the expression of meaning, and non-linguistic prosody that may indicate the emotional state of the speaker. These prosodic features provide the distinction between sentences being perceived as a question or a statement (linguistic), or the emotional state in a voice (non-linguistic). Abilities in these two areas of prosody are distinct; question and statement tasks have well-defined meaning and are associated with fairly broad intonation changes, whereas more subtle variations can greatly alter emotional meaning significantly (Thompson, Marin, & Stewart, 2012). Brain imaging with functional magnetic resonance imaging (fMRI) also supports this distinction, and found that linguistic prosody was linked to the left-sided language hemispheres, whereas emotional processing was linked to the activation of right hemisphere regions (Wildgruber et al., 2005).

The perception of question and statement identification is associated with intonation, and is perceived by listeners primarily as the variation in vocal pitch (mean F0 and variability) (Raphael, Borden, & Harris, 2007). Supplementary cues include loudness - related to the intensity of the sound (as previously discussed); the perception of stress - a combination of changes in frequency, intensity, and duration; and speech rhythm - the resulting combination of patterns from stressed and unstressed syllables in connected speech (Cruttenden, 1997). While NH listeners are able to use the dominant cue of F0, and lesser cues of intensity, stress, and rhythm; the poor encoding of F0 for CI recipients results in difficulty with question-statement prosodic identification as they are more reliant on weaker secondary cues (Meister et al., 2009; Peng, Chatterjee, & Lu, 2012). As such, CI recipients derive less benefit from natural intonation patterns, and are significantly poorer than NH listeners on question and statement tasks (Meister et al., 2009).

The parameters for determining emotional prosody of a speaker have a far higher requirement than question-statement tasks. The acoustic cues that encode vocal emotion are speech prosody, voice quality, and vowel articulation (Luo et al., 2007). In addition to the acoustic cues found in linguistic prosody (pitch, loudness, stress, and rhythm), cues relevant to emotional prosody include the first three formant frequencies, and the distribution of spectral energy (Luo et al., 2007). CI recipients perform poorly in the recognition of vocal emotions at approximately 45% correct compared to approximately 90% for NH listeners, as the multiple cues of pitch, vowel articulation, and spectral envelope cues, are all reduced as the result of poor spectral and temporal resolution provided by CI devices (Luo et al., 2007).

A study by Thompson et al. (2012) investigated emotional prosody in 12 individuals with congenital amusia. In general, this disorder is characterised by difficulties with music based tasks such as poor singing proficiency, difficulty with pitch perception, and poor melodic identification (Anderson et al., 2012). Results indicated that the amusic group was up to 20% poorer at decoding emotion than in the control group (Thompson et al., 2012). As the music perception characteristics of amusics are similar to CI recipients, it has been suggested that the training of musical skills may be beneficial for emotional prosody tasks (Petersen et al., 2012).

In general, facial expressions are a strong indicator for emotional states, but there are many occasions when visual cues are unavailable and only auditory cues can be relied upon (e.g. blocked line of sight, telephone conversation). Perception of emotions from nonverbal vocal utterances are critical to the understanding of emotional messages, which will feedback into how listeners' respond and react.

2.5.3 *Speech in noise*

The sonic landscape we have created as a result of modern industrial society is characterised by noise. For NH listeners, speech perception in noise is surprisingly robust due to various factors that prevent or compensate for distortion. Firstly, discrimination of phonemes is largely defined by salient formant peaks that are more robust to distortion than valleys, which would likely be masked by noise (Assmann & Summerfield, 2004). Second, waveform periodicity provides an effective means to identify and group sounds based on F0 (Bird & Darwin, 1998). Third, when noise is louder than the signal, formants of voiced sounds can disrupt the periodicity of competing sounds to aid in speaker separation (Summerfield & Culling, 1992). Finally, modulation of F0 may infer suprasegmental prosodic features that may aid in auditory stream segregation, and the identification of word boundaries (Summerfield & Culling, 1992). For many CI recipients, these factors are not a robust feature of their auditory processes, and as such, speech in noise perception remains the most commonly cited problem (e.g. Nelson et al., 2002; Plomp, Festen, & Bronkhorst, 1990).

The effect of noise on speech perception is a function of noise intensity as well as the type of noise itself. In steady-state noise, Plomp et al. (1990) found that hearing impaired listeners (mild to moderately severe, 35 to 65 dB hearing loss) required at least 2.8 dB higher signal-to-noise ratio (SNR) than their NH counterparts to achieve equal intelligibility. In fluctuating noise, the difference was even larger, with hearing impaired listeners requiring a 5.5 dB increase in SNR (Plomp et al., 1990). Temporally fluctuating noise is far more indicative of an everyday listening situation, and NH listeners are able to "listen in the dips" of the modulating noise to extract speech information far more readily than CI recipients (Nelson et al., 2002). Compared to findings by Shannon et al. (1995) that showed as few as 4 spectral channels were required for adequate speech perception in quiet, it is estimated more than 12 channels are required for maximum speech in noise performance in CI recipients (Dorman et al., 1998; Fu et al., 1998). While the mechanisms for noise disrupting speech are not yet completely understood, it is thought that device limitations, poor pitch perception, and poor spectral-fine structure representation all play a role in speech in noise performance for CI recipients.

When the intelligibility of the target is overlapped with a masker such as noise, reverberation, or additional distracting speakers, there is the tendency for speech perception to be greatly reduced. The ability to spatially separate a target from maskers is known as spectral release from masking (SRM). A target-masker interaction occurs when there are temporal fluctuations, due to breaks between words and sentences, and regions of silence from stop consonants. In these fluctuations, it is possible to use glimpsing - the extraction of a brief "snapshot" of a continuous temporal sequence that can link the signal that has been separated due to the masker. This allows for the signal to be reconstituted and interpreted with its original intent (Assmann & Summerfield, 2004; Mattys et al., 2012). In general, SRM is much poorer for CI recipients than NH listeners (Feston & Plomp; Peters et al, 1998). Nelson et al. (2003) suggests the difference in performance between quiet and noisy settings for CI recipients using envelope cues is because ideal perception is severely disrupted by the random envelope of noise. Multiple studies have also concluded that CI recipients' speech in noise perception is negatively affected by the lack of fine structure due either to the limited number of spectral channels, or as a result of the type of speech processing strategy (Fu & Nogaki, 2005).

2.5.4 *Cognitive factors*

As the efficacy of the CI device itself has improved over the years to provide better speech perception outcomes for recipients, there has been a move to explore and better understand cognitive factors such as learning, attention, and memory; and how they contribute to speech encoding and perception (Pisoni, 2000). This information-processing approach is concerned with top-down, centralised processes that seek to identify how the nervous system translates sensory information (bottom-up processes) into meaningful speech perception (Bidelman et al., 2013).

The Ease of Language Understanding model (ELU model; Rönnberg et al., 2013) is one approach that posits working memory (WM - defined as a temporary store of information processing, necessary for more complex tasks such as learning) is fundamental for understanding language, irrespective of modality (e.g. sign language, or speech in adverse conditions). With specific relation to speech perception, WM is correlated with the ability to selectively attend to one channel, divide attention between channels, and generally reduce interference from distractors – all of which are relevant to adverse listening situations where multiple stream segregation is required (Dalton, Santangelo, & Spence, 2009).

Early attentional processes are also known to interact with WM (Rönnberg et al., 2013). Although the question of precisely where (and how early) in the signal process this occurs remains unknown (Rönnberg et al., 2013), recent findings by Mesgarami and Chang (2012) have found cortical evidence for the representation of attentional focus, in relation to the listener's intended auditory target.

Most social environments are punctuated by multiple, competing speakers that act as maskers and distractors. In group conversations, the more people that partake will also lead to an increase in unpredictability between conversation topic and speaker, thus reducing the use of redundant cues such as context. Such (common) scenarios require dynamic changes in attentional focus to maintain a conversational connection (Shinn-Cunningham & Best, 2008). Unfortunately, many of the perceptual cues required for effective complex scene analysis (such as fine-structure) are diminished or absent in CI processing strategies, which in turn leads to greater cognitive processing demands (Shinn-Cunningham & Best, 2008). Additionally, listening difficulties in complex environments are exacerbated in older adult recipients, as attentional load affects speech perception in adverse conditions, typically resulting in mental fatigue (Tun, McCoy, & Wingfield, 2009).

2.6 Variation in auditory performance

For the majority of adult recipients, CI systems provide moderate to good levels of auditory-only speech (e.g. conversing over a telephone). Although speech perception outcomes are generally excellent for most CI recipients, there are significant levels of variability in outcomes. Blamey et al. (1996) found that the duration of deafness, and duration of CI experience had the greatest effect on auditory performance. Many investigators have found that age of implantation also has an effect, but Leung et al. (2005) suggest a more accurate measure is the ratio of duration-of-deafness to age-of-implantation for adults with symmetric losses.

Another factor is the extent to which a recipient can combine, utilise, or rely on non-auditory cues, such as lipreading. Lipreading is particularly effective when the acoustic signal is degraded, as it provides visual cues for articulators such as the lips and tongue. These in turn have acoustic correlates that can help infer the original signal (Summerfield, 1992). A recent approach to implantation is to partially insert the electrode array, leaving the apical section exposed. This allows for the retention of residual hearing in which low frequencies are acoustically stimulated, while the higher frequencies are electrically stimulated. This combined acoustic and electric (A + E) mode may improve pitch perception abilities for low frequencies, allowing for the F0 to be better perceived, enhancing speech in noise perception (Turner et al., 2004). However this is only suitable for a limited number of individuals with high-frequency hearing loss (Sampaio, Araújo, & Oliveira, 2011). Acoustic hearing can also be gained with the use of a hearing aid (HA) contralateral to the CI (bimodal hearing), potentially preserving low-frequencies (Cullington & Zeng, 2011).

Pitch reversals (e.g. an apical electrode provides a higher pitch relative to a basal electrode) may also occur due to various reasons. The most probable reason is due to a mismatch between the electrodes and the cochlea, due to the sub-optimal placement of the electrode array, both in the terms of the depth of insertion, as well as their proximal position to nerve fibres (Vandali et al., 2005; Zeng, 2004). Yukawa et al. (2004) found that recipients with deeper insertions tended to score higher on speech perception tasks, though this also increases risk of damage to the cochlea, potentially affecting the use of residual hearing as a result (Zeng et al., 2008).

2.7 Music

Music provides many roles, whether performative, as a passive listener, or active participant. Music is a particularly powerful construct for the expression of emotion that can help regulate mood, enhance memory, and create important social bonds (Gfeller, 2008; McDermott & Hauser, 2005); and is important for developing a sense of identity of self, particularly in young people (Bennett, 2000), all of which contribute greatly to quality of life outcomes.

In light of this, it is unfortunate that such a large difference exists between NH listeners and CI recipients for music perception and appreciation. A questionnaire based study by Gfeller et al., (2000) found that CI recipients were far less likely to spend time engaged in music listening after implantation when compared to pre-implantation, with many recipients citing that music no longer sounded natural, pleasant, or enjoyable, such that music was deliberately avoided or dismissed. Additionally, in separate studies by Mirza et al. (2003) and Lassaletta et al. (2008), CI recipients reported that their enjoyment of music was significantly lower at post-implantation compared with pre-implant ratings.

2.8 Music perception

Music perception is the broad recognition of patterns that are organised across the four major dimensions of pitch, loudness, timbre, and duration (Gfeller et al., 1997; Krumhansl & Iverson, 1992). The combination, interplay, and relative contributions of these dimensions are what provide our understanding of music (Limb & Rubenstein, 2012; Looi, Gfeller, & Driscoll, 2012). As complex waveforms, speech and music both share many features that are defined primarily by their pitch, loudness, temporal, and timbral features. Limb and Roy

(2014) emphasize that music remains the most complex auditory stimulus in existence, and as such provides a remarkable paradigm for fully appreciating and understanding the auditory system.

Results by Shannon et al. (1995) showed that as few as four spectral channels were required for adequate speech perception in quiet, while effective music listening was estimated at requiring at least 64 channels (Smith et al., 2002). As of yet, no CI system comes close to providing this degree of fine-structure information, and without this sufficient number of channels, music perception remains severely limited for the majority of CI recipients (Cooper et al., 2008; Crew et al., 2012).

Pitch has a particularly special role in music, as musical scales are derived from it, and because it is crucial for melody - defined as the organisation of consecutive pitches (Tovey, 1975). Importantly, melody recognition requires identification of contours and its relative pitch relationships (rather than absolute pitch). Pitch or melodic contours are the overall direction of a musical phrase, and all melodies can be broadly described in terms of a combination of rising, falling, or steady contours. Melodic contour identification (MCI) consisting of 5-note sequences forming 9 basic patterns (e.g. “rising”, “flat”, and “falling”, or a combination thereof) was explored across various studies in CI recipients by Galvin, Fu, and Nogaki (2007) and Galvin, Fu, and Shannon (2009). Results were highly variable among the 11 participants tested (90.7% to 14% correctly identified). Additionally, top performers were found to have greater pre or post-implant music experience, indicating that both experience and training may improve perceptual abilities.

Limited pitch perception by CI recipients has perhaps the greatest detrimental effect across music perception in general (Limb & Roy, 2014). NH listeners can accurately discriminate between a one semitone difference, whereas CI recipients have a much larger variation, with discriminability anywhere between one to eight semitones (Kang et al., 2009). A study by Looi, McDermott, McKay, and Hickson (2008) found that as a group, CI recipients performed at chance level for 3 semitones intervals. The frequency range and dynamic range is also far wider in music than in non-tonal speech. As CI recipients have a limited use of F0 information and poor ability to resolve many lower order harmonics, the result is a far greater impact on music perception than for speech (Galvin et al., 2007; Looi et al., 2012).

Rhythm is the gross temporal state of music (i.e. over a period of seconds rather than milliseconds) over which a sequence of beats and rests form a consistent pattern over time (Limb & Roy, 2014). On the whole, multiple studies have confirmed that rhythm performance is near equivalent for both CI and NH groups (Looi et al., 2008). As rhythm cues are the most readily perceived by CI recipients, rhythm often play a compensatory role for poor pitch perception, aiding in the recognition of common, simple melodies such as “Happy Birthday” that are not only distinct in pitch, but also in rhythmic pattern (Gfeller et al., 1997).

Another fundamental feature of music is timbre, a multidimensional attribute that combines spectral-temporal features to characterise tone qualities (Houtsma, 1997). It is timbre that is closely linked to instrument identification, as it helps distinguish instruments such as a flute from a guitar, when other features such as pitch and loudness are the same (Handel, 1995). A novel approach to investigating the relative contributions between spectral and temporal envelopes was studied using chimeras (waveforms that consisted of the envelope of one instrument, combined with the fine-structure features of another instrument). Both cues could be used by NH listeners, but because fine-structure cues are not available to CI recipients, only envelope cues could be used, resulting in much poorer timbral appraisal (Heng et al., 2012).

2.9 Training and transfer

Despite the many issues surrounding its validity and interpretation, a study by Rauscher, Shaw and Ky (1993) subsequently dubbed by popular media as the “Mozart effect”, found that short term spatial IQ gains were conferred simply by listening to Mozart’s piano sonata for 10 minutes, instigating a great deal of interest on the effect that music may have on behaviour. While those results are too temporary and more likely indicative of arousal and mood changes (Thompson, Schellenberg, & Hussain, 2001); more recent studies (with more appropriate interpretation) have been motivated by advances in the understanding of neural plasticity, new technologies for increasingly detailed brain imaging (Besson et al., 2011), evidence that musicians have perceptual auditory advantages compared to non-musicians (Besson et al., 2011; Musacchia et al., 2007; Parbery-Clark et al., 2012), and various theoretical frameworks to provide context (e.g. Patel, 2011). All of these contribute to the understanding and exploration of connections between music, and perceptual and cognitive enhancement.

2.9.1 *Brain plasticity*

Brain (or neural) plasticity is the function of the brain to modify its neural connections and patterns, either as the result of internal (e.g. genetic), and/or external (e.g. environmental) forces. While learning is fundamental for our capacity to succeed and survive (Green & Bavalier, 2008), and is thought to play the primary role in the formation of neurophysiological and behavioural changes; both exist in a mutual co-dependency in which both influence and affect each other (Ryugo & Limb, 2000). While the nature of learning is a continuous process, there are also well defined critical periods in which there is enhanced sensitivity to neurosensory functions. In the case of cochlear implantation, plasticity is a function of age, with maximal plasticity at around 3.5 to 7 years old (Sharma, Dorman, &

Spahr, 2002). But the majority of adult CI recipients are postlingually deafened (ASHA, 2004). As such, these recipients have an existing pre-implant mental lexicon or “reference” of auditory stimuli, so their auditory system must undergo both organisation of new networks and reorganisation of existing pathways (Harrison, Gordon, & Mount, 2005). While the perceptual abilities for most postlingually implanted CI recipients are generally poorer than prelingual CI child recipients, and despite the diminishing effect of age on plasticity, the benefit to quality of life outcomes and capacity to learn is evidence this dynamic process exists in adults.

The functions of learning are broad, and this thesis will consider the function of skill learning specifically, defined as the changes that can be measured in perceptual, cognitive, or motor performance that persist for several weeks or more (Green & Bavalier, 2008). A common approach to investigate plasticity is to examine differences between populations with specific learned experiences (e.g. musicians and non-musicians), evaluating a range of skillsets, making comparisons, and most interestingly, interpreting their influence on unrelated processes (Moreno & Bidelman, 2014).

Barnett and Ceci (2002) categorised transfer effects as near or far, relative in proximity to context and content. Intuitively and empirically, the majority of tasks do transfer to highly similar contexts (i.e. near transfer). For example, a study by Thompson, Schellenberg, and Hussain (2004) found perceptual enhancement of prosodic cues for children who were provided with music lessons. But the results for far-transfer effects are mixed, for example Ho, Cheung and Chan (2003), showed that music training improved verbal memory, but not visual memory; while Schellenberg (2004) found an overall improvement for general IQ scores in children irrespective of training group (keyboard, vocal, drama, and control) over the period of a year, but those in the music-based groups had the largest increase.

2.9.2 *The musician effect*

Studies that have focussed on the effect of music training have found significant neurophysiological and behavioural differences between NH musicians and non-musicians. This “musician effect” is the observation that NH musically-trained individuals generally have perceptual, lingual, and cognitive enhancements relative to their non-musician controls (Fuller et al., 2014; Moreno et al., 2009). The mechanisms for such enhancements have been the interest of many studies, across a range perspectives and disciplines. From a neurobiological basis, music training is associated with enhanced general (i.e. not music specific) auditory processing, in which cortical responses are more highly differentiated to smaller acoustic changes, due to a greater volume of white matter - responsible for cortical connectivity, as well as musicians having more grey matter in both sensorimotor and auditory cortices (Gaser & Schlaug, 2003; Schmithorst & Wilkie, 2001).

The transfer of these effects to speech is of great interest due to its obvious application for remediation. Using auditory brainstem responses (ABR), Parbery-Clark et al. (2012) showed musicians had enhanced neural responses to improve speech syllable distinction, providing an advantage at the phoneme level for /ba, da, ga/. In addition, Strait and Kraus (2014) also found that musicians were both faster and more precise with the encoding of VOT and formant transition for consonants, though there was no difference for vowels. Perhaps the most significant auditory difference between young adult musicians and non-musicians, is that there is no significant advantage for speech perception in quiet environments, but musicians are faster to encode speech harmonics in more difficult and complex listening situations, resulting in positive correlations between years of music training and greater speech-in-noise perception (Parbery-Clark et al., 2009; Strait & Kraus, 2014).

There are limitations to the conclusions that can be drawn by the analysis of musicians versus non-musicians, as the disentanglement of training and predisposition is challenging. But positive correlations have been found between earlier commencement and duration of musical training, and overall auditory enhancement (Zendel & Alain, 2014); although indirect, it is suggestive of a causal relationship (Moreno & Bidelman, 2014). Longitudinal approaches are also indicative of causality. In a study by Francois et al. (2013), NH 8 year old children were assigned to either a music or painting group over the course of 2 years. Children provided with the music training showed behavioural and electrophysiological measures were improved for speech-segmentation skills, indicating that music training had a positive effect for overall language outcomes. However, recent findings by Mosing et al. (2014) in twins found that associations between music practice and music ability could be explained by genetic, rather than causal factors. It should be noted that these results can be interpreted as associations with general auditory abilities, and more specific, targeted music training (such as learning a musical instrument) should be reflected with enhancement of specific skills. Nonetheless, musicians provide an interesting model of auditory learning, with many implications for skill learning which may be of particular benefit for CI recipients.

2.9.3 *OPERA hypothesis*

The OPERA hypothesis (Patel, 2011), provides a framework consisting of five conditions that suggest why music training may drive perceptual speech gains. These are: Overlap - acoustic features relevant to both speech and music are encoded on overlapping brain networks; Precision - the requirements for music perception are higher than for speech; Emotion - the musical activity should elicit a strong positive emotion; Repetition - promotion of plasticity within the relevant networks must be subject to repeated action; Attention - focussed attention and engagement with the task will enhance the encoding process. When these conditions are met, they will drive the network to function at the degree of precision required for music, and as this precision is higher than for speech, there should be a flow on effect resulting in performance gains for speech perception.

In separate studies, Besson et al. (2011) and Kraus et al. (2012) showed that musicians had enhanced measures of auditory attention and working memory, arguing that both processes were particularly influential for effective transfer effects between music and speech perception. In light of these findings, Patel (2014) expanded the OPERA hypothesis to include: sensory or cognitive features such as attention and WM must overlap neural networks, as an additional requirement for effective transference of music skills to speech perception skills.

2.9.4 *Musical Training*

Recent efforts have been focussed on using auditory training as a tool for enhancing speech perception in CI recipients, potentially maximising the benefit derived from their CI device, mediated by changes in brain plasticity (Fu & Galvin, 2012; Patel, 2014). In particular, music training may be warranted, to improve pitch perception by helping CI recipients resolve conflicts inherent to existing CI processer strategies (Sucher & McDermott, 2007). The question arises as to why music would be more beneficial than a more direct, domain specific approach of speech training. The OPERA hypothesis offers the explanation that greater brain plasticity may occur due to the integration of positive emotions and sensorimotor actions that drive neural changes more effectively (Patel, 2014). In addition, musical activities are typically enjoyable, and as such participants may be more willing and motivated to participate in such tasks.

Gfeller et al. (2000) examined the effect of a structured, computer-based training program consisting of 48 modules (approximately 30 minutes each) over 12 weeks. A total of 24 CI recipients (assigned to training or control groups) trained on a series of simple and complex melodies, in highly familiar and unfamiliar contexts. Stimuli were generated with a Musical Instrument Digital Interface (MIDI) acoustic piano, so no linguistic (i.e. lyric cues) were provided, although an on-screen visual representation of the melody was presented. Results between pre and post-test measures of simple melody recognition, complex song recognition, and complex song appraisal indicated that structured training could provide benefits to recognition and appraisal of complex songs.

Petersen et al. (2012) investigated the effects of a 6 month one-to-one musical ear training program for 9 recently implanted recipients with matched controls. The training involved the development of pitch, rhythm, and timbre through singing, playing and listening exercises. Participants were tested on a musical test battery, the Hagerman speech perception test (HAG) - a speech in noise test, and an emotional prosody recognition test (EPR). Results showed that while training improved musical performance, there was no post-training improvement for any speech tasks, although the onset of improvement for emotional prosody was seen in the first 3-months - earlier than the controls. Thus, while there were no perceptual speech benefits, music training may facilitate the acceleration of brain plasticity.

Another form of music training that has been investigated is the use of melodic contours as the training stimuli. A preliminary study as described in Patel (2014), trained CI recipients to play melodic contours (a total of 9 patterns, consisting of 5-note sequences as used in Galvin et al. (2007), on a piano keyboard. Participants trained with semitone spacing between 1 and 3 semitones, with the hypothesis that practice in this task should develop greater precision for MCI. Length of training was approximately 30 minutes a day, 5 days a week, for 4 weeks. Measures taken at baseline and post-training were MCI (in a frequency range different to the training task), prosody perception of sentences that had a modified F0 trajectory over the last syllable (an upward trajectory sounds like a question, a downward trajectory sounds like a statement), and a Hearing in Noise Test (HINT). Although the results are preliminary and based only on two participants, early indications show melodic contour training may improve intonation prosody and speech in noise perception.

2.9.5 *Melodic contour training: interval size and note durations*

The majority of music training programs tend to focus on improving pitch perception for CI recipients. However, sounds are dynamic and multi-dimensional by nature, and the perception of pitch is necessitated by changes across time. Both music and speech can be conceptualised as hierarchical, syntactic structures (Patel, 2003). For example, a sentence corresponds with a melodic phrase, while a word corresponds to individual notes. With this in mind, a comparison between the average rate at which a musical note or spoken utterance is produced at, reveals some interesting results. An analysis of 5000 MIDI melodies by Watt and Quinn (2006) found mean note durations were 280 ms in length; an analysis of 16,000 syllables in American English by Greenberg (1996) found mean syllable utterances were 191 ms in length. Thus the time provided to extract cues is much shorter in speech than in music, generally. The perception of various consonants that use VOT contrasts (e.g. the distinction between voiced and unvoiced stops), or formant trajectory discrimination (e.g. to identify stops from within the voiced class such as /b/ from /g/) also rely on the extraction of cues across very short periods, 5 to 50 ms for VOT, and 5 to 50 ms for F2 trajectories, for effective perception (Vandali, 2001). As such, in order to fulfil the condition of "Precision" as described by the OPERA hypothesis, the exploration of shorter (and thus more difficult) note durations is proposed as a mechanism for effective transfer of musical training to improve speech perception in CI recipients.

The purpose of the present study was to develop and evaluate two take-home, computer based melodic contour training programs for CI recipients. The programs were self-adaptive, and differentiated by how the difficulty of the melodic contour was controlled: 1. Interval - the interval size was adjusted and note duration was fixed; 2. Duration - the note durations were adjusted and interval size was fixed. A key goal was to explore the transference of non-linguistic musical skills to specific aspects of speech perception and attention. Using a baseline and post-training paradigm we can measure the relative efficacy of each musical mechanism. It was hypothesised that:

- Both training programs should improve intonation perception, as melodic contours and intonation are acoustically similar with transitioning F0 cues required for question/statement identification.
- The training program with varying note durations should improve CI recipients' ability to process frequency changes more quickly, improving VOT and F2 trajectory perception, resulting in greater accuracy for stop consonants and nasal stops.
- The training program with varying pitch intervals will improve pitch perception, resulting in enhanced F0 tracking, improving emotional prosody perception, and potentially improving speech in noise perception.
- Both training programs should enhance the cognitive function of attention.

3 Materials and Methods

Approval for this study was granted by the Macquarie University Faculty of Human Sciences Human Research Ethics Sub-Committee (reference: 5201400348).

3.1.1 *Participants*

Sixteen adult postlingually implanted CI recipients (11 female, 5 male) that ranged in age from 26 to 86 (*M* = 58 years, *SD* = 15 years), and CI experience from 1 to 20 years (*M* = 9 years, *SD* = 7 years) participated in the study. For performance reference purposes, 12 NH adults (6 female, 6 male) that ranged in age from 21 to 42 (*M* = 27 years) were also included in the study. All participants were native Australian English speakers. On each day of testing, all NH adults had normal hearing thresholds (≤ 30 dB HL) measured in octave steps between 500 to 4,000 Hz.

All CI recipients used Cochlear Ltd. Implants, most of which were Nucleus devices, although a few participants used Freedom devices. Participants used CIs in unilateral, bilateral, or bimodal (with a hearing aid) configuration. Relevant demographic information can be found in Table 3.1.

Table 3.1. Demographic information for Cochlear implant (CI) recipients

ID	Age	Gender	CI	Processor	Strategy	No. of electrodes activated	Unilateral/ Bilateral/ Bimodal	No. of years implanted	Training Program
1	80	Female	L - CI24M R - CI24RE(CA)	L - CP810 R - CP810	ACE	L - 16 R - 22	Bilateral	20	Interval
2	26	Female	L - CI24RE(ST) R - HA	L - CP810	ACE	L - 22	Bimodal	1	Interval
3	66	Male	L - CI422 R - HA	L - CP810	ACE	L – 21	Bimodal	2	Duration
4	56	Female	L - CI24M R - CI24RE(CA)	L - CP810 R - CP810	ACE	L - 18 R - 22	Bilateral	14	Duration
5	35	Female	R - CI24RE(ST)	R - CP810	ACE	R - 22	Unilateral	1	Duration
6	61	Male	L - CI24R(ST) R - CI24RE(ST)	L - CP810 R - CP810	ACE	L - 22 R - 14	Bilateral	12	Duration
7	47	Female	L - CI24RE(CA) R - CI24R(ST)	L - CP810 R - CP810	ACE	L - 22 R - 20	Bilateral	10	Interval
8	86	Female	L - CI24RE(CA) R - CI24RE(ST)	L - CP810 R - CP810	ACE	L - 22 R - 18	Bilateral	8	Interval
9	52	Female	L - CI24RE(CA) R - CI24RE(CA)	L - Freedom for N24 R - Freedom for N24	ACE	L - 22 R - 18	Bilateral	10	Interval
21	54	Male	L - HA R - CI422	R - CP810	ACE	R - 21	Bimodal	2	Duration
22	48	Male	R - CI512	R - CP810	ACE	R - 22	Unilateral	4	Interval
23	69	Female	L - CI24RE R - CI24M	L - CP910 R - Freedom	ACE	L - 21 R - 22	Bilateral	15	Interval
24	66	Female	L - CI512 R - CI24RE(CA)	L - CP910 R - CP810	ACE	L - 19 R - 22	Bilateral	18	Duration
26	60	Male	L - CI24RE(CA) R - HA	L - CP810	ACE	L - 22	Unilateral	2	Duration
27	67	Female	L - CI24RE(CA) R - CI24M	L - CP810 R - CP810	ACE	L - 22 R - 20	Bilateral	15	Interval
28	55	Female	L - CI22 R - CI422	L - CP900 R - Freedom	ACE	L - 22 R - 15	Bilateral	19	Duration

3.1.2 *Melodic Contour Training Program (MCTP)*

Two take-home PC-based training programs were created using the Microsoft.NET framework: MCTP (Interval) and MCTP (Duration). The training paradigm was adaptive with the stimuli becoming more difficult after a correct response, and easier after every incorrect response, with a total of 7 levels of difficulty. The task was to listen to a randomly selected melodic contour, and identify it with a four-alternative forced choice paradigm (4AFC). The melodic contours were sequences of 5 consecutive notes that formed a total of 9 patterns: “Rising”, “Rising-Flat”, “Rising-Falling”, “Flat-Rising”, “Flat”, “Flat-Falling”, “Falling-Rising”, “Falling-Flat”, and “Falling” as used in Galvin et al. (2007). Figure 3.1 shows the melodic contours used in the training programs.

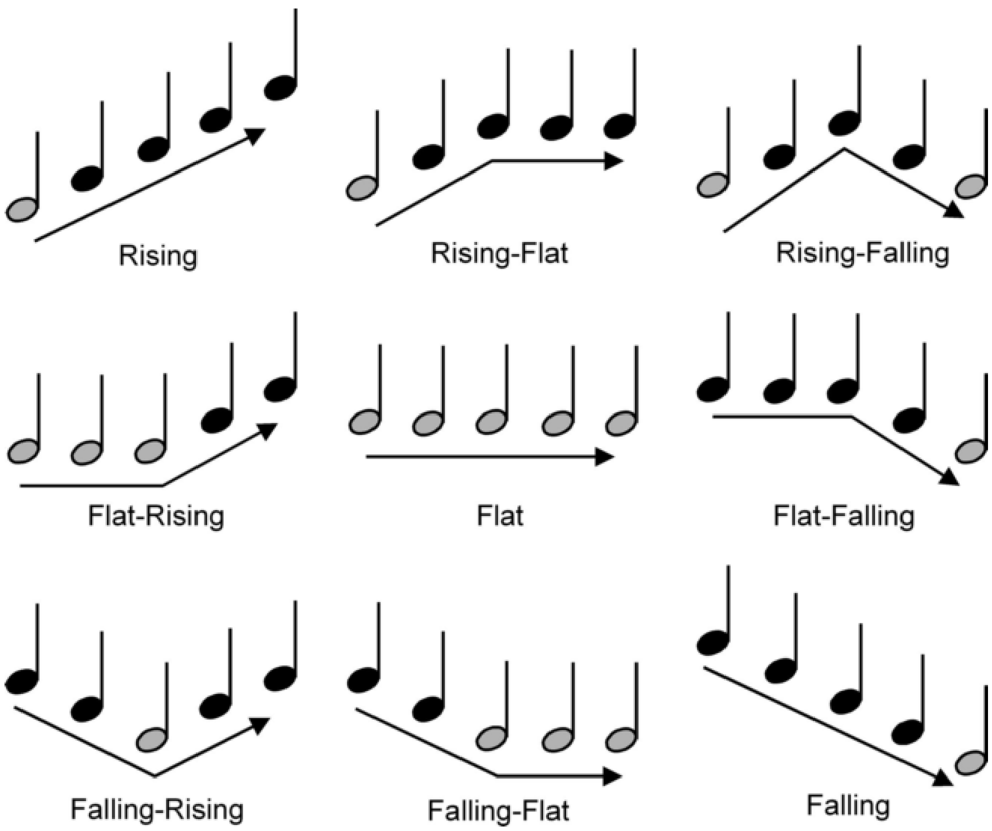


Figure 3.1. The 9 melodic contours used in the Melodic Contour Training Program. Root notes are marked in grey. From Galvin et al. (2007).

The only difference between the two programs was how difficulty was controlled. In the MCTP (Interval), note duration was fixed at 250 ms, and the interval size between consecutive notes was manipulated between 1 and 7 semitones that increased or decreased by 1 semitone. For example, the program would begin by presenting a randomly selected melodic contour with a 7 semitone interval between consecutive notes. If the participant correctly identified the melodic contour, the program would increment the difficulty and the next presentation would be at 6 semitones, and so on. If a melodic contour was incorrectly identified, the next presentation would be decreased in difficulty (e.g. presentation from 6 semitones to 7 semitones).

The melodic contours were created relative to a root (or lowest) note set at A4 for both programs - these are marked in light grey in Figure 3.1. The F0 range for the MCTP (Interval) was 440 to 2218 Hz, and the MCTP (Duration) was 440 to 1397. A full list of contours, frequency ranges and note durations can be found in Table 7.1 and 7.2, in Appendix A.

A melodic contour with a large interval size should stimulate a greater number of frequency channels than a smaller interval size. Thus the recipient should be able to use changes across multiple channels as their primary cue for melodic contour identification (MCI). For smaller intervals, less frequency channels are stimulated and users must attend to changes across a few channels, and may have to rely on temporal encoding on the channel as an additional (or primary) cue. In the MCTP (Duration), interval size between notes was fixed at 5 semitones, and the duration of each note was manipulated between 50 and 450 ms. Difficulty was increased by reducing the duration of the notes (7 levels: 450, 350, 250, 200, 150, 100, 50 ms). Thus as the participants perform better at MCI, the note durations become shorter, and participants have less time to process changes across a fairly wide range of channels.

The stimuli were created using MuseScore - a MIDI-based composition and notation software. MIDI was used as it is robust to the manipulation we required (i.e. modifying the durations of notes). A Yamaha Disklavier Pro soundfont (i.e. our digital instrument) was chosen specifically, as it was a fairly realistic MIDI representation of an acoustic grand piano. The YDP soundfont is freely available as a Creative Commons Attribution 3.0 license from (http://zenvoid.org/audio/).

The program has two modes: “Practice” and “Training” (see Figure 3.2 and Figure 3.3). In the Practice Mode, participants are provided with all 9 melodic contours on their screen, with 3 tabs labelled “Easier”, “Moderate” and “Harder”. In the MCTP (Interval) these tabs represented 7, 4, and 1 semitone intervals, and in the MCTP (Duration) these tabs represented 450, 250, and 50 ms respectively. Practice Mode was designed so that participants could listen (and see) all 9 melodic contours available, and listen to their differences. The main task for participants however, was to complete at least one Training session (consisting of 25 melodic contours), 4 days a week, for 6 weeks. In the training task, participants were presented with a melodic contour sound stimulus (which they could repeat), and four buttons representing one correct answer matching the presented contour, and three other options that were randomly selected from the pool of 9 contours. Feedback was provided after each response. If they were incorrect, the correct response would be highlighted, and they were then permitted (and encouraged) to listen for the differences between their selected and correct responses.

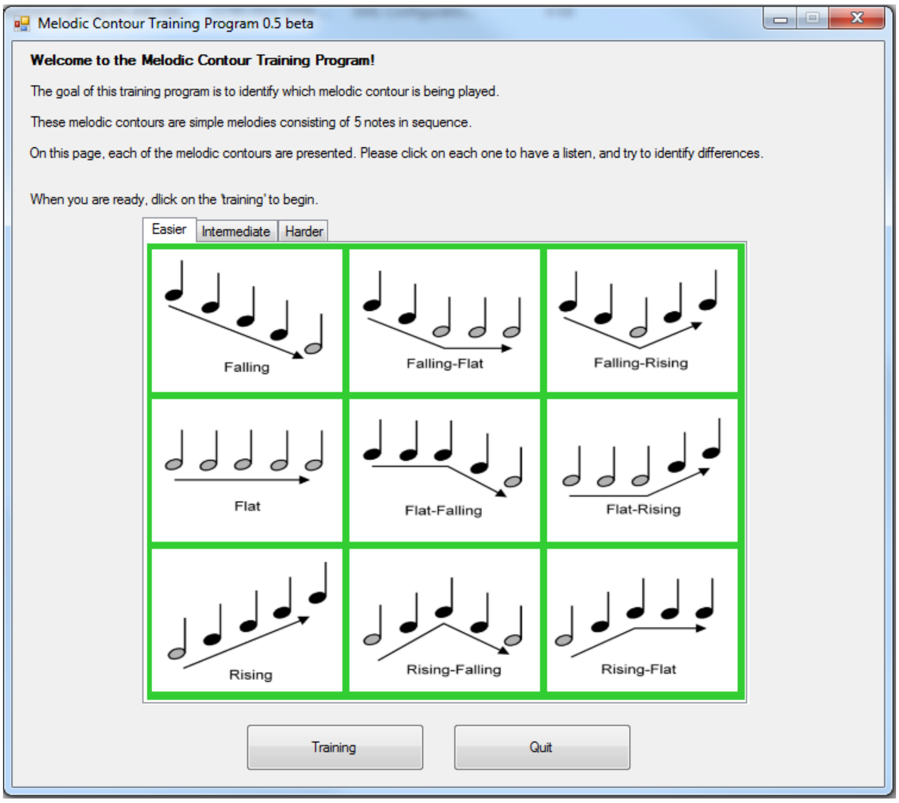


Figure 3.2. Melodic Contour Training Program in Practice Mode

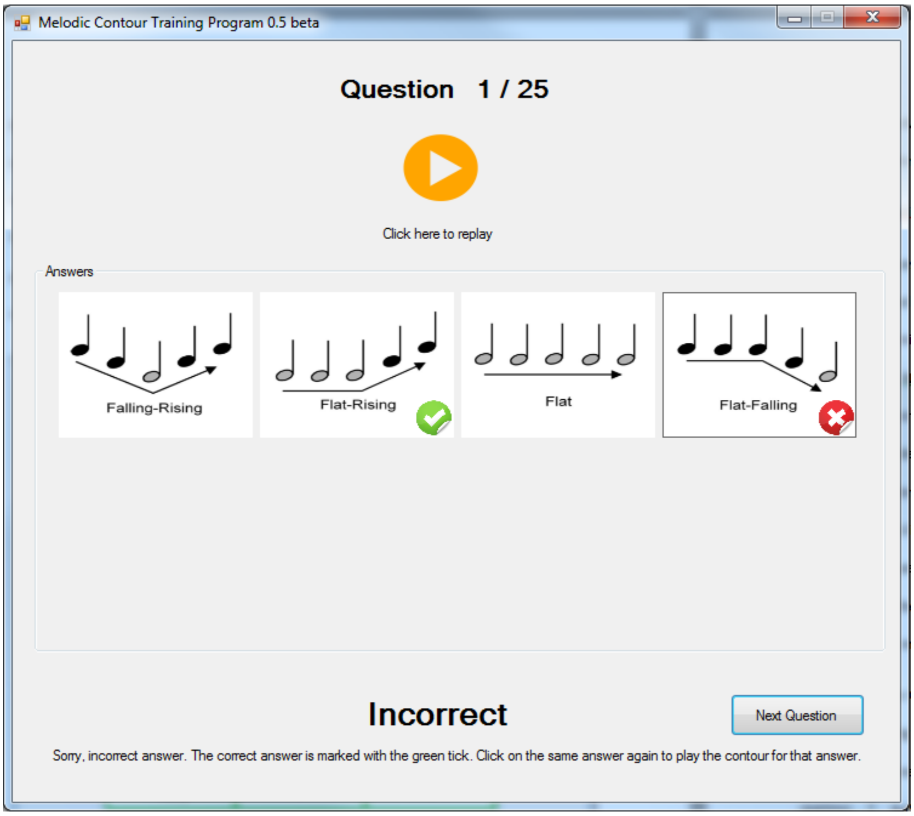


Figure 3.3. Melodic Contour Training Program in Training mode. In this example, the answer selected is incorrect. The participant is provided feedback, and encouraged to switch between the correct and incorrect responses, listening for differences.

Data logging tracked the progress of each participant's session. For the MCTP (Interval), a melodic contour interval threshold was calculated - the interval size (measured in semitones) at which 50% of contours were correctly perceived (similar to an SRT). Similarly for the MCTP (Duration), a melodic contour duration threshold was calculated - the duration of each note (measured in ms) at which 50% of contours were correctly perceived.

3.1.3 *Materials*

The Victoria Stroop Test (VST) was used as a measure of selective attention and cognitive flexibility (Regard, 1981). The Victoria version was selected specifically as it requires a short administration time. The VST consists of three tasks (naming the colour of dots, neutral words, and colour words), with each task consisting of 24 items each. The VST was administered on computer as an individual subtest of the Psychology Experiment Building Language (PEBL) Test Battery software (Mueller & Piper, 2014). In task 1, the participant was presented with on-screen instructions to select the colours of the dots as quickly as they could. Task 2 required selecting the colour of neutral words such as "when, hard, and, or over". Task 3 required selecting the colour of coloured words such as "red, blue, green, or yellow". The duration taken to complete each task was recorded and an efficiency score was calculated by taking the ratio of coloured words divided by neutral words (Task 3/Task 2).

A pitch ranking task was developed by Swanson (2008), and was used as a measure of ability to rank notes in their correct order. In each trial, a pair of notes (each 500 ms in duration) was presented consecutively with a 250 ms gap. Participants had to rank the notes by identifying if the final note was 'rising' or 'falling'. There were two conditions, differentiated by their fundamental frequency: 131 Hz (C3) harmonic tone, and 440 Hz (A4) harmonic tone. As the MCTP stimuli was centred at 440 Hz, the decision to include the 131 Hz condition was to see if any pitch ranking improvements were specific to the range they were trained at, or if they were more generalizable. There were 6 note pairs used for the 131 Hz condition (C-D, C-G, C-A, D-G, D-A, G-A), and for the 440 Hz condition (A-B, A-E, A-F#, B-E, B-F#, E-F#) all of which represent 2, 5, 7, and 9 semitone differences. The ranking task had 6 note pairs, and 8 trials per pair, for a total of 48 trials in each condition.

The Australian Sentences Test in Noise (AuSTIN) is an adaptive speech in noise test developed specifically for Australian CI recipients (Dawson, Hersbach, & Swanson, 2013). AuSTIN uses a speaker at a fixed level, with 4 talker babble (4TB) adaptively modified to obtain a speech reception threshold (SRT - defined as the signal to noise ratio at which 50% of words within a sentence were correctly perceived). For this study, AuSTIN was configured as a sentence list of 16 sentences randomly selected and spoken by a female speaker in the presence of time-locked 4TB, presented through one loudspeaker at 0^{o} azimuth. Compared to competing noise that is static, 4TB represents a dynamic masker that is more indicative of everyday listening and provides informational masking. By using an adaptive procedure, ceiling and floor effects that may provide low or high intelligibility scores can also be avoided (Bode & Carhart, 1974). Furthermore, AuSTIN has shown high test-retest reliability, with a standard deviation of 0.76 dB for SRTs (Dawson, Hersbach, & Swanson, 2013), making it highly suitable for use in pre- and post-test measurement. Participants were asked to repeat the sentence as best they heard, and were scored on a morpheme level (e.g. catching consists of two morphemes: catch, and -ing). In each session, two lists were completed, a speech reception threshold (SRTs, defined as the signal to noise ratio at which 50% of words were correctly perceived) was calculated by averaging the two results.

A short Consonant Discrimination Test was developed for the purposes of this study, using a set of 12 commonly used consonants that provide contrasts between voicing, manner of articulation (stops, sibilants, fricatives, and nasal), and place of articulation (bilabial, labio-dental, alveolar, and velar) contrasts. Using an initial /CV/ context, the consonants used were /pa, ta, ka, ba, da, ga, fa, va, sa, za, ma, and na/. The speech materials consisted of one male speaker, and were validated for clarity and level-balance by two professional linguists. Lists consisting of 32 consonants in random order were created in two conditions: 1. quiet, and 2. noise with 4TB (10dB SNR). Participant scores were calculated as percentage-correct scores, and errors were analysed in terms of confusion matrices. All acoustic analyses were performed using Praat Version 5.4 (Boersma & Weenink, 2014) Spectrograms for voiced stop consonants are presented in Figure 3.4., and acoustic features are listed in Table 3.2.

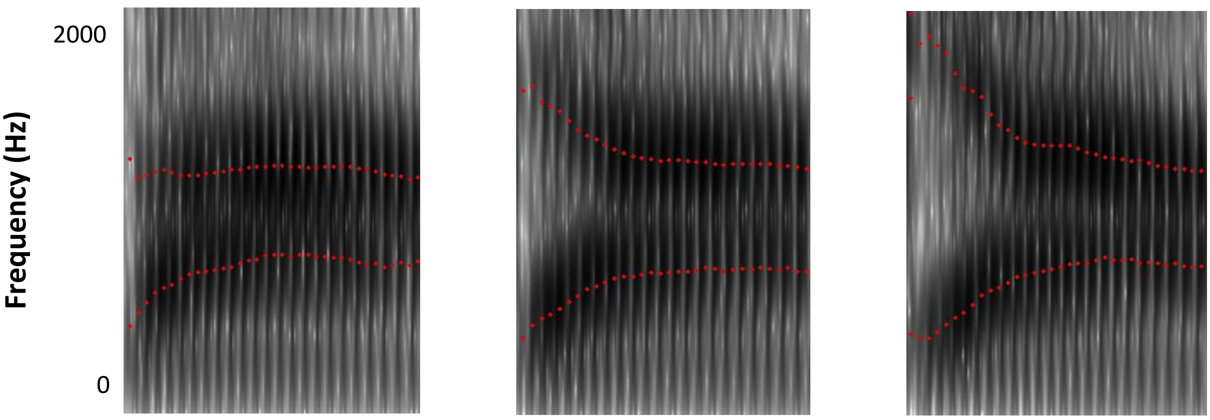


Figure 3.4. Spectrograms for voiced stop consonants /ba, da, ga/ with F1 and F2 labelled.

Table 3.2. Acoustic features for stop consonants

Consonant	Voiced/Unvoiced	VOT	Onset of F2 (Hz)	Offset of F2 (Hz)	Duration of F2 transition
/ba/	Voiced	190 ms	1140	1171	~50 ms
/da/	Voiced	180 ms	1563	1212	~75 ms
/ga/	Voiced	180 ms	1821	1308	~100 ms

The Macquarie Battery of Emotional Prosody (MBEP; Thompson et al., 2012) consists of sentences that vary in emotional prosody, consisting of 4 female and 4 male speakers uttering semantically neutral sentences such as “the room is down the hallway, and the pen is in the drawer”. Each sentence was 14 syllables in length and spoken with the intention of emoting: happy, sad, tender, irritated, afraid, or neutral. For this study, the task was simplified to 4 emotions, with 2 lists consisting of 32 happy/sad sentences, and 2 lists consisting of 32 fearful/irritated sentences. The happy/sad sentences were representative of an easier task relative to the afraid/irritated sentences, as the acoustic features are more perceptually distinct. The acoustic features are outlined in Table 3.3. Participant scores were calculated as percentages of correct responses. The speech materials are available for download from William Forde Thompson’s website (www.psy.mq.edu.au/me2).

A short questionnaire was also included (Thompson et al., 2012), in which participants were asked to self-evaluate their ability to perceive emotional prosody in day-to-day activities. Participants were asked to indicate their agreement with three statements: (1) When speaking on the telephone, I cannot tell how someone feels just by listening to their voice; (2) When talking to people, I mostly rely on their facial expressions to understand their mood and feelings; and (iii) When people are talking to me, I do not realize when they are being sarcastic. These statements were designed to determine CI recipients’ ability to interpret emotional speech prosody when it is not supplemented by facial expressions and gestures (statement 1); their reliance on facial expressions and gestures (statement 2); and their perceived ability to interpret subtle aspects of speech prosody (statement 3). Ratings were assigned on a 5-point scale with the following response options: strongly disagree, disagree, unsure, agree, and strongly agree.

Table 3.3. Acoustic features of the Macquarie Battery of Emotional Prosody

Emotion	F0 (Hz)	*SD* (Hz)	Contour changes	Slope	Duration (s)	Intensity (dB)
Happy						
M	93.44	*3.92*	8.13	5.00	2.85	73.99
Sad						
M	87.49	2.88	6.94	-11.98	3.10	68.76
Afraid						
M	93.46	1.69	7.56	-17.54	2.31	74.80
Irritated						
M	91.98	2.97	5.63	-30.15	2.43	73.76

An individual subtest (turn-end reception) was selected from the Profiling Elements of Prosody in Speech-Communication (PEPS-C) (Peppé and McCann, 2003), as a means to assess simple question and statement prosodic discrimination. Participants were presented with single word utterances such as ‘carrot’ or ‘milk’ that varied with intonation. Rising intonations indicated a question, while falling intonations indicated statements. Participants were tested on 16 single word utterances, and were scored as a percentage of correct responses. Intonation features for each statement utterance are detailed in Table 3.4, and for each question utterance in Table 3.5. Intonation curves are also presented in Figure 3.5.

Table 3.4. Acoustic features for statement intonations

Word	F0 maximum (Hz)	F0 minimum (Hz)	Intonation duration (ms)	Word duration (ms)
cheese	273	190	33	57
orange	221	183	42	64
lemon	252	182	55	56
pear	242	185	33	44
cabbage	239	159	35	65
tea	252	203	31	41
milk	242	181	29	63
raisins	230	215	25	83
MEAN	244	187	35	59

Table 3.5. Acoustic features for question intonations

Word	F0 maximum (Hz)	F0 minimum (Hz)	Intonation duration (ms)	Word duration (ms)
apple	460	232	43	52
cake	448	261	19	44
honey	428	261	33	43
jam	426	256	42	52
leeks	476	235	27	73
water	384	261	41	47
carrots	474	138	27	76
honey	396	213	37	49
MEAN	437	232	34	55

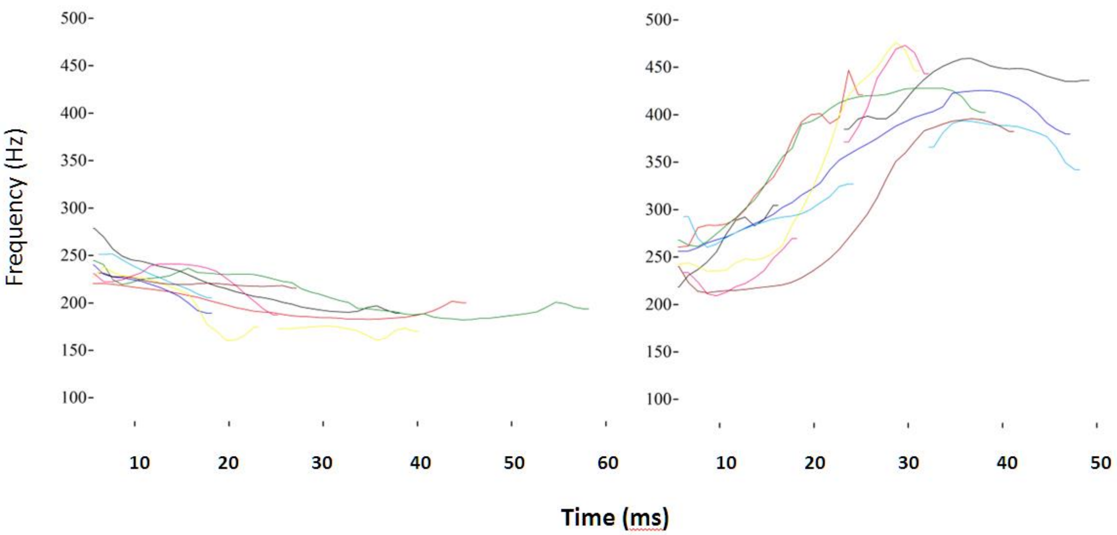


Figure 3.5. Intonation curves for PEPS-C stimuli
(statement utterances on the left, question utterances on the right)

The 12-item version of the Speech, Spatial and Qualities of Hearing Scale (SSQ12; Noble et al., 2013) was used to provide complementary, subjective analysis of hearing ability. The SSQ12 is a short-form questionnaire that uses a subset of 12-items adapted from the complete 49-item scale (SSQ). Designed for clinical purposes, it is quick to administer, providing essential information for day-to-day hearing abilities. Each item is marked on a scale of 0-10, and an overall SSQ score is calculated as the unweighted average of each item.

Finally, a short music questionnaire was administered to determine participants' musical background. Participants were asked to self-evaluate their: interest in music; general music listening habits at pre and post-implant stages; participation in formal and informal music training experiences; and ability regarding music theory, music reading, instrument playing, and singing.

3.1.4 *Procedures*

Testing occurred in an acoustically treated test booth in the Macquarie University Speech and Hearing Clinic, and in an acoustically treated room at Sydney Cochlear Implant Centre (SCIC), Gosford, NSW. The test battery was administered using a Toshiba Tecra R850 laptop computer running a Windows 7 operating system. A Yamaha Audiogram 3 USB audio interface provided the sound signal, and was connected to a Behringer Truth B3030A loudspeaker. Stimuli were presented at 65 dBA as measured with a sound level meter from the participant's listening position, located 1 metre in front of the loudspeaker. CI recipients were asked to use their regular, everyday settings and adjust their volume to a comfortable sound level on their cochlear device and hearing aid. Once set, participants were requested to refrain from modifying any settings.

Both the CI and NH groups were presented with test stimuli for both the baseline and post-training sessions in the following order: 1. VST, 2. Pitch Ranking (131Hz) 3. AuSTIN, 4. short Consonant Discrimination test in quiet, 5. MBEA (Happy/Sad), 6. Pitch Ranking (440Hz), 7. AuSTIN, 8. short Consonant Discrimination Test with 4TB, 9. MBEA (Fear/Irritation), 10. PEPS-C (turn-end reception sub-test), 11. SSQ12. As the AUSTIN was used twice, SRTs were averaged between the two tests for each session. The test battery took an average time of 1 hour to complete, with participants being offered a short break prior to the second Pitch Ranking (440Hz) task at the halfway point to avoid fatigue. Participants showed no signs of fatigue, and all declined the offer for a rest break.

Each participant was randomly assigned a program and trained for a period of 6 weeks, 4 days a week, for 30 minutes a day. Following the baseline battery, participants were randomly assigned either the Interval or Duration program for the MCTP and were provided with instructions for their training. The training required the completion of one set of the “Training Mode” (25 melodic contours - requiring approximately 15 to 30 minutes, depending on the participants’ ability), 4 days a week, for a total duration of 6 weeks. All participants were provided with a set of Edifier M1250 USB powered loudspeakers to use during their training. At the end of the post-training session, participants were asked to evaluate their experience with the MCTP.

3.1.5 *Statistical methods*

Analysis was performed with IBM SPSS Statistics version 21. Unless stated otherwise, each test was analysed using a repeated-measures analysis of variance (ANOVA), with session (baseline and post-training) as the within-subject factor, and program (interval or duration) as the between-groups factor. Additionally, the post-training scores were compared between the CI group and the NH group using independent sample *t* tests.

4 Results

Group means and statistical data have been tabulated and are presented in Table 4.1. Individual participant results for each test are provided in Appendix B.

Table 4.1. Main effects of Session, Program, and Interactions for all tests

Test	*t or F*(df)	*p*
MCTP (Interval)		
• Session	2.75 (6)	0.033*
MCTP (Duration)	3.00 (6)	0.024*
• Session		
VST		
• Session	1.20 (1, 13)	0.301
• Program	3.10 (1, 13)	0.102
• Session/Program	0.07 (1, 13)	0.796
Pitch ranking (131 Hz)		
• Session	0.08 (1, 5)	0.787
• Program	1.53 (1, 5)	0.271
• Session/Program	0.03 (1, 5)	0.861
Pitch ranking (440 Hz)		
• Session	0.08 (1, 5)	0.794
• Program	1.78 (1, 5)	0.240
• Session/Program	0.01 (1, 5)	0.936
AuSTIN		
• Session	2.46 (1, 14)	0.139
• Program	0.01 (1, 14)	0.925
• Session/Program	0.01 (1, 14)	0.914
Consonant discrimination (quiet)		
• Session	6.00 (1, 14)	0.028*
• Program	0.03 (1, 14)	0.868
• Session/Program	2.69 (1, 14)	0.123
Consonant discrimination (4TB)		
• Session	0.48 (1, 14)	0.500
• Program	0.08 (1, 14)	0.779
• Session/Program	0.62 (1, 14)	0.444
MBEP (Happy/Sad)		
• Session	3.18 (1, 14)	0.096
• Program	0.24 (1, 14)	0.634
• Session/Program	1.54 (1, 14)	0.235
MBEP (Fear/Irritation)		
• Session	1.05 (1, 14)	0.323
• Program	0.06 (1, 14)	0.814
• Session/Program	0.39 (1, 14)	0.544
PEPS-C		
• Session	9.31 (1, 14)	0.009*
• Program	0.01 (1, 14)	0.978
• Session/Program	0.90 (1, 14)	0.359
SSQ12		
• Session	0.22 (1, 14)	0.650
• Program	1.43 (1, 14)	0.252
• Session/Program	0.01 (1, 14)	0.927

*Indicates significance at alpha = 0.05

Participants were randomly assigned a training program, but to confirm there were no statistically significant differences on key variables between those assigned the Interval program compared with the Duration program, independent sample *t*-tests were calculated across: age, CI experience, musical experience (interest in music, formal training, and informal training), and all baseline scores (VST, AuSTIN, Consonant perception (quiet and 4TB), MBEP (Happy/Sad, and Fear/Irritation), and PEPS-C). There were no statistically significant differences found, and the two groups can be considered broadly equivalent prior to the training program. Results for individual tests are as follows.

Melodic Contour Training Program (Interval) - Figure 4.1 shows the mean interval threshold (semitones) for each week of training. Interval threshold was calculated with paired *t*-tests. The post-training session threshold (M = 1.7 ± 1.2 semitones) was significantly better compared with baseline (M = 2.5 ± 1.7 semitones), $t(6) = 2.75$, $p = 0.033$. These results indicate that CI recipients were able to identify melodic contours with smaller interval sizes at post-training than at baseline. Participant 1 did not complete all training sessions, completing 4 weeks (out of the full 6 required), but did train with extra sessions in the weeks she did train; unfortunately the data logs were corrupted and thus unavailable for participant 9; participant 12 was an outlier in terms of interval threshold, but also showed the greatest improvement.

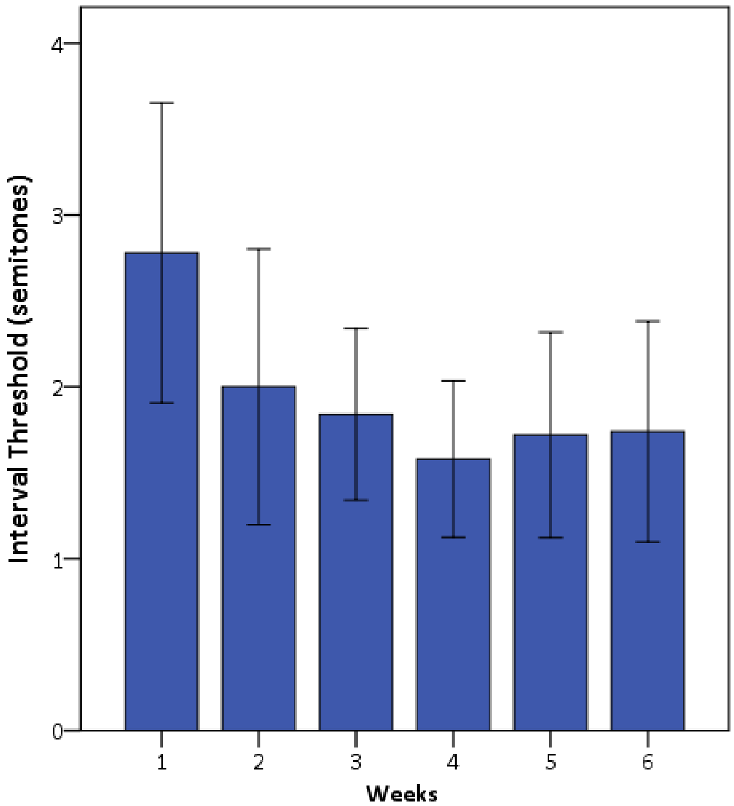


Figure 4.1. Week-to-week interval threshold scores for the Melodic Contour Training Program (Interval group). Error bars indicate 1 standard error.

Melodic Contour Training Program (Duration) - Figure 4.2 shows the mean duration threshold (ms) for each week of training. Duration threshold was calculated with paired *t*-tests. The post-training session threshold (M = 72 ± 13 ms) was significantly better compared with baseline (M = 108 ± 39 ms), $t(6) = 3.00$, $p = 0.024$. These results indicate that CI recipients were able to identify melodic contours with shorter note durations at post-training than at baseline. Ceiling performance was observed in 3 participants; participants 5 and 10 did not complete all training sessions, completing 4 and 2 weeks respectively (out of the full 6 required).

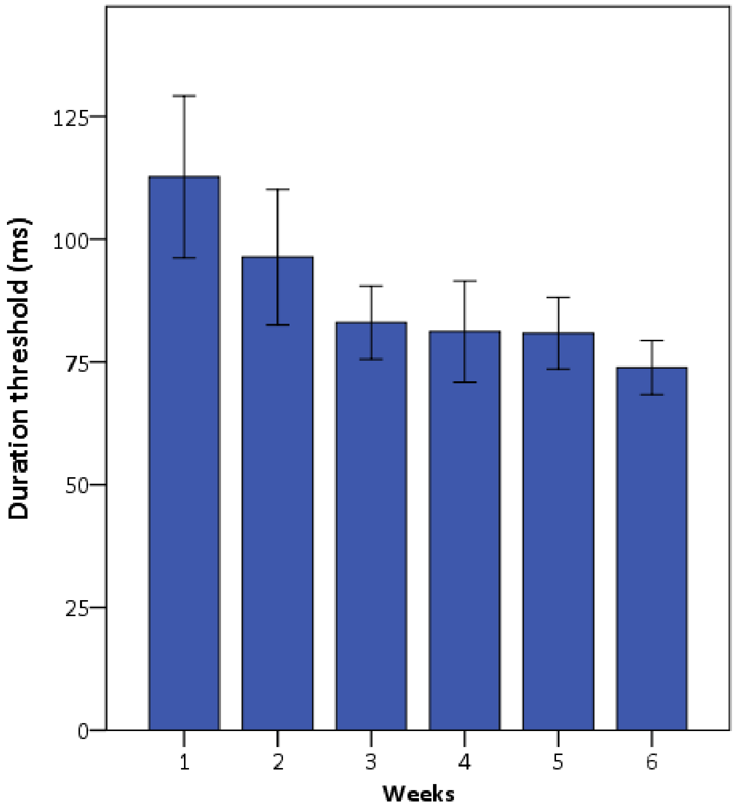


Figure 4.2. Week-to-week duration threshold scores for the Melodic Contour Training Program (Duration group). Error bars indicate 1 standard error.

Victoria Stroop Test - Figure 4.3 shows the mean scores (efficiency) for baseline and post-training on attention. There were some small improvements in efficiency for the Interval (-0.23) and Duration (-0.14) groups from baseline to post-training, but the main effect of session was non-significant [$F(1, 13) = 1.20$, $p = 0.301$], the main effect of program was non-significant [$F(1, 13) = 3.10$, $p = 0.102$], and there were no interaction effects [$F(1, 13) = 0.07$, $p = 0.796$]. Efficiency scores at the post-training session for the CI group ($M = 1.2 \pm 0.4$) compared with the NH group ($M = 1.1 \pm 0.5$) were not statistically significant, $t(25) = 0.58$, $p = 0.567$. These results indicate that the CI recipients and NH listeners were equivalent based on a broad measure of attention and executive function. Participant 4 was unable to be tested with the VST due to colour blindness and cone dystrophy.

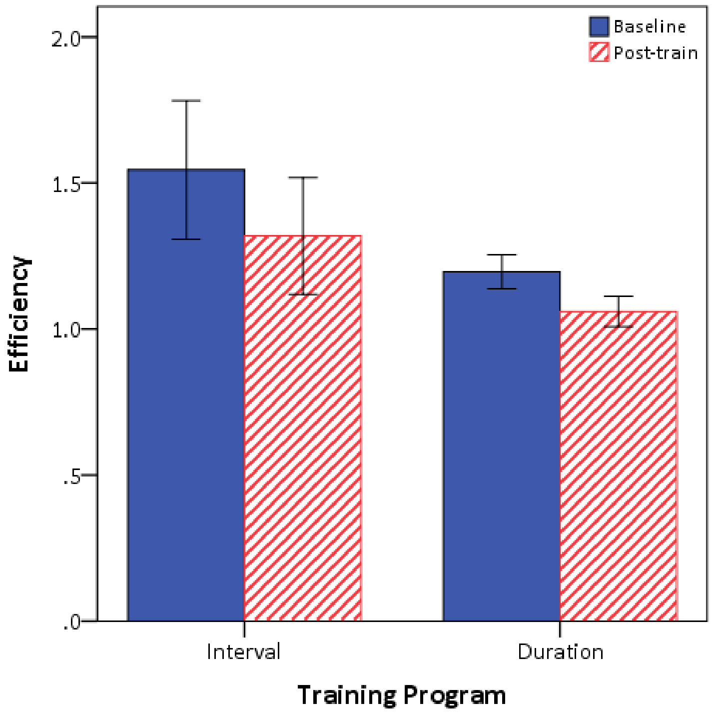


Figure 4.3. Baseline and post-training efficiency scores for the Victoria Stroop Test. Error bars indicate 1 standard error.

Pitch ranking (131 Hz) - the main effect of session was non-significant [$F(1, 5) = 0.08$, $p = 0.787$], the main effect of program was non-significant [$F(1, 5) = 1.53$, $p = 0.271$], and there were no interaction effects [$F(1, 5) = 0.03$, $p = 0.861$]. Pitch ranking (131 Hz) scores at the post-training session showed the CI group was significantly lower ($M = 74 \pm 16\%$) compared with the NH group ($M = 86 \pm 9\%$), $t(25) = -2.16$, $p = 0.041$.

Pitch ranking (440 Hz) - the main effect of session was non-significant [$F(1, 5) = 0.08$, $p = 0.794$], the main effect of program was non-significant [$F(1, 5) = 1.78$, $p = 0.24$], and there were no interaction effects [$F(1, 5) = 0.01$, $p = 0.936$]. Pitch ranking (440 Hz) scores at the post-training session showed the CI group was statistically significantly lower ($M = 71 \pm 17\%$) compared with the NH group ($M = 91 \pm 7\%$), $t(25) = -4.01$, $p = 0.001$. It should be noted that only 7 participants completed the pitch ranking task for both baseline and post-training sessions. But all participants completed the pitch ranking task at the post-training session. Figure 4.4 shows the mean scores (percent correct) for baseline and post-training on pitch ranking.

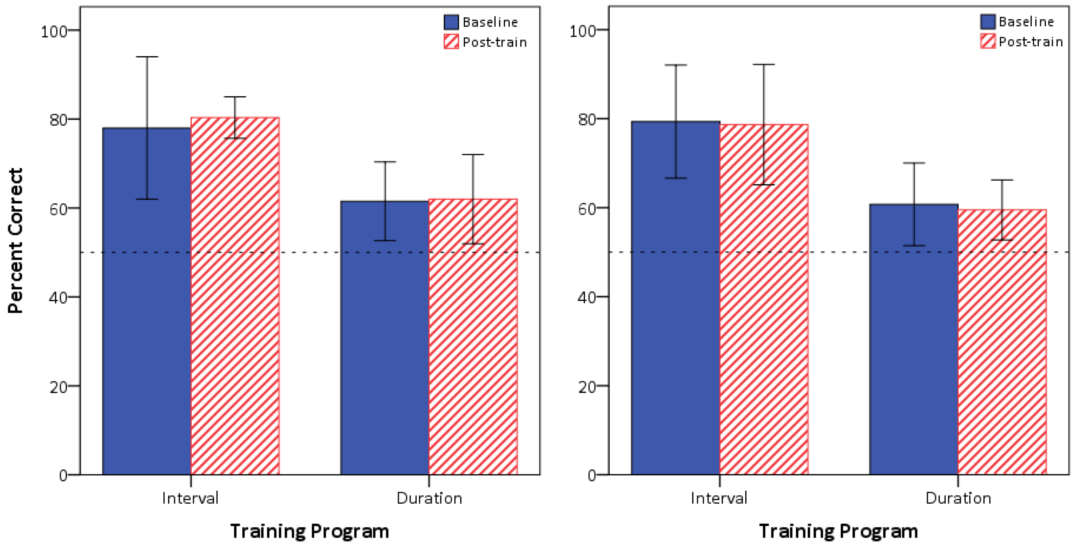


Figure 4.4. Baseline and post-training performance for the pitch ranking test (131 Hz on the left, 440 Hz on the right).The dashed line indicates the chance score.

Australian Sentence Test in Noise - Figure 4.5 shows the mean scores (SRTs) for baseline and post-training on speech-in-noise. The main effect of session was non-significant [$F(1, 14) = 2.46$, $p = 0.139$], the main effect of program was non-significant [$F(1, 14) = 0.01$, $p = 0.925$], and there were no interaction effects [$F(1, 14) = 0.01$, $p = 0.914$]. SRT scores at the post-training session showed the CI group was significantly higher ($M = 4.4 \pm 2.2$ dB) compared with the NH group ($M = -4 \pm 0.9$ dB), $t(25) = 11.85$, $p < 0.001$.

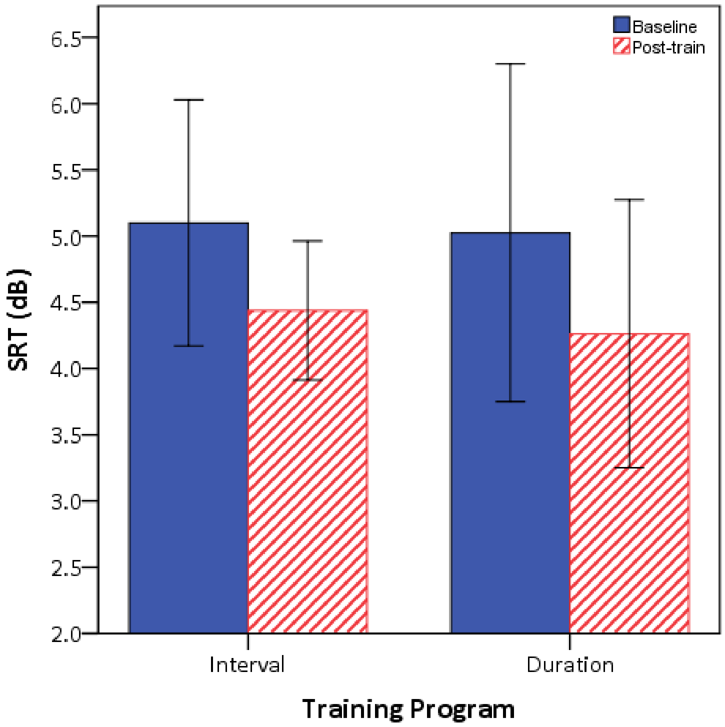


Figure 4.5.Baseline and post-training SRTs for AuSTIN.

Consonant discrimination in quiet - the main effect of session was statistically significant [$F(1, 14) = 6.00$, $p = 0.028$], the main effect of program was non-significant [$F(1, 14) = 0.03$, $p = 0.868$], and there were no interaction effects [$F(1, 14) = 2.69$, $p = 0.123$]. Consonant scores in quiet at the post-training session showed the CI group was significantly lower ($M = 87 \pm 15\%$) compared with the NH group, with all NH individuals performing at ceiling ($M = 100\%$), $t(26) = -3.58$, $p = 0.003$.

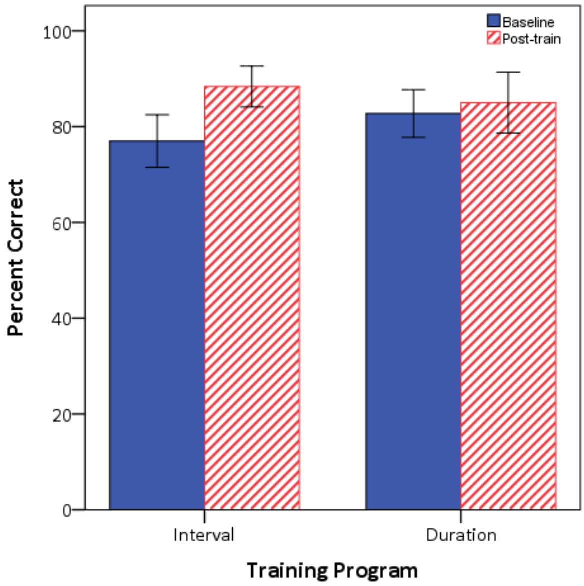


Figure 4.6. Baseline and post-training performance for consonant discrimination in quiet.

Further analysis using confusion matrices for individual consonants reveals place of articulation was most improved for both training programs. To reconcile the analysis, only confusions greater than 10% (5 or more confusions) at baseline were considered.

In the Interval group, analysis of individual consonants showed large improvements in the perception of stop consonants in which a 30% increase was observed for /p/, and a 23% increase for /d/, and in the nasal stop an increase of 33% was observed for /n/. A large reduction was observed for stop consonants, in which a 23% decrease was observed for /p/ perceived as /k/, and a 13% decrease was observed for /d/ perceived as /g/, in fricatives a 13% decrease was observed for /s/ perceived as /z/, and in the nasal stop an 18% decrease was observed for /m/ perceived as /n/. Pooled confusion matrices at baseline and post-training for the Interval group are presented in Figure 4.7.

	Response												
Stimulus	p	t	k	b	d	g	f	v	s	z	m	n	-
p	**23**		11			1							5
t		**40**											
k			**30**	1			1						8
b				**37**	2		1						
d			1	4	**22**	12		1					
g			1		4	**35**							
f	2			1			**34**		1	1			1
v				5				**28**			3		4
s							3		**29**	8			
z						3		1		**36**			
m											**31**	3	6
n						3					9	**25**	3

	Response												
Stimulus	p	t	k	b	d	g	f	v	s	z	m	n	-
p	**35**	3	2										
t		**35**	5										
k			**36**										4
b				**34**	2			4					
d				2	**31**	7							
g				1		**39**							
f	1			3			**36**						
v		1		1				**32**		2	2		2
s							2		**35**	3			
z								1		**39**			
m											**35**	2	3
n											2	**38**	

Figure 4.7. Confusion matrix for Interval group. Baseline is on top, post-train on the bottom. Significant confusions in baseline have been marked in grey, and this is carried over to post-train matrix for easier visual identification of confusion decreases.

In the Duration group, analysis of individual consonants showed large improvements in the perception of stop consonants in which a 25% increase was observed for /p/, and in the fricative a 25% increase was observed for /v/, and in the nasal a 33% increase was observed for /n/. A large reduction was observed for stop consonants, in which a 13%

decrease was observed for /g/ perceived as /d/, in fricatives an 18% decrease was observed for /v/ perceived as /m/, and in the nasal stop a 13% decrease was observed for /m/ perceived as /n/. Pooled confusion matrices at baseline and post-training for the Duration group are presented in Figure 4.8.

Stimulus \ Response	p	t	k	b	d	g	f	v	s	z	m	n	-
p	**18**	3	9		1	1	1						7
t		**39**	1										
k	5		**35**										
b	1			**33**	3			2		1			
d					**32**	8							
g			1	2	5	**31**					1		
f	1		1				**34**	1	3				
v			1		1		2	**22**		7	7		
s									**38**	2			
z										**40**			
m											**26**	11	3
n											2	**38**	

Stimulus \ Response	p	t	k	b	d	g	f	v	s	z	m	n	-
p	**28**	1	6										5
t		**39**	1										
k	6		**30**										4
b		1		**34**	2	1		2					
d				1	**34**	5							
g				5	0	**35**							
f	1			2		1	**30**	2	4				
v							1	**35**		4	0		
s							4		**36**				
z								4		**36**			
m		1									**32**	6	1
n											1	**39**	

Figure 4.8. Confusion matrix for Duration group. Baseline is on top, post-train on the bottom. Significant confusions in baseline have been marked in grey, and this is carried over to post-train matrix for easier visual identification of confusion decreases.

Consonant discrimination with 4TB - the main effect of session was non-significant [$F(1, 13) = 0.48$, $p = 0.500$], the main effect of program was non-significant [$F(1, 13) = 0.08$, $p = 0.779$], and there were no interaction effects [$F(1, 13) = 0.62$, $p = 0.444$]. Consonant scores with 4TB at the post-training session showed the CI group was significantly lower ($M = 63 \pm 16\%$) compared with the NH group, with all NH individuals near ceiling performance ($M = 99 \pm 1\%$), $t(26) = -9.08$, $p < 0.001$. In the baseline session, participant 12 did not complete the task - citing difficulty perceiving any consonants in noise. However, in the post-training session after completion of training, the participant was able to complete the task, scoring 57% consonants correct. Figure 4.9 shows the mean scores (percent correct) for baseline and post-training on consonant perception.

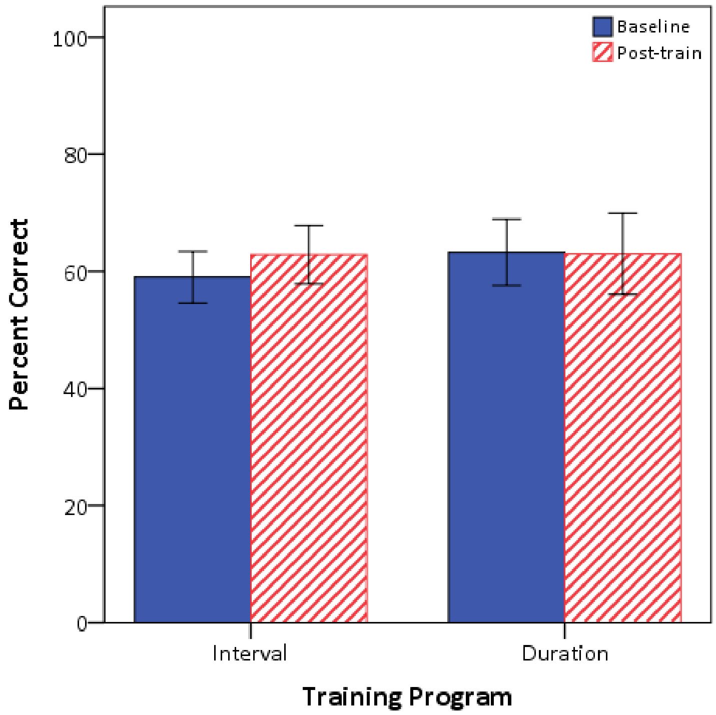


Figure 4.9.Baseline and post-training performance for the consonant discrimination with 4TB.

Macquarie Battery of Emotional Prosody (Happy/Sad) - the main effect of session was non-significant [$F(1, 14) = 3.18$, $p = 0.096$], the main effect of program was non-significant [$F(1, 14) = 0.24$, $p = 0.634$], and there were no interaction effects [$F(1, 14) = 1.54$,

p = 0.235]. Emotional prosody scores at the post-training session showed there was a statistical significance between the CI group (M = 95 ± 3%) compared with the NH group (M = 99% ± 1%), but this difference was negligible with the majority of recipients near ceiling performance, $t(26) = -3.80$, $p = 0.001$.

Macquarie Battery of Emotional Prosody (Fear/Irritation) - the main effect of session was non-significant [$F(1, 14) = 1.05$, $p = 0.323$], the main effect of program was non-significant [$F(1, 14) = 0.06$, $p = 0.814$], and there were no interaction effects [$F(1, 14) = 0.39$, $p = 0.544$]. Emotional prosody scores at the post-training session showed the CI group was significantly lower (M = 78 ± 12%) compared with the NH group, with all NH individuals performing at ceiling (M = 100%), $t(26) = -6.15$, $p < 0.001$. Figure 4.10 shows the mean scores (percent correct) for baseline and post-training on emotional prosody.

Results from the questionnaire on participants' self-evaluated ability to perceive emotional prosody indicated the CI recipients and NH group differed significantly for all three statements posed. Firstly, mean ratings for statement 1 were significantly higher for the CI group (M = 2.8 ± 1.3) than for the NH group (M = 1.8 ± 0.9), t(26) = 2.48, $p = 0.020$. These results indicate that the CI recipients have more difficulty using auditory-only cues to accurately perceive emotion than NH listeners. Mean ratings for statement 2 were significantly higher for the CI group (M = 3.6 ± 1.3) than for the NH group (M = 2.3 ± 1.1), $t(26) = 2.66$, $p = 0.01$, indicating that the CI recipients are more reliant on facial expressions and gestures to perceive emotion than NH listeners. This reliance on non-auditory cues would be expected among individuals who have reduced capacity to perceive emotion from speech prosody. Finally, mean ratings for statement 3 were significantly higher for the CI group (M = 2.3 ± 1.2) than for the NH group (M = 1.6 ± 0.5), $t(26) = 2.271$, $p = 0.033$, indicating that CI recipients have more difficulty than NH listeners when interpreting more subtle aspects of emotional prosody.

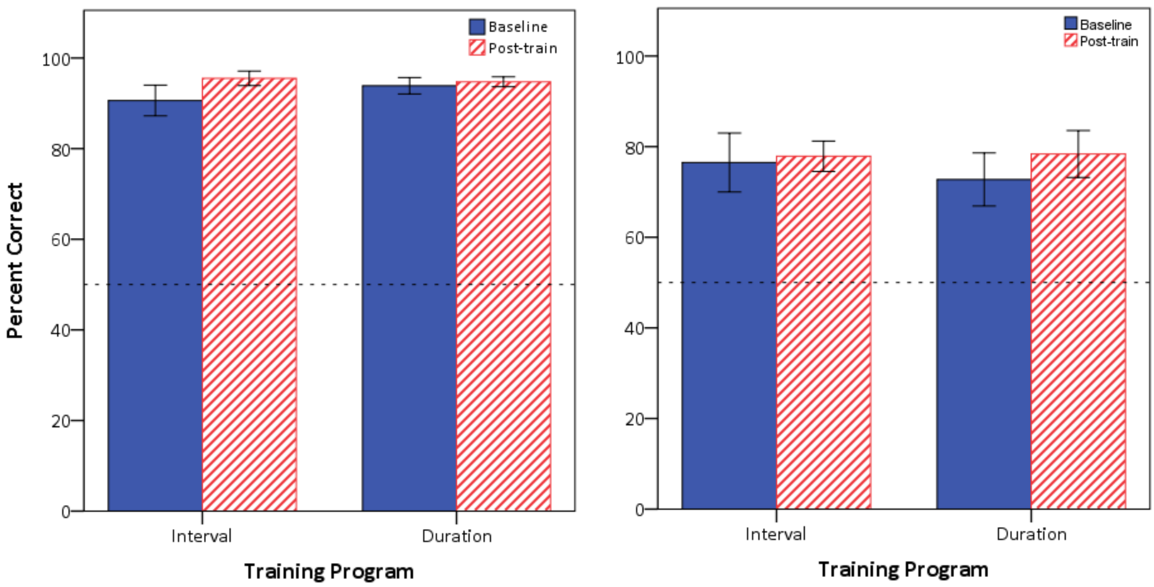


Figure 4.10. Baseline and post-training performance for the MBEP (Happy/Sad condition on the left and Fear/Irritation condition on the right). The dashed line indicates the chance score.

Profiling Elements of Prosody in Speech-Communication (Turn-end reception) - Figure 4.11 shows the mean scores (percent correct) for baseline and post-training on question/statement prosody. The main effect of session was statistically significant [$F(1, 14) = 9.31$, $p = 0.009$], the main effect of program was non-significant [$F(1, 14) = 0.01$, $p = 0.978$], and there were no interaction effects [$F(1, 14) = 0.90$, $p = 0.359$]. Prosody scores at the post-training session showed the CI group was significantly lower (84 ± 18%) compared with the NH group, with all NH individuals performing at ceiling (100% accuracy), $t(26) = -3.42$, $p = 0.004$. These results indicate a significant post-training improvement for prosody perception using intonation cues.

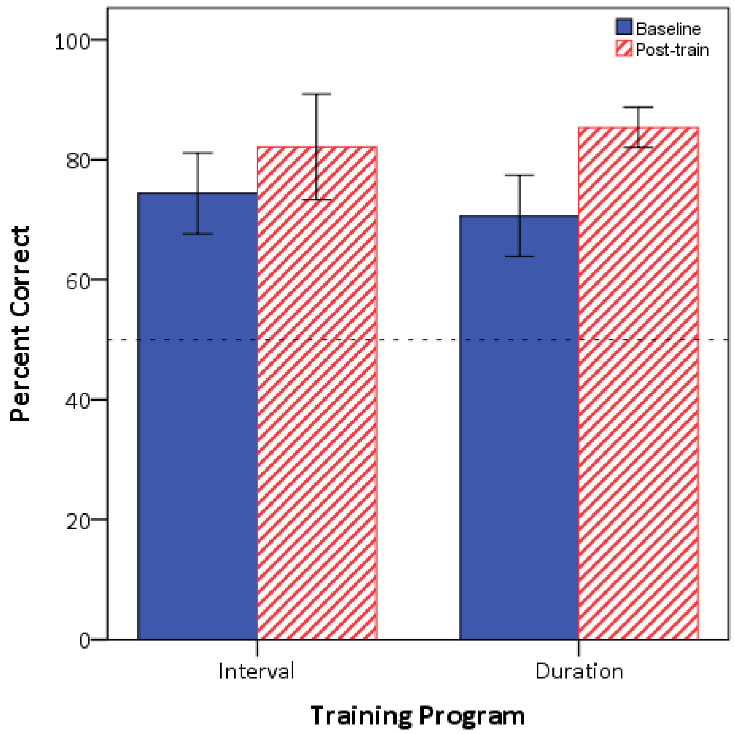


Figure 4.11. Baseline and post-training performance for the PEPS-C (Turn-end reception task). The dashed line indicates the chance score.

Speech, Spatial and Qualities of Hearing Scale 12 - Figure 4.12 shows the mean SSQ scores for baseline and post-training on self-evaluated general hearing ability. The main effect of session was non-significant [$F(1, 14) = 0.21$, $p = 0.650$], the main effect of program was non-significant [$F(1, 14) = 1.43$, $p = 0.252$], and there were no interaction effects [$F(1, 14) = 0.01$, $p = 0.927$]. SSQ scores at the post-training session showed the CI group was significantly lower (5.1 ± 2.1) compared with the NH group (8.6 ± 0.7), $t(26) = -6.26$, $p < 0.01$.

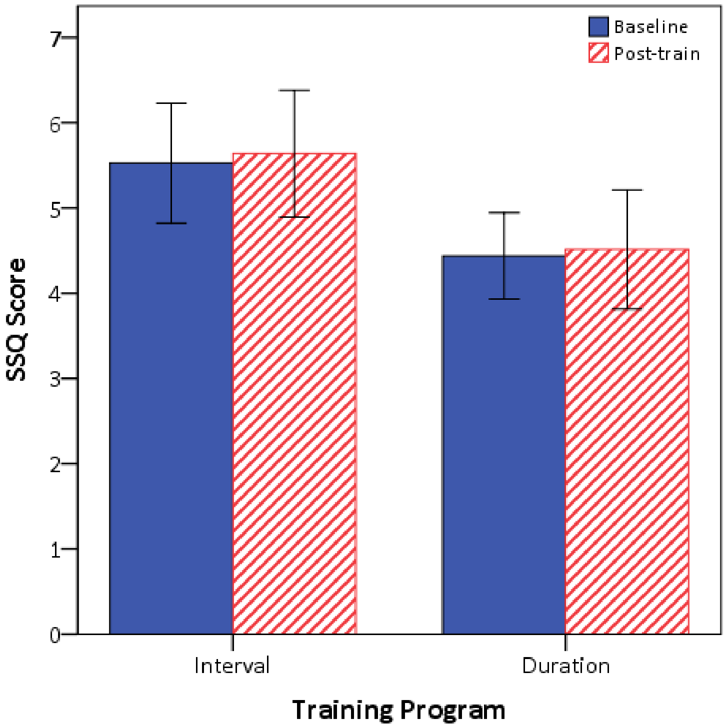


Figure 4.12. Baseline and post-training performance for the SSQ12.

5 Discussion and conclusion

The results indicate that melodic contour training can significantly improve some, but not all, aspects of speech perception in CI recipients. The main findings from this study can be summarised as follows. Firstly, while all tests measured group mean effects comparing baseline to post-train performance and efficacy of each training program, as with many studies involving CI participants, there was significant variation in performance between individuals. Second, the data logged results from CI recipients indicate that MCI performance was significantly improved after six-weeks of training in both Interval and Duration programs. Third, for all tests, there was no significant effect for the type of program assigned to each participant. Fourth, there were significant improvements for the perception of consonants in quiet, and for the identification of questions and statements using only speech intonation cues. Fifth, CI recipients performed poorer than NH listeners in all tasks except attention. Finally, there were no significant group gains for attention, SIN perception, consonant perception in 4TB, or emotional prosody.

While there were overall group improvements for both training programs, there was also considerable variation among individual CI recipients. The program had two tasks: ‘practice’ and ‘training’, but data logging was only taken for the training task. As such, the week-to-week improvements can only be interpreted broadly, as it is impossible to determine how much practice an individual completed. Additionally, participants were informed to do at least 4 training sessions a week, but were not discouraged from doing more. Nonetheless, irrespective of the rate of improvement, there were significant gains from baseline to post-train.

On all tests of speech perception and attention there was no statistical difference between either training programs. These findings indicate that CI recipients were able to improve their pitch perception and temporal processing abilities in the context of melodic contour identification. While the relative efficacy between both mechanisms of interval size and note duration was non-significant, comparisons were also difficult to make due to the small sample size, resulting in a lack of statistical power.

As the two training programs used significantly different musical mechanisms, it was surprising that the improvement in consonant perception in quiet for both groups had similar patterns. In particular, confusions between place of articulation cues in voiced, unvoiced, and nasal stops were the most reduced, despite these cues being typified as the poorest speech production feature (McKay & McDermott, 1993).

While the VOT distinction between the voiced /b/ and unvoiced /p/ stops is typically the most salient cue, it is often used as an example of the redundant nature of speech, as other factors such as semantic and syntactic context, vowel duration, and F0 (as examples), also provide distinction between /p/ and /b/ (Patel, 2014). However, as this was a /CV/ task there was no semantic and syntactic content, and because the pattern of confusions is consistent (i.e. confusions occurred primarily between voiced stop consonants /ba, da, ga/, between unvoiced stop consonants /pa, ta, ka/, and nasal stops /ma, na/), we can also remove intensity differences due to voicing. In the case of the unvoiced stop consonants, VOT was 130 ms for /pa/, 110 ms for /ta/, and 130 ms for /ka/, and in voiced stop consonants, VOT was 190 ms for /ba/, 180 ms for /da/, and 180 ms for /ga/, it is clear a VOT distinction would only occur across the voicing cue (i.e. between /ba/ and /pa/), but this was not where the confusions occurred. Hence, the likely cue that recipients more effectively used post-training was the formant trajectory itself. Acoustic analysis on the stop consonant trajectories show frequency ranges between 1563 to 1211 Hz for /da/, and 1821 to 1308 Hz for /ga/ that when converted into semitones, equate to an approximate F2 interval of 3 semitones for /ba/, and approximately 6 semitones for /ga/. On average, the CI recipients in the Interval group had a mean post-train interval threshold less than 2 semitones (with fixed note durations of 250 ms), and the Duration group has duration threshold of about 75 ms (at a fixed note interval of 5 semitones). Acoustically, the melodic stimuli appears to be within the bounds of what is required to identify the F2 trajectory, and it is plausible that post-training, the F2 trajectory cue is more available for CI recipients. In addition, further investigation at the subcortical level may confirm these findings, as it has been observed that musicians have enhanced neural differentiation for the F2 trajectories in /ba, da, ga/, most likely the result of years of musical practice that enhance fine spectral resolution (Parberry-Clark et al., 2012).

Vandali (2001) investigated the perceptual benefits of the transient emphasis spectral maxima (TESM), a speech processing strategy that was designed to improve short durational, low intensity cues which are typically encoded poorly for CI recipients. As such, signals such as fricatives and stop consonants (particularly in the noise burst) would be targeted and amplified. Although monosyllabic consonant-nucleus-consonant (CNC) word lists were used for that study, the pattern of confusions is strikingly similar to the present findings for consonant perception in quiet. Vandali (2001) found that the enhancement on short duration, low intensity cues led to a decrease in confusions based on place of articulation errors, reducing confusions of /m/ for /n/ by 12.5%, and /p/ for /k/ by 7.3%. On combined Interval and Duration pooled scores, our results are 15% and 30% respectively. A significant limitation in our study is comparatively few tokens, with only 40 for each consonant per training group, such that outliers can easily skew our patterns, whereas the

Vandali (2001) study had approximately 200 tokens for each consonant, allowing for the robust removal of outliers.

Both groups also showed significant improvement for the question-statement task that required cues of speech intonation. Firstly, it must be noted that the stimuli were single words consisting of one or two syllables, and the intonation pattern occurred over the final (or only) syllable. As such, there were no syntactic or semantic cues available and the improvement from training is most likely due to the mechanism of enhanced F0 tracking, although it is possible recipients also used duration and intensity cues across syllable boundaries as a distinction. Additionally, as question-utterances rarely consist of just one word, the applicability of this enhancement to a more realistic question-statement identification task such as with sentences, or in adverse conditions, is limited.

Surprisingly, improvement with melodic contour identification in either training group did not also result in better pitch ranking discrimination. It is likely that melodic contour training confers improvements in the very specific domain of pitch tracking over continuous (i.e. melodic), rather than discrete levels. However, the sample for baseline and post-train pitch ranking task was small with only 6 CI recipients, so we cannot generalise this finding. Acoustically, the mean range for the rising intonation was 232 to 437 Hz, or approximately 11 semitones (or close an octave), well within the acoustic parameters of both Interval and Duration melodic contour training stimuli.

In a training program to improve second-language learning for NH English speakers, Song et al. (2008) used embedded F0 contours (Mandarin tones of high-level, rising, and falling-rising) in the syllable /mi/ as the stimuli for investigation of pitch tracking enhancement. Examination of the auditory brainstem responses to the three tones found significant pitch tracking improvement, but only for the unfamiliar falling-rising tone. As the CI recipients in this study were postlingually implanted, it would be assumed that a rising intonation should be a familiar response. But as there were significant improvements for the intonation task, it is likely their degraded spectral cues limit access to the intonation curve. As such, the melodic contour training provides an effective form of rehabilitation, and such results could be confirmed with a similar paradigm as used by Song et al. (2008). Additionally, paediatric CI recipients are often reported as having poor intonation and recognition of melodies (See et al., 2013), and melodic contour training may be especially effective and beneficial for habilitation purposes.

There was no improvement on emotional prosody tasks, though the happy/sad condition was possibly limited to due to ceiling effects observed for almost half of all participants. Another explanation for the non-significant findings is the trained stimuli itself. Assuming a syntactic structure between speech and music holds (i.e. a music note can be classified as corresponding to a phoneme); as the emotional prosody task required interpretation of a whole sentence, a simple 5-note melodic contour may not be sufficient. Additionally, unlike the intonation task, emotional prosody required decoding of intonation, stress and intensity cues, to which our results only support improvement on the cue of intonation. More complex melodic contour training using combined 5-note sequences into phrase structures, such as “Rising”, followed immediately with “Falling-Rising” and then “Flat”, with varying pitch intervals and note durations within each sequence may provide a more effective training paradigm.

Based on preliminary results by Patel (2014) that indicated the possibility of improvement for speech in noise perception as a result of melodic contour training (albeit with a sensorimotor component through the act of using a piano keyboard), similar gains were anticipate for the current study, but our findings indicate that as a group, there was no significant improvement for consonant perception in noise, or with the AuSTIN. Despite this, certain individuals (Participants 2 and 10), showed large improvement in SRTs, although they were both bimodal listeners using a contralateral HA. This may suggest that HA users, with more access to acoustic, F0, and fine-structure cues, find melodic contour training particularly effective for speech in noise improvement, but warrants further investigation.

While our findings indicate some level of F0 improvement, primarily for intonation, such enhancement is only accessible in quiet, indicating that maskers significantly disrupt F0 cues for CI recipients that only have access to gross temporal envelope. Effective speech in noise perception is also reliant on auditory stream segregation processes to perceptually group and separate multiple sources (Bregman, 1990). As the melodic contours were a single-stream melody, it is unlikely that it would confer any benefit for segregation tasks. This remains a key area for the improvement of speech in noise perception for CI recipients, but early indications by Zhu et al. (2011) suggest segregation will remain a difficulty until spectral resolution is improved.

There was no change in attentional processing attributable to the training. The VST was chosen specifically for its quick administration time, and it was thought that a visual based task may reduce the confounding variable of hearing loss with an auditory test of attention. However, there is evidence that attention is domain-specific between vision and audition (Woodruff, 1996). While the VST may be useful as a broad measure of attention and executive function, tests of WM capacity and auditory tests of attention may be more suitable for future studies with auditory stimuli, and provide evidence of top-down processing.

Finally, the subjective measures of improvement using the SSQ12 did not provide evidence of overall day-to-day improvement, but as our results have indicated, the gains from the training have been limited to consonant perception in quiet and speech intonation. The SSQ12 also places an emphasis on stream-segregation tasks, limiting its usefulness for our test battery. Unsolicited correspondence from several participants after completion of the program indicated improvements that were not necessarily captured with our test battery.

> "One thing the study appears to have done is improve my concentration as I am hearing [my wife] clearer through the Cochlear - even have to ask that she speaks quieter."

> "I find that as I concentrate, I have better listening performance and have had no trouble listening to it with environmental background noises. I still make some mistakes though, but am now quicker to pick up different tones when listening to music. I've noticed I'm also getting good at distinguishing between different instruments that I haven't picked up before, like a guitar string in the background."

The overall response to the training program was positive. Participants found the task engaging, as evidenced by high-compliance. The decision to use an adaptive program was motivated by the advantage that it avoids ceiling and floor effects, pushing the perceptual boundaries in small steps, and adjusting its difficulty according to the participant's performance. But this system was interpreted very differently by participants, and future changes to the program will focus on providing more feedback and incentive to train.

"I have been persevering with the training and I guess there has been some improvement. I am a bit disappointed that I practically never get them all right. I think there have been one or two times that I have."

"It is very difficult but the program gives me some easier ones when I have made a few mistakes!"

This study was limited by a small sample that reduces our ability to evaluate the specific differences between the two melodic contour training programs that varied along the mechanisms of interval size and note duration. Interpretation of cues is also made difficult without objective measures as complimentary evidence. The CI and NH groups were also not age matched, and it would be of interest to see if melodic contour training may improve older NH listener's speech perception in noise. The test battery was fairly broad, but more robust baseline measures should be adopted ensuring stable asymptotic performance prior to training, such as introducing two or more spaced sessions prior to the training. Additionally, a follow-up study is planned that will help determine if the benefits of training are maintained.

The test battery was broad, as the potential benefits for all speech tasks was open to exploration. The findings suggest that perception of transition cues between consonants and intonation cues relevant for prosody derive the most advantage from melodic contour training. The consonant word list was also limited to one male speaker, and the prosody test with one female speaker; a more robust test will be incorporated in the future, with multiple speakers and tokens. Further confirmation of sensitivity to transition cues could also be elicited with CNC word lists, examining transitions in both initial and final contexts, as well as in sentences. As the melodic contours train pitch movement across time, it is hypothesised that if the improvement conferred by melodic contour training is specific to VOT and F2 trajectory, improvements for vowels which are indicated by relatively static F1 and F2 targets should be reduced. Synthesised stop consonants with varying VOT and F2 trajectories may also be particularly insightful.

In summary, both melodic contour training programs have had a beneficial outcome for CI recipients to perceive consonants in quiet, with a specific advantage conferred for voice-onset time and perception of second formant trajectories. Speech intonation cues have also been enhanced; the specific advantage was for continuous pitch tracking, as no improvement was shown for discrete pitch ranking. Future studies incorporating objective neural responses may further substantiate these findings.

6 References

American Speech-Language-Hearing Association. (2004). Cochlear implants [Technical Report]. Available from www.asha.org.

Assmann, P., & Summerfield, Q. (2004). The Perception of Speech Under Adverse Conditions. In S. Greenberg, A. N. Popper, W. A. Ainsworth, & R. R. Fay (eds) *Speech Processing in the Auditory System, New York: Springer.*

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612-637.

Bennett, A. (2000). *Popular music and youth culture: music, identity and place*. Macmillan Press Ltd.

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of Training between Music and Speech: Common Processing, Attention, and Memory. *Frontiers in Psychology*, *2*(May), 1-12.

Bird, J., Darwin, C. J. (1998) Effects of a difference in fundamental frequency in separating two sentences. In A. Palmer, A. Rees, Q. Summerfield, R. Meddis (eds) *Psychophysical and physiological advances in hearing*. London: Whurr Publishers.

Blamey, P. (1996). Factors Affecting Auditory Performance of Postlinguistically Deaf Adults Using Cochlear Implants. *Audiology & neurotology, 1*(5), 293-306.

Boersma, P., & Weenink, D. (2014). *Praat: doing phonetics by computer* [Computer program]. Version 5.4, retrieved 4 October 2014 from http://www.praat.org/

Boothroyd, A. (2007). Adult aural rehabilitation: what is it and does it work? *Trends in Amplification, 11*(2), 63-71.

Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound.* Massachusetts: The MIT Press.

Busby, P. A., Whitford, L. A., Blamey, P. J., Richardson, L. M., and Clark, G. M. (1994). Pitch perception for different modes of stimulation using the Cochlear multiple-electrode prosthesis. *The Journal of the Acoustical Society of America, 95*(5), 2658-2669.

Cooper, W. B., Tobey, E., & Loizou, P. C. (2008). Music perception by cochlear implant and normal hearing listeners as measured by the Montreal Battery for Evaluation of Amusia. *Ear and Hearing, 29*(4), 618-626.

Cox, F. (2012). *Australian English Pronunciation and Transcription*. Victoria: Cambridge University Press.

Crew, J. D., Galvin III, J. J., & Fu, Q. J. (2012). Channel interaction limits melodic pitch perception in simulated cochlear implants. *The Journal of the Acoustical Society of America, 132*(5), EL429-EL435.

Cruttenden, A. (1997). *Intonation*. Cambridge University Press.

Cullington, H., & Zeng, F. (2011). Comparison of bimodal and bilateral cochlear implant users on speech recognition with competing talker, music perception, affective prosody discrimination. *Ear and Hearing*, *32*(1), 16-30.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: a literature review. *Language and Speech*, *40 (2)*, 141-201.

Dalton, P., Santangelo, V., & Spence, C. (2009). The role of working memory in auditory selective attention. *Quarterly Journal of Experimental Psychology (2006), 62*(11), 2126-2132.

Donaldson, G. S., Rogers, C. L., Cardenas, E. S., Russell, B. a, & Hanna, N. H. (2013). Vowel identification by cochlear implant users: contributions of static and dynamic spectral cues. *The Journal of the Acoustical Society of America*, *134*(4), 3021-3028.

Dorman, M. F. (1993). Speech perception by adults. In R. S. Tyler (ed.) *Cochlear implants. Audiological foundations*. San Diego: Singular.

Dorman, M. F., Loizou, P. C., Fitzke, J., & Tu, Z. (1998). The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels. *The Journal of the Acoustical Society of America, 104*(6), 3583-3585.

Fitch, W. (2006). The biology and evolution of music: A comparative perspective. *Cognition, 100*(1), 173-215.

Festen, J. M., Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, 88(4), 1725-1736.

François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex*, *23*(9), 2038-2043.

Fu, Q.-J., & Galvin, J. J. (2007). Perceptual learning and auditory training in cochlear implant recipients. *Trends in Amplification, 11*(3), 193-205.

Fu, Q.-J, & Lii, J. J. G. (2012). Auditory Training for Cochlear Implant Patients, F. G. Zeng, A. N. Popper, & R. R. Fay (Eds.). (2012). *Auditory prostheses: New horizons* (Vol. 39). Springer.

Fu, Q.-J., & Nogaki, G. (2005). Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology: JARO, 6*(1), 19-27.

Fu, Q. J., Shannon, R. V, & Wang, X. (1998). Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing. *The Journal of the Acoustical Society of America, 104*(6), 3586-3596.

Fuller, C. D., Galvin, J. J., Maat, B., Free, R. H., & Başkent, D. (2014). The musician effect: does it persist under degraded pitch conditions of cochlear implant simulations? *Frontiers in Neuroscience, 8*, 179.

Galvin, J. J., Fu, Q.-J., & Nogaki, G. (2007). Melodic contour identification by cochlear implant listeners. *Ear and Hearing, 28*(3), 302-319.

Galvin, J. J., Fu, Q.-J., & Shannon, R. V. (2009). Melodic contour identification and music perception by cochlear implant users. *Annals of the New York Academy of Sciences*, *1169*, 518-533.

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *23*(27), 9240-9245.

Gfeller, K., Woodworth, G., Robin, D. A., Witt, S., & Knutson, J. F. (1997). Perception of rhythmic and sequential pitch patterns by normally hearing adults and adult cochlear implant users. *Ear and hearing, 18*(3), 252-260.

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

Grayden, D. B., & Clark, G. M. (2008). Implant design and development. In H. R. Cooper, & L. C. Craddock (ed.), *Cochlear Implants: A Practical Guide*. London: Whurr Publishers.

Green, C. S., & Bavelier, D. (2008). Exercising your brain: a review of human brain plasticity and training-induced learning. *Psychology and Aging, 23*(4), 692-701.

Greenberg, S., Ainsworth, W. A. (2004). Speech Processing in the Auditory System: An Overview. *In S. Greenberg, W. A. Ainsworth, A. N. Popper, R. R. Fay. Speech Processing in the Auditory system. New York: Springer*

Handel, S. (1995). Timbre perception and auditory object identification. In B.C.J. Moore (ed.), *Hearing*. New York: Academic Press.

Harrison, R. V, Gordon, K. A., & Mount, R. J. (2005). Is there a critical period for cochlear implantation in congenitally deaf children? Analyses of hearing and speech perception performance after implantation. *Developmental Psychobiology, 46*(3), 252-261.

Heng, J., Cantarero, G., Elhilali, M., & Limb, C. J. (2011). Impaired perception of temporal fine structure and musical timbre in cochlear implant users. *Hearing research, 280*(1), 192-200.

Ho, Y.-C., Cheung, M.-C., & Chan, A. S. (2003). Music training improves verbal but not visual memory: Cross-sectional and longitudinal explorations in children. *Neuropsychology, 17*(3), 439-450.

Houtsma, A. J. M. (1997). Pitch and timbre: Definition, meaning and use. *Journal of New Music Research, 26*(2), 104-115.

Huron, D. (2001). Is music an evolutionary adaptation? *Annals of the New York Academy of Sciences, 930*, 43-61.

Kiefer, J., Hohl, S., Stürzebecher, E., Pfennigdorff, T., & Gstöettner, W. (2001). Comparison of speech recognition with different speech coding strategies (SPEAK, CIS, and ACE) and their relationship to telemetric measures of compound action potentials in the nucleus CI 24M cochlear implant system. *Audiology : Official Organ of the International Society of Audiology*, *40*(1), 32-42.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews. Neuroscience, 11*(8), 599-605.

Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance, 18*(3), 739-751.

Leung, J., Wang, N.-Y., Yeagle, J. D., Chinnici, J., Bowditch, S., Francis, H. W., & Niparko, J. K. (2005). Predictive models for cochlear implantation in elderly candidates. *Archives of Otolaryngology--Head & Neck Surgery*, *131*(12), 1049-1054.

Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs: General and Applied*, *68*(8), 1-13.

Lieberman, P. (1988). *Speech physiology, speech perception, and acoustic phonetics*. Cambridge University Press.

Limb, C. J., & Roy, A. T. (2014). Technological, biological, and acoustical constraints to music perception in cochlear implant users. *Hearing Research*, *308*, 13-26.

Looi, V., Gfeller, K., & Driscoll, V. (2012). Music appreciation and training for cochlear implant recipients: a review. *Seminars in Hearing*, *1*(212), 307-334.

Looi, V., McDermott, H., McKay, C., & Hickson, L. (2008). Music perception of cochlear implant users compared with that of hearing aid users. *Ear and Hearing, 29*(3), 421-434.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. J. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(49), 18866-18869.

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

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7-8), 953-978.

McDermott, J., & Hauser, M. D. (2003). The evolution of the music faculty: A comparative perspective. *Nature neuroscience, 6*(7), 663-668.

Mckay, C. M., & McDermott, H. J. (1993). Perceptual Performance of Subjects with Cochlear Implants Using the Spectral Maxima Sound Processor (SMSP) and the Mini Speech Processor (MSP). *Ear & Hearing, 24*(5), 350-567.

McKay, C. M., McDermott, H. J., & Carlyon, R. P. (2000). Place and temporal cues in pitch perception: Are they truly independent? *Acoustics Research Letters Online, 1*(1), 25-30.

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.

Mesgarani, N., & Chang, E. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature, 485*, 233-237.

Mirza, S., Douglas, S. A., Lindsey, P., Hildreth, T., & Hawthorne, M. (2003). Appreciation of music in adult patients with cochlear implants: a patient questionnaire. *Cochlear Implants International, 4*(2), 85-95.

Moore, B. C. J., Glasberg, B. R. & Shailer, M. J. (1984). Frequency and intensity difference limens for harmonics within complex tones. *The Journal of the Acoustical Society of America,* 75, 550-561.

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cerebral Cortex, 19*(3), 712-723.

Mozziconacci, S. (2002). Prosody and Emotions. *Speech Prosody 2002*.

Mueller, S. T. & Piper, B. J. (2014). The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *Journal of Neuroscience Methods, 222*, 250-259.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of the United States of America, 104*(40), 15894-15898.

Nelson, P. B., Jin, S.-H., Carney, A. E., & Nelson, D.A. (2003). Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners. *The Journal of the Acoustical Society of America, 113*(2), 961-968.

Olson, A., & Canada, T. (2013) Using computerized auditory training clinically for adults with cochlear implants. *American Speech Language Hearing Association*. Available at: http://www.asha.org

Parbery-Clark, A., Tierney, A., Strait, D. L., & Kraus, N. (2012). Musicians have fine-tuned neural distinction of speech syllables. *Neuroscience, 219*, 111-119.

Patel, A. D. (2011). Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. *Frontiers in Psychology, 2*(June), 142.

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, *308*, 98-108.

Peppé, S., & McCann, J. (2003). Profiling Elements of Prosody in Speech-Communication (PEPS-C). *Clinical Linguistics & Phonetics, 17*(4-5), 345-354.

Peters, R. W., Moore, B. C., & Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *The Journal of the Acoustical Society of America, 103*(1), 577-587.

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), 134-151.

Plant, G. (1997). Testing visual and auditory-visual speech perception in Martin, M (2nd ed). *Speech Audiometry*. London: Whurr Publishers.

Plant, G. (2001). Manual of Speech Perception, Australian Hearing.

Qin, M. K., & Oxenham, A. J. (2005). Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear and Hearing, 26*(5), 451-460.

Raphael, L. J., Borden, G. J., & Harris, K. S. (2007). *Speech science primer: Physiology, acoustics, and perception of speech*. Lippincott Williams & Wilkins.

Regard, M. (1981). *Cognitive rigidity and flexibility: A neuropsychological study*. University of Victoria. Unpublished doctoral dissertation.

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*(July), 1-17.

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science, 15*(8), 511-514.

Schmithorst, V. J., & Wilke, M. (2002). Differences in white matter architecture between musicians and non-musicians: a diffusion tensor imaging study. *Neuroscience Letters, 321*(1-2), 57-60.

Schouten, J. F., Ritsma, R. J., & Cardozo, B. L. (1962). Pitch of the residue. *The Journal of the Acoustical Society of America, 34*(9B), 1418-1424.

Sharma, A., Dorman, M. F., & Spahr, A. J. (2002). A sensitive period for the development of the central auditory system in children with cochlear implants: implications for age of implantation. *Ear and Hearing, 23*(6), 532-539.

Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in Amplification, 12*(4), 283-299.

Skinner, M. W., Holden, L. K., Whitford, L. A., Plant, K. L., Psarros, C., & Holden, T. A. (2002). Speech Recognition with the Nucleus 24 SPEAK, ACE, and CIS Speech Coding Strategies in Newly Implanted Adults. *Ear and Hearing*, *23*(3), 207-223.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature, 416*, 87-90.

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *335*(1273), 71-78.

Summerfield, Q., Culling, J. F., & Fourcin, A. J. (1992). Auditory Segregation of Competing Voices: Absence of Effects of FM or AM Coherence. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 336*(1278), 357-366.

Stevens, K. N. (1980). Acoustic correlates of some phonetic categories. *The Journal of the Acoustical Society of America, 68*, 836-842.

Stickney, G. S., Zeng, F.-G., Litovsky, R., & Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *The Journal of the Acoustical Society of America*, *116*(2), 1081-1091.

Strait, D. L., & Kraus, N. (2014). Biological impact of auditory expertise across the life span: musicians as a model of auditory learning. *Hearing Research, 308*, 109-121.

Stroop, J. R. (1935). Studies of interference in serial verbal reaction. *Journal of Experimental Psychology, 18*, 643-662.

Swanson, B. (2008). *Pitch perception with cochlear implants*. Unpublished PhD.

Tang, Q., Benítez, R., & Zeng, F. G. (2011). *Spatial channel interactions in cochlear implants. Journal of neural engineering, 8*(4), 1-28.

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.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2001). Research Report Arousal, mood, and the Mozart effect. *Psychological Science, 12*(3), 248-251.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: do music lessons help? *Emotion, 4*(1), 46-64.

Tovey, D. F. (1975). *The Forms of Music*. Cleveland: Meridian Books.

Tun, P. A., Mccoy, S., & Wingfield, A. (2010). Aging, Hearing Acuity, and the Attentional Costs of Effortful Listening, *24*(3), 761-766.

Välimaa, T. T., Määttä, T. K., Löppönen, H. J., & Sorri, M. J. (2002). Phoneme recognition and confusions with multichannel cochlear implants: consonants. *Journal of Speech, Language, and Hearing Research*, *45*(5), 1055-1069.

Vandali, A. E. (2001). Emphasis of short-duration acoustic speech cues for cochlear implant users. *The Journal of the Acoustical Society of America*, *109*(5), 2049-2061.

Vandali, A. E., Sucher, C., Tsang, D. J., McKay, C. M., Chew, J. W. D., & McDermott, H. J. (2005). Pitch ranking ability of cochlear implant recipients: A comparison of sound-processing strategies. *The Journal of the Acoustical Society of America*, *117*(5), 3126-3138.

Vandali, A. E., Whitford, L. A, Plant, K. L., & Clark, G. M. (2000). Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system. *Ear and Hearing*, *21*(6), 608-624.

Watt, R., Quinn, S., (2006). Analysis of local and global timing and pitch change in ordinary melodies. In M. Baroni, et al. (Eds.), *Proceedings of the 9th International Conference on Music Perception & Cognition (ICMPC9)*, Bologna/Italy, August 22-26, 2006, 30-37.

Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., & Ackermann, H. (2005). Identification of emotional intonation evaluated by fMRI. *NeuroImage, 24*(4), 1233-1241.

Wilson, B., & Dorman, M. (2008). Cochlear implants: a remarkable past and a brilliant future. *Hearing Research*, *242*, 3-21.

Woodruff, P. W. R, Benson, R. R., Bandettini, P. A., Kwong, K. K., Howard, R. J., Talavage, T., Belliveau, & Rosen, B. R. (1996). Modulation of auditory and visual cortex by selective attention is modality-dependent. *NeuroReport, 7*, 1909-1913.

Yost, W. A. (2009). Pitch perception. *Attention, Perception & Psychophysics, 71*(8), 1701-1715.

Yukawa, K., Cohen, L., Blamey, P., Pyman, B., Tungvachirakul, V., & O'Leary, S. (2004). Effects of insertion depth of cochlear implant electrodes upon speech perception. *Audiology Neuro Otolaryngology, 9*(3), 163-172.

Zendel, B. R., & Alain, C. (2013). The influence of lifelong musicianship on neurophysiological measures of concurrent sound segregation. *Journal of Cognitive Neuroscience*, *25*(4), 503-516.

Zeng, F.-G. (2004). Trends in Cochlear Implants. *Trends in Amplification*, *8*(1), 1-34.

Zeng, F. G., Rebscher, S., Harrison, W., Sun, X., & Feng, H. (2008). Cochlear implants: system design, integration, and evaluation. *Biomedical Engineering, IEEE Reviews 1*, 115-142.

7 Appendix A

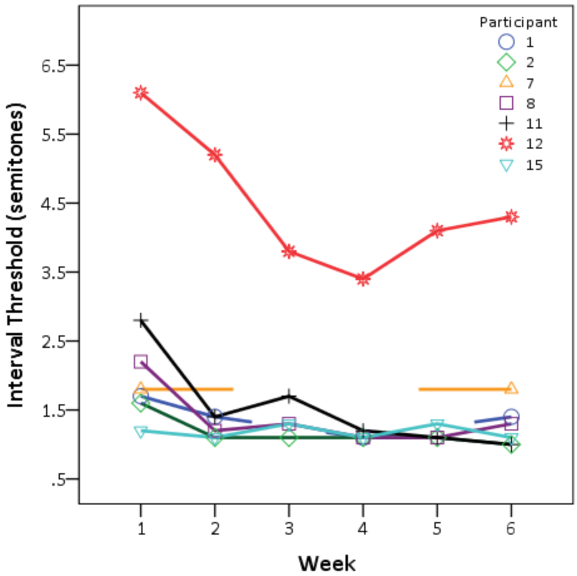
Table 7.1. MCTP (Interval) frequency and duration information

	Semi-tone range	F0 range (Hz)	Note duration (ms)	Total duration (ms)
Rising	28	440-2218	250	1250
Rising	24	440-1760	250	1250
Rising	20	440-1397	250	1250
Rising	16	440-1109	250	1250
Rising	12	440-880	250	1250
Rising	8	440-698	250	1250
Rising	4	440-554	250	1250
Rising-Flat	14	440-988	250	1250
Rising-Flat	12	440-880	250	1250
Rising-Flat	10	440-784	250	1250
Rising-Flat	8	440-698	250	1250
Rising-Flat	6	440-622	250	1250
Rising-Flat	4	220-544	250	1250
Rising-Flat	2	440-494	250	1250
Rising-Falling	14	440-988	250	1250
Rising-Falling	12	440-880	250	1250
Rising-Falling	10	440-784	250	1250
Rising-Falling	8	440-698	250	1250
Rising-Falling	6	440-622	250	1250
Rising-Falling	4	220-544	250	1250
Rising-Falling	2	440-494	250	1250
Flat-Rising	28	440-988	250	1250
Flat-Rising	24	440-880	250	1250
Flat-Rising	20	440-784	250	1250
Flat-Rising	16	440-698	250	1250
Flat-Rising	12	440-622	250	1250
Flat-Rising	8	220-544	250	1250
Flat-Rising	4	440-494	250	1250
Flat	0	440	250	1250
Flat	0	440	250	1250
Flat	0	440	250	1250
Flat	0	440	250	1250
Flat	0	440	250	1250
Flat	0	440	250	1250
Flat	0	440	250	1250
Flat-Falling	14	440-988	250	1250
Flat-Falling	12	440-880	250	1250
Flat-Falling	10	440-784	250	1250
Flat-Falling	8	440-698	250	1250
Flat-Falling	6	440-622	250	1250
Flat-Falling	4	220-544	250	1250
Flat-Falling	2	440-494	250	1250
Falling-Rising	14	440-988	250	1250
Falling-Rising	12	440-880	250	1250
Falling-Rising	10	440-784	250	1250
Falling-Rising	8	440-698	250	1250
Falling-Rising	6	440-622	250	1250
Falling-Rising	4	220-544	250	1250
Falling-Rising	2	440-494	250	1250
Falling-Flat	14	440-988	250	1250
Falling-Flat	12	440-880	250	1250
Falling-Flat	10	440-784	250	1250
Falling-Flat	8	440-698	250	1250
Falling-Flat	6	440-622	250	1250
Falling-Flat	4	220-544	250	1250
Falling-Flat	2	440-494	250	1250
Falling	28	440-988	250	1250
Falling	24	440-880	250	1250
Falling	20	440-784	250	1250
Falling	16	440-698	250	1250
Falling	12	440-622	250	1250
Falling	8	220-544	250	1250
Falling	4	440-494	250	1250

Table 7.2. MCTP (Duration) frequency and duration information

	Semi-tone range	F0 range (Hz)	Note duration (ms)	Total duration (ms)
Rising	20	440-1397	450	2250
Rising	20	440-1397	350	1750
Rising	20	440-1397	250	1250
Rising	20	440-1397	200	1000
Rising	20	440-1397	150	750
Rising	20	440-1397	100	500
Rising	20	440-1397	50	250
Rising-Flat	10	440-784	450	2250
Rising-Flat	10	440-784	350	1750
Rising-Flat	10	440-784	250	1250
Rising-Flat	10	440-784	200	1000
Rising-Flat	10	440-784	150	750
Rising-Flat	10	440-784	100	500
Rising-Flat	10	440-784	50	250
Rising-Falling	10	440-784	450	2250
Rising-Falling	10	440-784	350	1750
Rising-Falling	10	440-784	250	1250
Rising-Falling	10	440-784	200	1000
Rising-Falling	10	440-784	150	750
Rising-Falling	10	440-784	100	500
Rising-Falling	10	440-784	50	250
Flat-Rising	10	440-784	450	2250
Flat-Rising	10	440-784	350	1750
Flat-Rising	10	440-784	250	1250
Flat-Rising	10	440-784	200	1000
Flat-Rising	10	440-784	150	750
Flat-Rising	10	440-784	100	500
Flat-Rising	10	440-784	50	250
Flat	0	440	450	2250
Flat	0	440	350	1750
Flat	0	440	250	1250
Flat	0	440	200	1000
Flat	0	440	150	750
Flat	0	440	100	500
Flat	0	440	50	250
Flat-Falling	10	440-784	450	2250
Flat-Falling	10	440-784	350	1750
Flat-Falling	10	440-784	250	1250
Flat-Falling	10	440-784	200	1000
Flat-Falling	10	440-784	150	750
Flat-Falling	10	440-784	100	500
Flat-Falling	10	440-784	50	250
Falling-Rising	10	440-784	450	2250
Falling-Rising	10	440-784	350	1750
Falling-Rising	10	440-784	250	1250
Falling-Rising	10	440-784	200	1000
Falling-Rising	10	440-784	150	750
Falling-Rising	10	440-784	100	500
Falling-Rising	10	440-784	50	250
Falling-Flat	10	440-784	450	2250
Falling-Flat	10	440-784	350	1750
Falling-Flat	10	440-784	250	1250
Falling-Flat	10	440-784	200	1000
Falling-Flat	10	440-784	150	750
Falling-Flat	10	440-784	100	500
Falling-Flat	10	440-784	50	250
Falling	20	440-1397	450	2250
Falling	20	440-1397	350	1750
Falling	20	440-1397	250	1250
Falling	20	440-1397	200	1000
Falling	20	440-1397	150	750
Falling	20	440-1397	100	500
Falling	20	440-1397	50	250

8 Appendix B



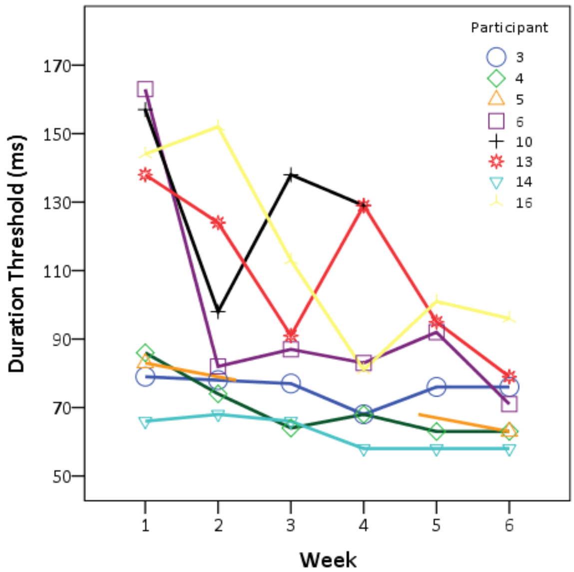


Figure 8.1. Individual results from training across weeks.
MCTP (Interval) on the left, and MCTP (Duration) on the right.
Broken lines indicate the participant did not train during that week.

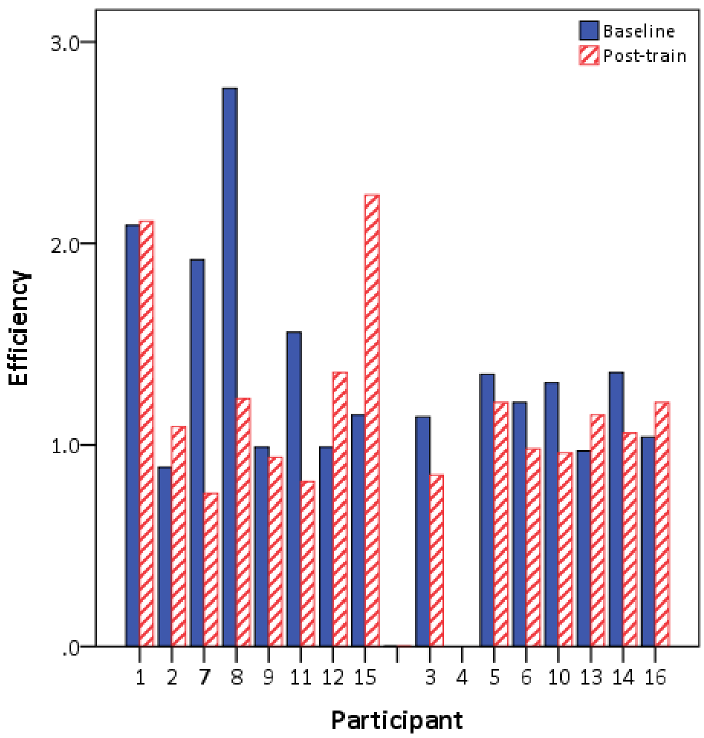


Figure 8.2. Individual results for VST. Participants that trained with MCTP (Interval) are nested on the left, with MCTP (Duration) on the right. Participant 4 was unable to complete this test due to colour blindness.

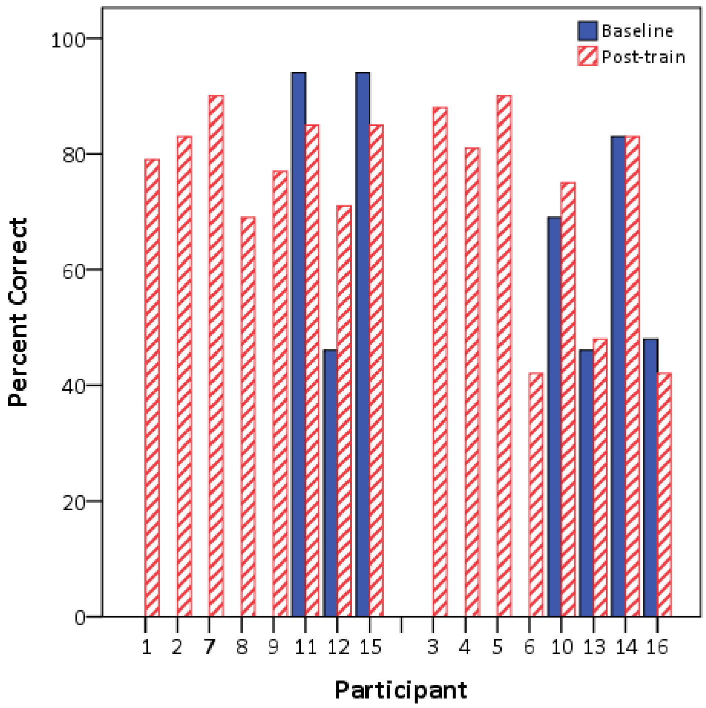


Figure 8.3. Individual results for Pitch Discrimination (131 Hz). Participants that trained with MCTP (Interval) are nested on the left, with MCTP (Duration) on the right. All participants completed baseline testing, but only 7 participants completed both baseline and post-train sessions.

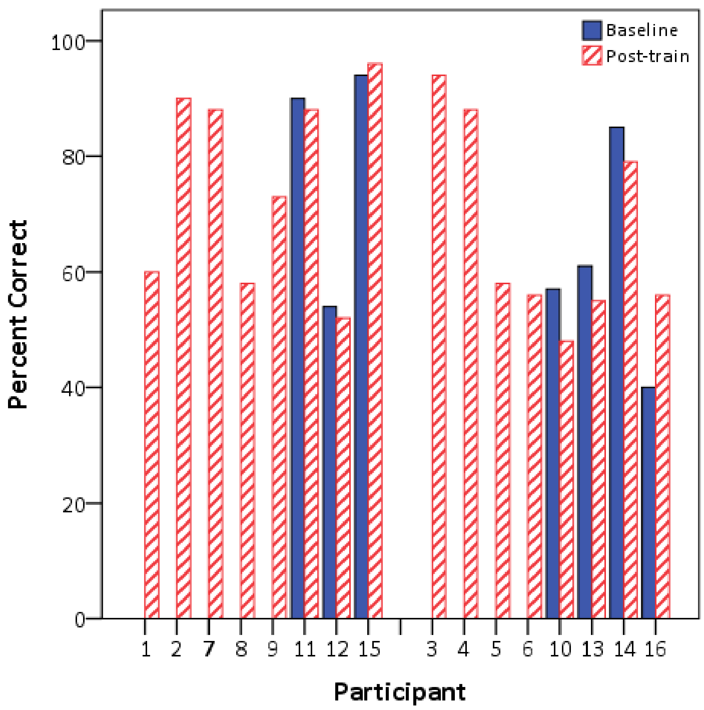


Figure 8.4. Individual results for Pitch Discrimination (440 Hz). Participants that trained with MCTP (Interval) are nested on the left, with MCTP (Duration) on the right. All participants completed baseline testing, but only 7 participants completed both baseline

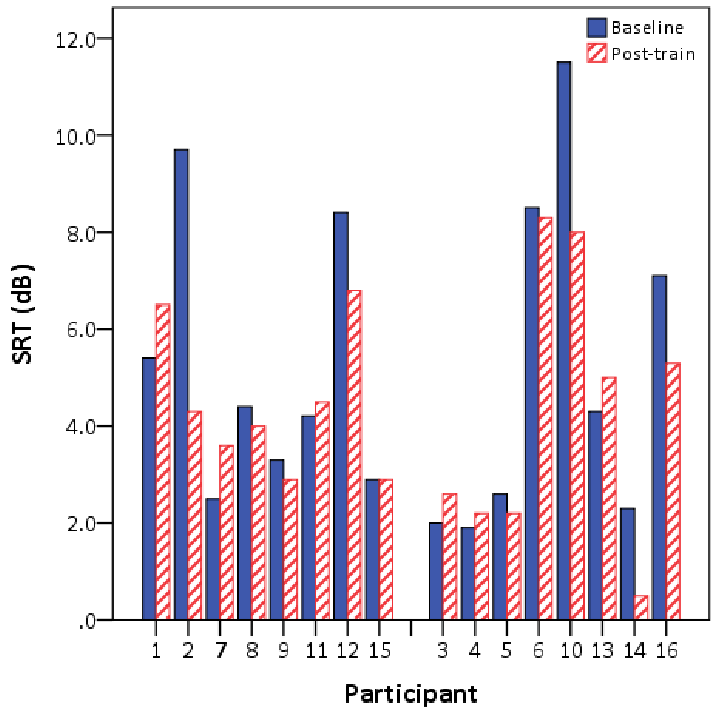


Figure 8.5. Individual results for AuSTIN.
Participants that trained with MCTP (Interval) are nested on the left, with MCTP (Duration) on the right.

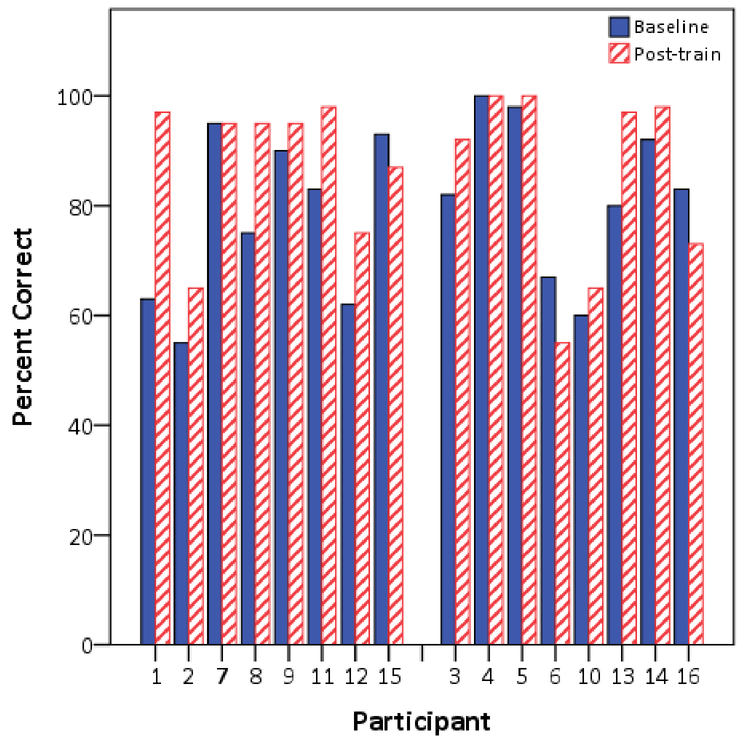


Figure 8.6. Individual results for consonant discrimination in quiet.
Participants that trained with MCTP (Interval) are nested on the left, with MCTP (Duration) on the right.

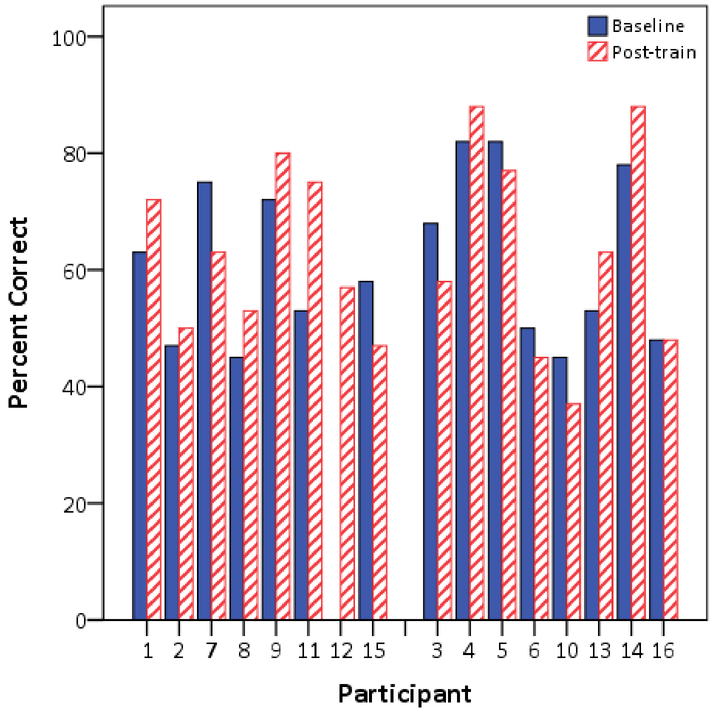


Figure 8.7. Individual results for consonant discrimination with 4TB. Participants that trained with MCTP(Interval) are nested on the left, with MCTP(Duration) on the right. Participant was unable to complete the task at baseline, citing difficulty, but was able to complete the task post-training.

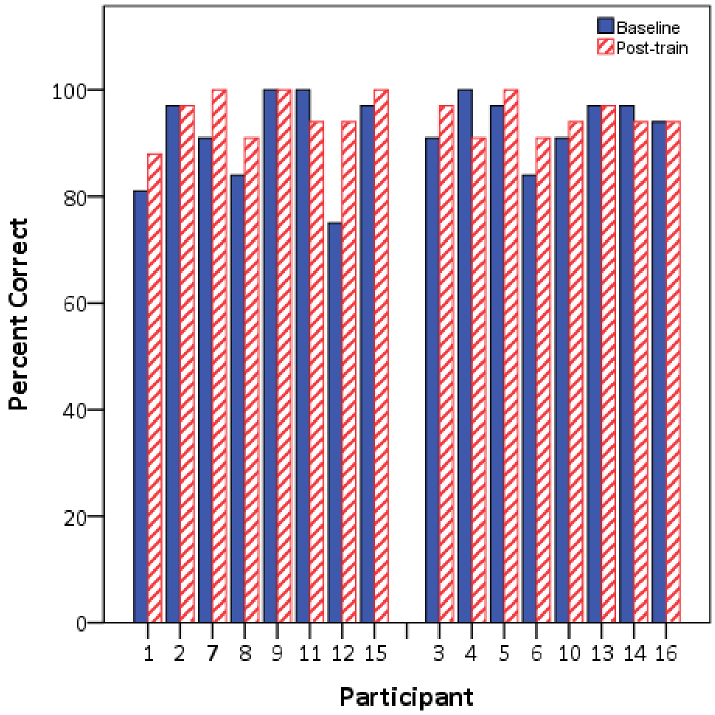


Figure 8.8. Individual results for MBEP(Happy/Sad).
Participants that trained with MCTP(Interval) are nested on the left, with MCTP(Duration) on the right.

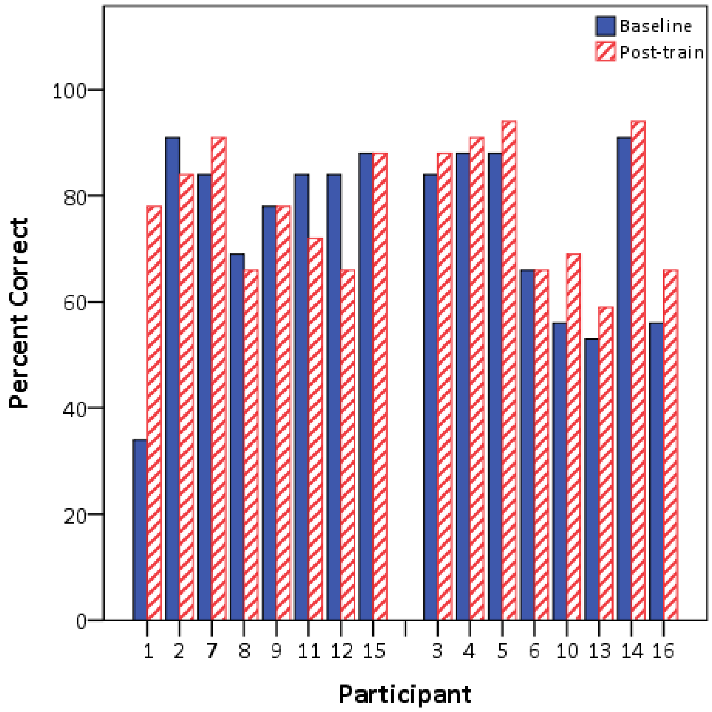


Figure 8.9. Individual results for MBEP(Fear/Irritation).
Participants that trained with MCTP(Interval) are nested on the left, with MCTP(Duration) on the right.

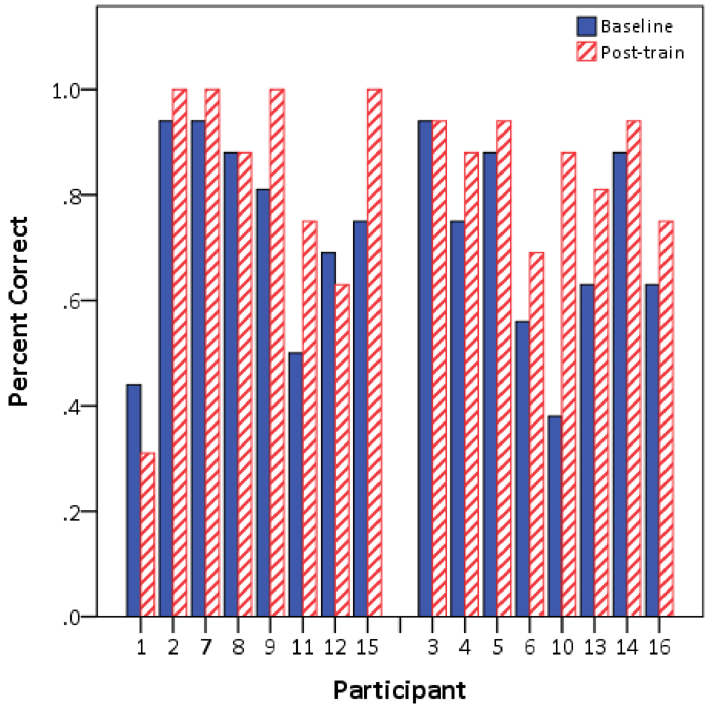


Figure 8.10. Individual results for prosody discrimination (question/statement).
Participants that trained with MCTP(Interval) are nested on the left, with MCTP(Duration) on the right.

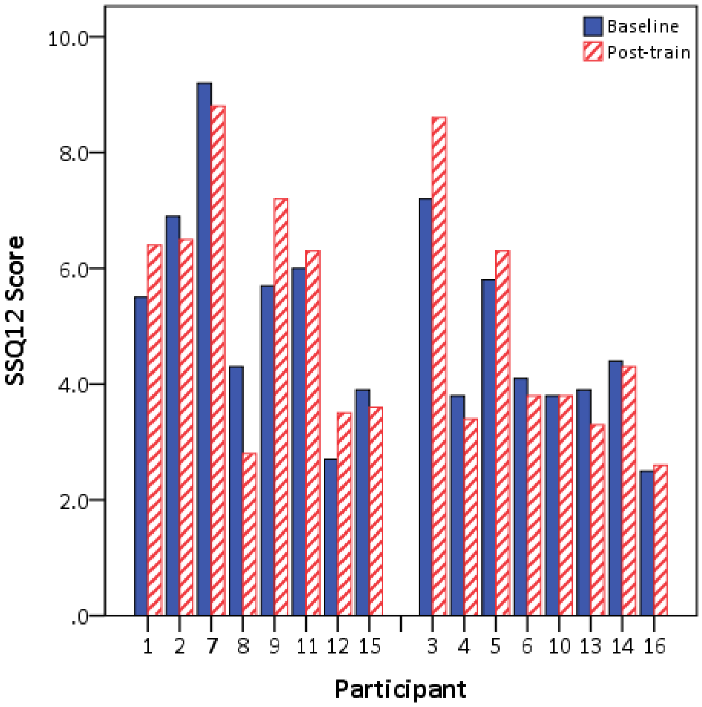


Figure 8.11. Individual results for SSQ12 scores.
Participants that trained with MCTP(Interval) are nested on the left, with MCTP(Duration) on the right.

9 Appendix C

Dear A/Prof McMahon,

Re: "Manipulation of interval size and note duration for melodic contour training and its effect on speech perception for cochlear implant users"(5201400348)

Thank you for your recent correspondence. Your response has addressed the issues raised by the Faculty of Human Sciences Human Research Ethics Sub-Committee and approval has been granted, effective 24th April 2014. This email constitutes ethical approval only.

This research meets the requirements of the National Statement on Ethical Conduct in Human Research (2007). The National Statement is available at the following web site:

http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/e72.pdf.

The following personnel are authorised to conduct this research:

A/Prof Catherine M McMahon

Professor Bill Thompson

Dr Valerie Looi

Mr Chi Yhun Lo

Please note the following standard requirements of approval:

1. The approval of this project is conditional upon your continuing compliance with the National Statement on Ethical Conduct in Human Research (2007).
2. Approval will be for a period of five (5) years subject to the provision of annual reports.

Progress Report 1 Due: 24th April 2015

Progress Report 2 Due: 24th April 2016

Progress Report 3 Due: 24th April 2017

Progress Report 4 Due: 24th April 2018

Final Report Due: 24th April 2019

NB. If you complete the work earlier than you had planned you must submit a Final Report as soon as the work is completed. If the project has been discontinued or not commenced for any reason, you are also required to submit a Final Report for the project.

Progress reports and Final Reports are available at the following website:

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/forms

3. If the project has run for more than five (5) years you cannot renew approval for the project. You will need to complete and submit a Final Report and submit a new application for the project. (The five year limit on renewal of approvals allows the Sub-Committee to fully re-review research in an environment where legislation, guidelines and requirements are continually changing, for example, new child protection and privacy laws).

4. All amendments to the project must be reviewed and approved by the Sub-Committee before implementation. Please complete and submit a Request for Amendment Form available at the following website:

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/forms

5. Please notify the Sub-Committee immediately in the event of any adverse effects on participants or of any unforeseen events that affect the continued ethical acceptability of the project.

6. At all times you are responsible for the ethical conduct of your research in accordance with the guidelines established by the University. This information is available at the following websites:

http://www.mq.edu.au/policy

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/policy

If you will be applying for or have applied for internal or external funding for the above project it is your responsibility to provide the Macquarie University's Research Grants Management Assistant with a copy of this email as soon as possible. Internal and External funding agencies will not be informed that you have approval for your project and funds will not be released until the Research Grants Management Assistant has received a copy of this email.

If you need to provide a hard copy letter of approval to an external organisation as evidence that you have approval, please do not hesitate to contact the Ethics Secretariat at the address below.

Please retain a copy of this email as this is your official notification of ethics approval.

Yours sincerely,

Dr Simon Boag

Acting Chair

Faculty of Human Sciences

Human Research Ethics Sub-Committee

FHS Ethics

Faculty of Human Sciences

C5C Research Hub East

Level 3, Room 324

Macquarie University

NSW 2109 Australia

T: +61 2 9850 4197

F: +61 2 9850 4465

E: fhs.ethics@mq.edu.au