Spring 2018

Melodic contour identification and speech recognition by school-aged children

Michael P. Morikawa
James Madison University

Melodic Contour Identification and Speech Recognition by School-Aged Children

Michael P. Morikawa

A dissertation submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Audiology

Communication Sciences and Disorders

May 2018

FACULTY COMMITTEE:

Committee Chair: Yingjiu Nie, Ph.D.

Committee Members/ Readers:

Ayaskanta Rout, Ph.D.

Rory Depaolis, Ph.D.

Acknowledgements

I'd like to express my sincere appreciation and gratitude to the faculty at James Madison University. Specifically, thank you to my advisor, Dr. Yingjiu Nie, for her tireless work and the countless hours she has spent lending her knowledge and expertise to this project and myself. Thank you to Dr. Qian-Jie Fu for sharing the Sung Speech Corpus program. Thank you to my committee members, Dr. Ayaskanta Rout and Dr. Rory DePaolis for their support and contributions. My sincerest appreciation and love is extended to the Au.D. Class of 2018—thank you for the friendship, support, and positivity. Special thanks to my wonderful research participants and members of the lab, specifically Victoria André, Sarah Troy, and Lindsey Seyfried for without them this project would be impossible. A final thank you is extended to my family and friends who have been so encouraging since the beginning. This project's existence is a reflection of your unwavering love and support.

This work was partially supported by the Faculty Fund and by the Roger Ruth Memorial Grant from the Department of Communication Sciences and Disorders at James Madison University.

Table of Contents

List of Figures

Abstract

Using the Sung Speech Corpus (SSC), which encompasses a single database that contains musical pitch, timbre variations and speech information in identification tasks, the current study aimed to explore the development of normal-hearing children's ability to use the pitch and timbre cues. Thirteen normal hearing children were recruited for the study ages ranging from 7 to 16 years old. Participants were separated into two separate groups: Younger (7-9) and Older (10-16). Musical Experience was taken into account as well. The Angel Sound ™ program was utilized for testing which was adopted from previous studies, most recently Crew, Galvin, and Fu (2015). Participants were asked to identify either pitch contour or a five word sentence while the one not being identified was manipulated in quiet. Each sentence recognition task was also tested at three different SNRs (-3, 0, 3 dB). For sentence recognition in quiet, children with musical training performed better than those without. A significant interaction between Age-Group and Musical Experience was also seen, such that Younger children showed more benefit from musical training than Older, musically trained children. Significant effect of pitch contour on sentence recognition in noise was found showing that naturally produced speech stimuli were easier to identify when competing background noise was introduced for all children than speech stimuli with an unnatural pitch contour. Significant effect of speech timbre on MCI was found which demonstrates that as the timbre complexity increases, the MCI performance decreases. The current study concluded that pitch and timbre cues interfered with each other in child listeners, depending on the listening demands (SNR, tasks, etc.). Music training can improve overall speech and music perception.

Manuscript

1. Introduction

Pitch and timbre are two of the main attributes of sounds that are important for speech perception. The acoustic correlate of perceived pitch in spoken English, the fundamental frequency (F0), provides information that allows for speaker identification (Carey, Parris, Lloyd-Thomas, & Bennett, 1997), intent (Grant, 1996), and emotion (Murray, 1993). Extensive evidence has shown that F0 contour facilitates segregation of target speech from competing maskers (e.g., Assmann & Summerfield, 1990; for a review see, Darwin, 2008; Drullman & Bronkhorst, 2004). The acoustic correlates of perceived timbre involve the distributions of energy over time and frequency such as features of spectral or temporal envelope (Moore, 2003). These correlates have been widely studied in musical timbre (e.g., McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995; J. M. Grey, 1977) which is typically referred to as an attribute that allows listeners to distinguish instruments (e.g., piano versus violin) playing the same the note with the same loudness and duration (e.g., Grey, 1975). The acoustic correlates of timbre are also important cues for speech recognition (e.g., Ardoint, Agus, Sheft, & Lorenzi, 2011, for temporal envelope; Keurs, Festen, & Plomp, 1992 for spectral envelope). For example, variation in the positions of amplitude peaks on the frequency spectrum (i.e., formant position) of a synthetic vowel may alter its identification to listeners (Delattre, Liberman, Cooper, & Gerstman, 1952; Klatt, 1982; Molis, 2005; Swanepoel, Oosthuizen, & Hanekom, 2012) and shifts in formant positions distort phonetic judgments of vowel similarity (Carlson & Granstrom, 1979; Klatt, 1982). In addition, rise time difference in the temporal envelope has been noted to vary consonant identification (e.g., Goswami, Fosker, Huss, Mead, & Szűcs, 2011). In short, for speech perception, pitch and pitch contours provide suprasegmental information as well as cues for separation of target speech from maskers, whereas acoustic correlates of timbre are important for speech recognition by providing cues for the identification of segmental elements of speech, such as phonemes. The present study focuses on the perception of pitch contour and timbre with respect to speech recognition.

While research has shown mutual interference between pitch and timbre perception using non-speech stimuli (Allen & Oxenham, 2014), studies using speech stimuli have mainly examined the effect of variations of pitch contour on the processing of speech timbre (reflected by speech recognition) (e.g., Miller and Schlauch & Watson, 2010). In general, when pitch contours of utterances are altered away from the natural linguistic representations to some extent, significant reduction of speech recognition in the presence of noise have been widely documented (Binns & Culling, 2007; Miller, Schlauch, & Watson, 2010). Only recently, the effect of variations of speech timbre on the identification of pitch contour was examined in adult musician and non-musician listeners both with normal hearing (Crew, Galvin & Fu, 2015). In that study, to address the concerns that different stimuli (e.g., speech versus musical notes) and test procedures (e.g., spoken emotion discrimination versus melodic contour identification) had been used across studies (e.g., Chatterjee et al., 2015 for emotion discrimination; Galvin, Fu, & Oba, 2009 for melodic contour identification) when assessing the contribution of pitch and timbre to speech perception, the authors developed the Sung Speech Corpus (SSC) that allowed for the examination of pitch and timbre perceptions using the same set of stimuli. The SSC is a closed-set of stimuli comprising sentences of five spoken monosyllabic words. The fundamental frequency (F0) across the words is either varied to attain desired melodic contours or it remains in the natural-speech pattern. The Melodic Contour Identification (MCI) is measured while the consistency of words within and across sentences is varied resulting in different levels of timbre complexity. Conversely, sentence recognition is measured to assess timbre processing (i.e. sentence recognition) while the F0 contour is varied in the alternatives of different melodic contours and the natural-speech contour. The authors found that, for MCI, non-musicians performed less accurately as the timbre condition became more complex, whereas musicians reached near-perfect scores regardless of the complexity of timbre conditions, showing a musician advantage. In contrast, for the processing of timbre (measured as sentence recognition), both listener groups scored near perfect regardless of the variations of the F0 contour. In short, for NH adult non-musicians, higher timbre complexity produced more interference on their ability to track pitch contours than variations of pitch contour did on their ability to process timbre. Additionally, musical experience was found to facilitate counteracting the adverse effect of timbre complexity on the perception of pitch contour meaning that musician listeners performed comparably well in the MCI task across various levels of timbre complexity.

Research on children's ability to identify pitch contours of complex stimuli, such as musical notes or speech, is still growing, although the ability to discriminate speech intonations has been evidenced in infancy (for review, see Vihman, 2014). A recent study (Stalinski, Schellenberg, & Trehub, 2008) suggested that NH children may have reached an adult-like level of pitch contour identification at around 8 years of age. In that study, participants were asked to judge whether the target note, which occurred in the middle of the sequence, was higher or lower in pitch than the two reference notes after being presented with the sequence of 3 synthesized piano notes. Thus, this study focused more on pitch ranking, per se, rather than pitch contour identification. However, research is emerging to study the identification of pitch contours in pediatric cochlear implant (CI) users (See, Driscoll, Gfeller, Kliethermes, & Oleson, 2013; Tao et al., 2015). Consequently, better understanding of such identification in NH children is warranted to lay a baseline for studies on children with hearing impairment.

Additionally, in NH children, little is known regarding the effect of variations of pitch contours on speech recognition (i.e., processing of speech timbre) and the effect of variations of timbre on identification of pitch contours. Evidence of these effects is particularly informative for understanding pediatric CI users' pitch and timbre perception. These robust cues for NH listeners' pitch perception are different from those for timbre perception, the former including temporal fine structure and harmonic resolution (e.g., McDermott & Oxenham, 2008; Oxenham, Bernstein, & Penagos, 2004), while the latter involving attack time (extracted from temporal envelope) and spectral centroid (contained in spectral envelope)—a noise-robust estimate of how the dominant frequency of a signal changes over time (e.g., Caclin, McAdams, Smith, & Winsberg, 2005; Elliott, Hamilton, & Theunissen, 2013; McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995; Massar, Fickus, Bryan, Petkie, and Terzuoli, 2010). Similar to NH listeners, CI users rely on both temporal envelopes and spectral envelopes for timbre perception (Kong, Mullangi, Marozeau, & Epstein, 2011; Macherey & Delpierre, 2013). However, different from NH listeners, CI users rely heavily on spectral envelope cues for pitch perception (Crew, Galvin, & Fu, 2012) due to the lack of access to temporal fine structure. Such dependence on the same cues (i.e., spectral envelopes) for both pitch and timbre perception may make CI users' perception of one attribute susceptible to the variations of the other attribute. Using the SSC stimuli, Crew, Galvin, and Fu (2016) have provided evidence supporting this notion in adult CI users. In addition, comparing findings in their NH peers studied with the same SSC stimuli (Crew, Galvin, & Fu, 2015), the alternations of pitch contour were suggested to have more severely degraded adult CI users' sentence recognition (i.e., timbre processing), where NH listeners scored near-perfect regardless of the variations of the pitch contour as opposed to CI users, who scored worse when the pitch contour was unnatural rather than natural. Pediatric cochlear implant users differ from the general adult in many aspects related to hearing, such as onset age of hearing loss, duration of acoustic hearing prior to cochlear implantation, plasticity of the auditory system, etc. It would be of interest to study how pediatric CI user's perception of pitch and timbre is affected by the variations of the other attribute and whether such effects differ between children with NH VS CIs.

In this study, the interdependent relationship between the processing of speech timbre and melodic contour was examined in NH children with an age range between 7 and 16 years to 1) provide a baseline for such studies in children with hearing impairment to compare with; 2) investigate the differences between younger children and older children's ability to identify pitch contour and timbre while the other is varied and 3) assess the musician advantage in this age range.

2. Materials and Methods

2.1 Participants

Thirteen normal hearing subjects participated in this study. These participants were paid volunteers recruited through the Communication Sciences and Disorders department at James Madison University using an e-mail blast asking for willing participants. All participants had pure tone thresholds at 15 dB HL or better at all audiometric frequencies from 250 Hz to 8000 Hz in their right ear. Participants were divided into two groups: Musicians (M) and Non-Musicians (NM). These two groups were defined by their musical experience, training, and confidence based on a questionnaire completed before participation. The musician group was determined by at least three years of formal musical training. All participants also reported their musical confidence ranging from 1-10 with 1 being least confident and 10 being most confident. Once these groups were established they were once again parsed down into smaller groups – ages 7-9 (Y) and ages 10-16 (O). Prior to participation, informed consent and assent were obtained from participants' authorized caregivers and the participants respectively, in accordance with a protocol approved by the Institutional Review Board at James Madison University.

2.2 *Stimuli*

The stimuli were generated via the Angel Sound ™ program (http://angelsound.emilyfufoundation.org) controlled by a DELL computer routed through the High Definition Sound Device soundcard and a DAC1 D/A converter, and presented through a Tucker-Davies Technologies (TDT) RZ-6 headphones buffer driving a HDA 200 circumaural headphone.

The Sung Speech Corpus (SSC) (Crew, Galvin, & Fu, 2015; Crew, et al., 2016) is made up of 50 sung monosyllabic words produced by an adult male speaker that creates simple sentences with syntax of: 'name', 'verb', 'number', 'color', 'clothing' (ex: "Bob wears four brown belts"). Each of the five categories contained ten words and each word was sung at all thirteen pitches from 110 Hz to 220 Hz in different semitone steps. This allows for a five-word sentence containing a five-note melody, which is used for both the Sentence Recognition and Melodic Contour Identification (MCI) conditions. Natural speech was also produced for each word to allow for comparison between natural production of words and sung speech. The stimuli used were all 500 ms in duration with minimal adjustments made after recording in order to obtain an exact F0 and amplitude.

The other set of stimuli was adopted from previous MCI studies (e.g., Crew, Galvin, Landsberger, & Fu, 2015; Crew, Galvin, & Fu, 2015; Galvin, et al., 2008) and consisted of sequences of five synthesized piano notes. The F0's of the notes were generated to form one of the nine melodic contours as in the SSC. The F0 range, F0 difference between successive notes, duration of each note, and silent gap between successive notes were identical as those set in the SSC.

2.3 *Procedure*

Testing took place over two days with reasonable amounts of breaks taken to account for fatigue. On the first day of testing, a hearing screening was conducted from 250 to 8000 Hz to ensure normal hearing across these frequencies (< 15 dB HL). Tympanometry was also performed to assess middle ear function. In order to proceed with the experiment, all hearing thresholds and tympanograms had to be within normal limits. After the hearing screening, participants were asked to fill out a survey, which was used to classify each participant's amount of musical experience on scale of 1-10, with ten being the most musically experienced. The survey was developed and scored by the experimenters. The participant's parents were also asked to sign permission forms for their children to be able to be a part of the study.

Prior to the experimental conditions, participants received a minimum of four practice sessions for the MCI test using the synthesized piano notes to assure the performance on the last two practice sessions was within 5 percentage points. The experimental conditions were blocked between the two tasks and presented in a random order under each test condition (i.e., sentence recognition test or MCI test. The signals were presented at a nominal level of 60 dB A unilaterally to the right ear. In the conditions under the sentence recognition test with the presence of background noise, the overall level was kept at 60 dB A, rendering the stimuli levels of 53, 54, 55, and 56 dB A at -3, 0, +3 dB SNRs.

For the Melodic Contour Identification (MCI), participants were asked to identify the pitch contour of a given sequence while the timbre varied amongst being piano notes, the same word, different words that were fixed across sequences (i.e., trials), or randomly selected words. There were nine different pitch contour choices. For the sentence recognition conditions participants were asked to identify five random words of 50 possible options (i.e., processing of the speech timbre), presented in a sentence with a syntax structure of name-verb-number-color-clothing with different pitch contours. Each condition contained 27 trials. Scores were calculated by percent correct. Both the MCI and the sentence recognition tasks were presented in quiet. The sentence recognition tasks were also tested at different SNRs (listed above).

After the practice runs were finished the MCI conditions were tested. There were four subtests in the MCI conditions, which were – Piano, Fixed Word, Fixed Sentence, and Random Sentence. There were nine different melodic contours in which the stimulus could be presented and the participant made a choice from: flat, rising, falling, rising-falling, rising-flat, flat-rising, falling-rising, falling-flat, and flat-falling (as shown in Figure 1).

Figure 1

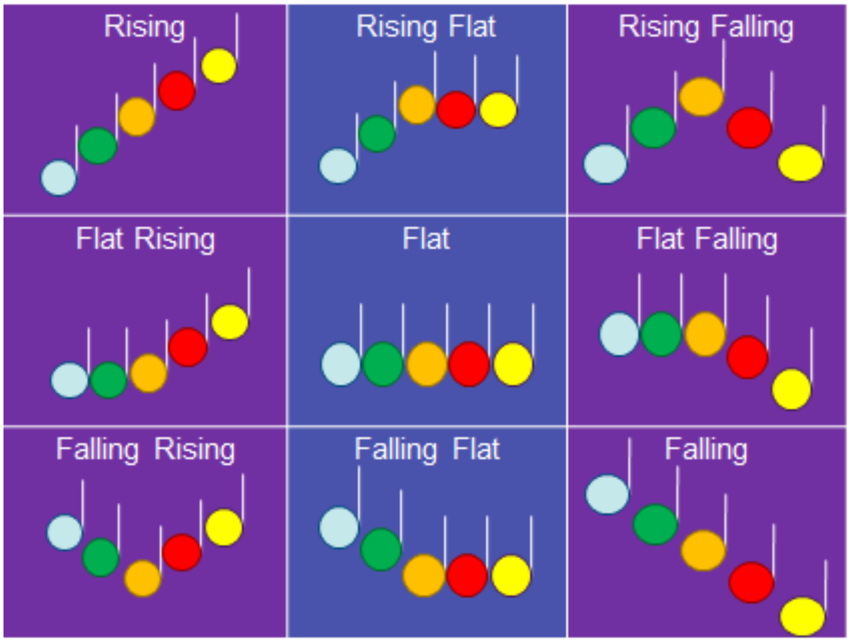
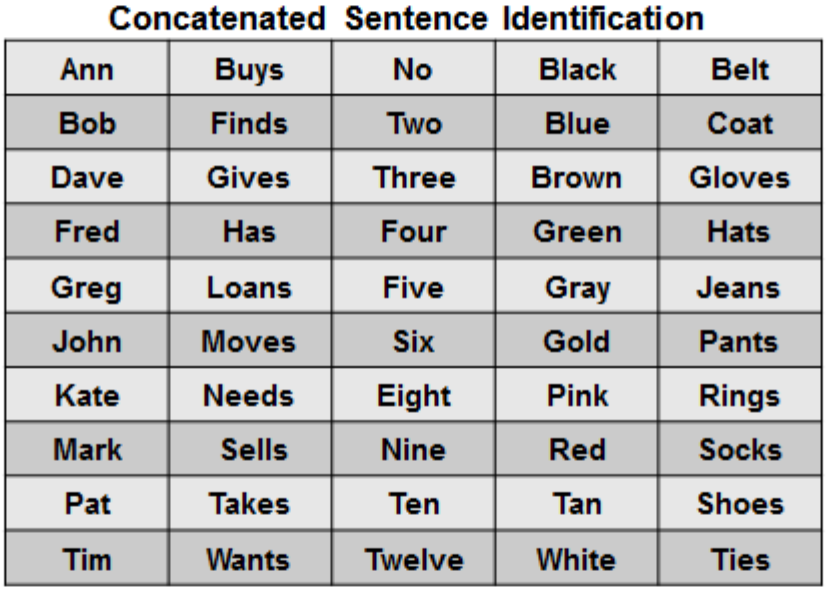


Figure 1 – Melodic Contour Identification choices. Participants were presented with the stimuli (5 piano notes or 5-word sentences with varying pitch contours) and asked to choose the pitch contour that matched to the best of their ability from the choices listed.

The Piano condition used synthesized piano notes as the stimuli being played in the different melodic contours. Fixed Word used the same word (Bob) said in the different melodic contours. Fixed Sentence used a five-word sentence said in different melodic contours and the Random Sentence used five random words, which were said in the different melodic contours. Each of these conditions: Piano, Fixed Word, Fixed Sentence, and Random Sentence were completed once each in a random order. The participant was asked to say their choice out loud and the experimenter would select their choice for them. This was done for efficiency. The experimenter was not able to hear the stimuli in order to reduce experimenter bias.

Once the MCI conditions were completed, the Sentence Recognition conditions were started. The Sentence Recognition conditions asked the participant to listen to a five-word sentence spoken whilst the pitch contour was being changed. The choices that were able to be selected are illustrated in Figure 2.

Figure 2



Concatenated Sentence Identification

Ann	Buys	No	Black	Belt
Bob	Finds	Two	Blue	Coat
Dave	Gives	Three	Brown	Gloves
Fred	Has	Four	Green	Hats
Greg	Loans	Five	Gray	Jeans
John	Moves	Six	Gold	Pants
Kate	Needs	Eight	Pink	Rings
Mark	Sells	Nine	Red	Socks
Pat	Takes	Ten	Tan	Shoes
Tim	Wants	Twelve	White	Ties

Figure 2 – Sentence Recognition choices listed in a 5 X 10 matrix. Participants were asked to listen to the stimuli (5 word sentence) and choose one word from each column to match the stimuli to the best of their ability.

Unlike the MCI conditions the participants did not have to identify the pitch contour, rather needed to identify the sentence being presented. These conditions were further manipulated to be presented in speech-shaped noise at different Signal-to-Noise Ratios (SNRs). The different SNRs tested were: quiet, +3, 0, and -3 dB. The competing noise was routed through the same headphone that the speech was presented.

The three different subtests of the Sentence Recognition conditions were Spoken, Random, and Flat. The ‘Spoken’ subtest used a normal speech-like utterance, which resembles everyday spoken speech as the pitch contour. The ‘Random’ subtest used a random pitch contour from the nine different options as shown in the earlier MCI conditions. The ‘Flat’ subtest used a flat, constant pitch contour across all of the stimuli.

3. Results

3.1 Sentence Recognition in Quiet

The left panel in Figure 3 illustrates average sentence recognition scores in the three different pitch contour conditions for the data collapsed across both age groups and musical experience groups. On average, in quiet, participants scored in the ranges of 78.9% (SE, 2.2%), 80.1% (3.1%), and 86.4% (2.5%) in the three pitch contour conditions—Flat, Random, and Spoken, respectively. The right panel of Figure 3 shows the average sentence recognition scores in three pitch contour conditions for the older and younger groups of participants with data collapsed across musical experience groups. On average, the older group scored 91.1% (2.6%), 90.6% (3.9%), and 94.4% (3.0%) respectively in the flat, random, and spoken pitch contour conditions, while the younger group scored 66.7% (3.4%), 69.6% (5%), and 74.7% (3.9%).

Figure 3

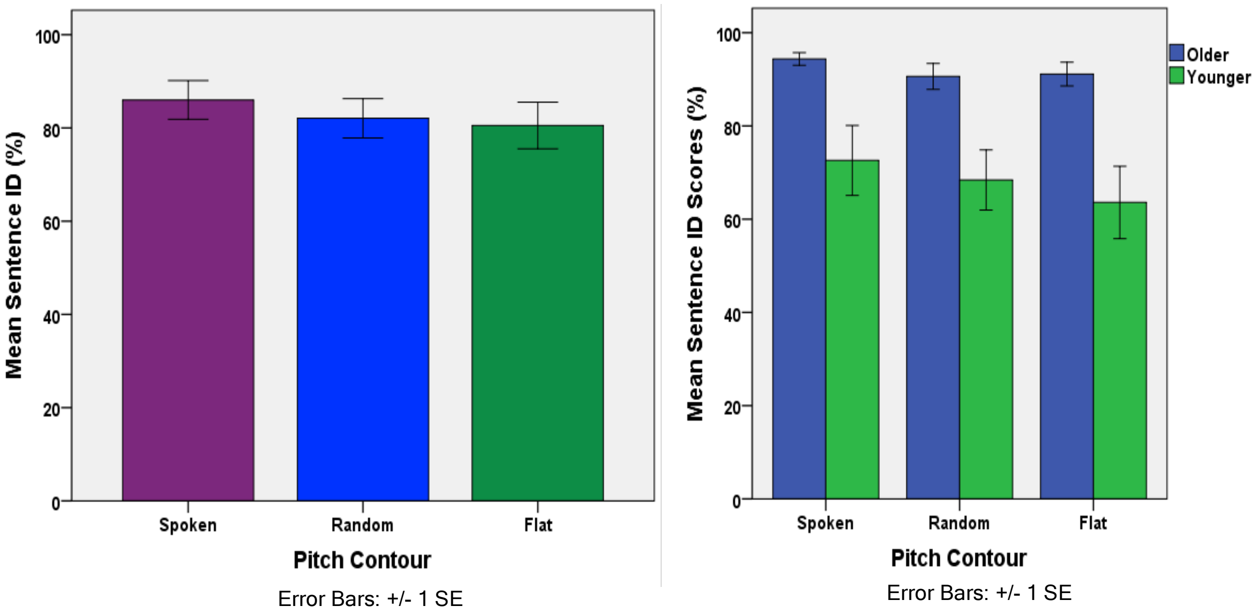


Figure 3. Sentence Recognition performance in Quiet: Left panel – Average sentence recognition score for the 3 different pitch contour conditions for the data collapsed across both age groups and musical experience groups. Right panel – Average sentence recognition scores in the 3 different pitch contour conditions between Older and Younger groups.

Repeated measures analyses of variance (RM-ANOVA) were performed with the dependent variable of sentence recognition score in quiet. The independent variable was a within-subject factor of pitch contour condition (Random, Flat, and Spoken) and between-subject factors of age group (Y- 7-9 yo and O- 10-16 yo), and musical experience (Musical experience—M and No Musical experience--NM). Significant effect of pitch contour condition on sentence recognition was not found (F (2, 18) = 2.344, p=.125). With respect to the effects of between-subject factors, the older age group scored significantly higher than the younger group [F (1, 9) = 27.213, p = .001] There was also a significant effect of musical experience on correct sentence identification versus non-musician [F (1,9) = 6.233, p = .034.] A significant interaction was found between Age Group and Musical Experience [F (1,9) = 6.536, p = .031]. This interaction was shown between the effects of musical experience on the two different age groups and can be seen visually in Figure 4. The 'O' group's sentence recognition scores in quiet were comparable regardless of musical training. On the other hand, the 'Y' group performed significantly different with the musically experienced members in the 'Y' group scoring higher on average than their non-musically trained peers.

Figure 4

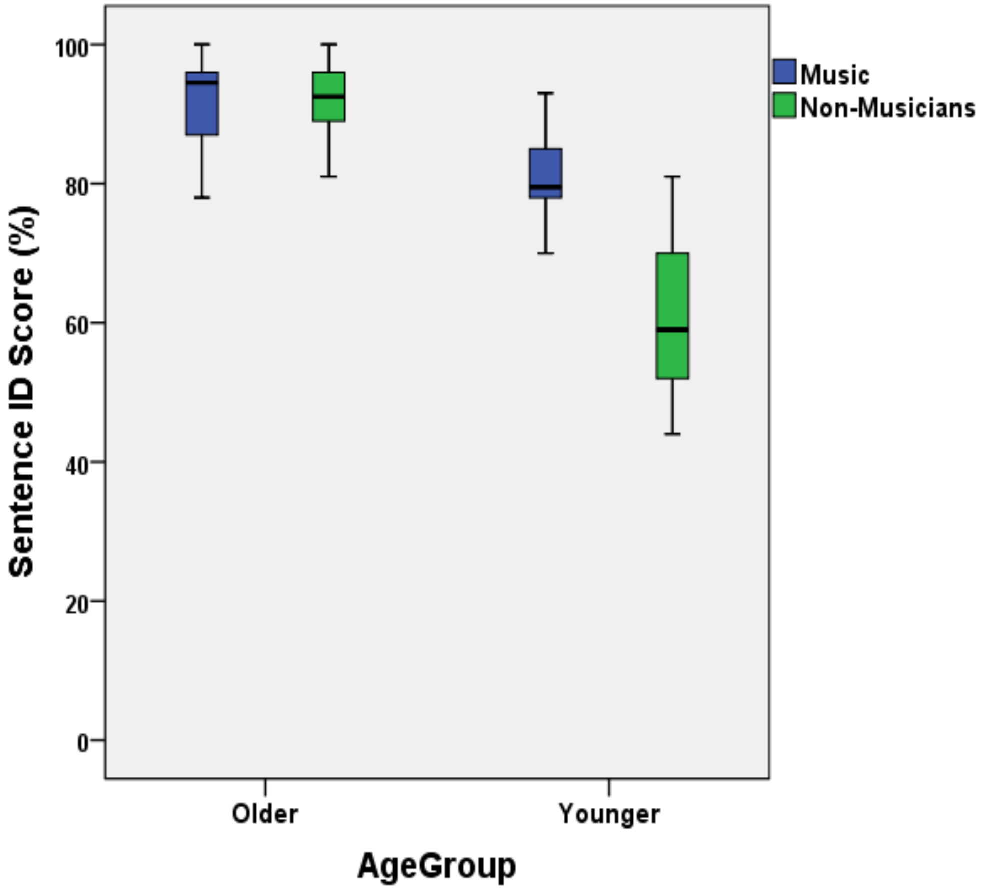


Figure 4. Box plots for sentence recognition in quiet by the Older and Younger children divided between musicians and non-musicians for each group. The boxes show the 25th and 75th percentile, the error bars show the 10th and 90th percentiles, and the solid line shows the median.

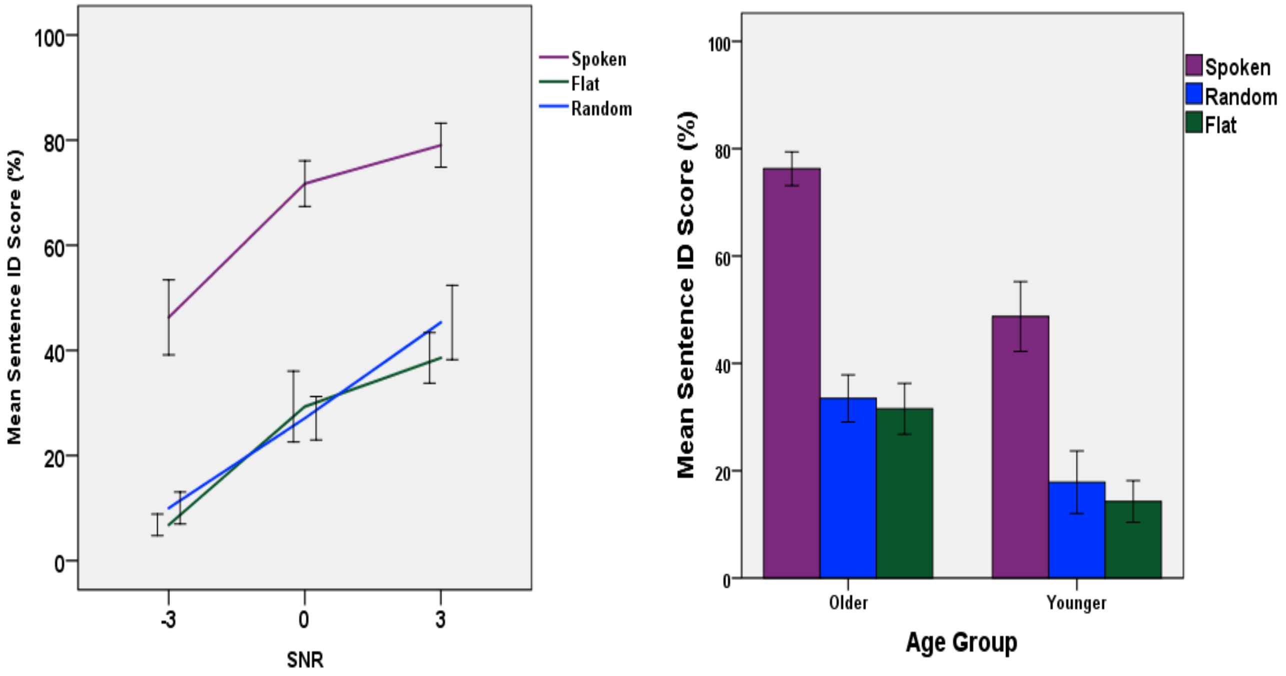
3.2 Sentence Recognition in Background Noise

The left panel in Figure 5 illustrates average sentence recognition scores as a function of SNR in the three different pitch contour conditions for the data collapsed across both age groups and musical experience groups.

On average, at -3 dB SNR, participants scored in the ranges of 5.8% (SE, 2.1%), 8.6% (3.2%), and 40.5% (3.8%) in the three pitch contour conditions—Flat, Random, and Spoken, respectively. For the 0 dB SNR condition participants scored in the ranges of 26% (SE, 6%), 24.2% (2.9%) and 68.8% (3.6%) in the respective three pitch contour conditions. Finally, for the +3 dB SNR condition participants scored in the ranges of 36.1% (SE, 3.6%), 45.6% (7%), and 77.2% (3.5%) for the aforementioned three pitch contour conditions.

The right panel of Figure 5 shows the average sentence recognition scores in three pitch contour conditions for the older and younger groups with data collapsed across musical experience groups. On average, the older group scored 31.5% (2.4%), 33.5% (4.5%), and 76.3% (3.2%) respectively in the Flat, Random, and Spoken pitch contour conditions, while the younger group scored 13.7% (3.1%), 18.9% (5.8%), and 48.1% (4.1%)

Figure 5



Error Bars: +/- 1 SE

Error Bars: +/- 1 SE

Figure 5. Sentence Recognition Performance in Background Noise: Left panel – Average sentence recognition score in noise for the 3 different pitch contour conditions for the data collapsed across both age groups and musical experience groups. Right panel – Average sentence recognition scores in noise in the 3 different pitch contour conditions between Older and Younger groups.

A RM-ANOVA was performed with the dependent variable of sentence recognition score. The independent variable was within-subject factors of pitch contour condition (Flat, Random, and Spoken) and SNR. Between-subject factors were age group (Y & O) and musical experience (M and NM). Significant effect of pitch contour condition on sentence recognition was found [F (2, 18) = 110.969, p <.001. Pairwise comparisons (with the Bonferroni correction) showed that sentence recognition scores were significantly higher in the Spoken condition compared to the Random condition (p = < .01) or the Flat condition (p < .01). Significant effect of SNR on sentence recognition was found [F (2,18 = 51.889, p < .001]. Pairwise comparisons (with Bonferroni correction) also showed that sentence recognition scores were significantly different from each other (p < .001) with the +3 dB SNR condition being scored the highest followed by 0 dB SNR, and finally -3 dB SNR. With respect to the effects of between-subject factors, the 10-16 year old group scored significantly higher than the younger group [F(1, 9) = 20.185, p = .002.] There was not a significant effect of musical experience on correct sentence recognition versus non-musician [F (1,9) = 0.99, p =.761].

Three-way interactions were significant for pitch contour X age group X SNR [F(4,36) = 2.818, p = .039] which is represented visually in the left panel of Figure 6 and also for musical experience X age group X SNR [F(2,18) = 6.437, p=.008] which is represented visually in the right panel of Figure 6. The interaction of the pitch contour X age group X SNR interaction revealed that, as SNR increased, the older group improved their performance at a comparable rate amongst the three pitch contour conditions, while the younger group experienced a larger improvement with the naturally spoken sentences than with the sentences produced in the other two pitch contours as SNR increased from -3 to 0 dB. The other two SNRs (0 dB & 3 dB) showed that the improvement rates were comparable amongst the three pitch contours. The musical experience X age group X SNR interaction illustrated in the right panel of Figure 6 revealed that in the 'O' group, the musically-experienced participants improved their sentence recognition at a faster rate as SNR increased from -3 to 0 dB than their no-music-experience peers, while both musical experience groups performed comparably at +3 dB SNR. A different trend was observed for the 'Y' group: although both musical experience groups improved their sentence recognition at a comparable rate when the SNR increased from -3 to 0 dB, the musically-experienced group appeared to continue to make improvement as the SNR further increased to +3 dB while their no-music-experience peers group reached their ceiling performance at 0 dB SNR.

Figure 6

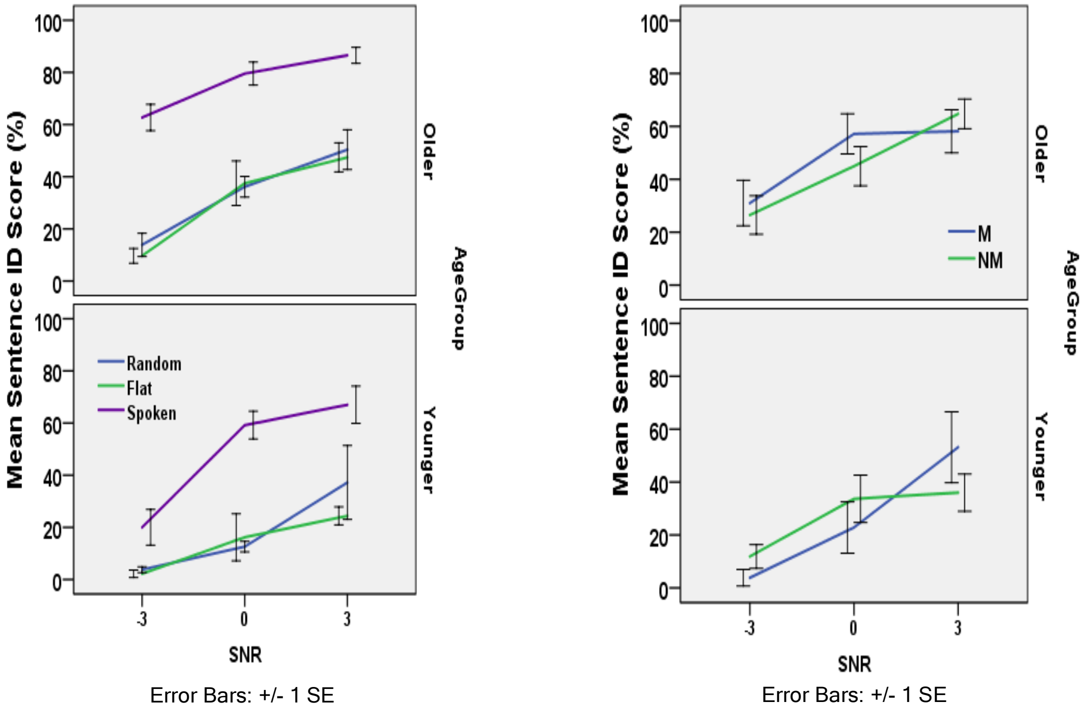


Figure 6. Sentence identification scores as a function of signal-to-noise ratio (SNR) for the two age groups in different pitch contour conditions (left panel) and with different musical training experiences (right panel).

All the remaining two-way, three-way, and four-way interactions were not found significant (all p values > .05).

3.3. Melodic Contour Identification

A RM-ANOVA was performed with the dependent variable of MCI score. The independent variable included within-subject factors of timbre complexity: Piano, Fixed Word, Fixed Sentence and Random Sentence and between-subject factors of age group (‘Y’ and ‘O’) and musical experience (M and NM). For the within-subject factors, significant effect of speech timbre conditions on MCI was found [F (3, 27) = 9.521, p < .001.] This is shown in the left panel of Figure 7. Pairwise comparisons (with the Bonferroni correction) showed that MCI scores were higher when asked to identify musical notes represented by piano keys rather than in the Fixed Word (p = .039) or Random Sentence condition (p = .029), but not different than the Fixed Sentence condition (p =.476). With respect to the effects of between-subject factors, the 10-16 year old group scored higher than the 7-9 year old group [F(1, 9) = 12.969, p = .006] seen visually in the right panel of Figure 7, while no effect of musical experience was revealed [F(1,9) = 1.708, p = .224]. All interactions amongst the factors were also found not significant (all p values > .05).

Figure 7

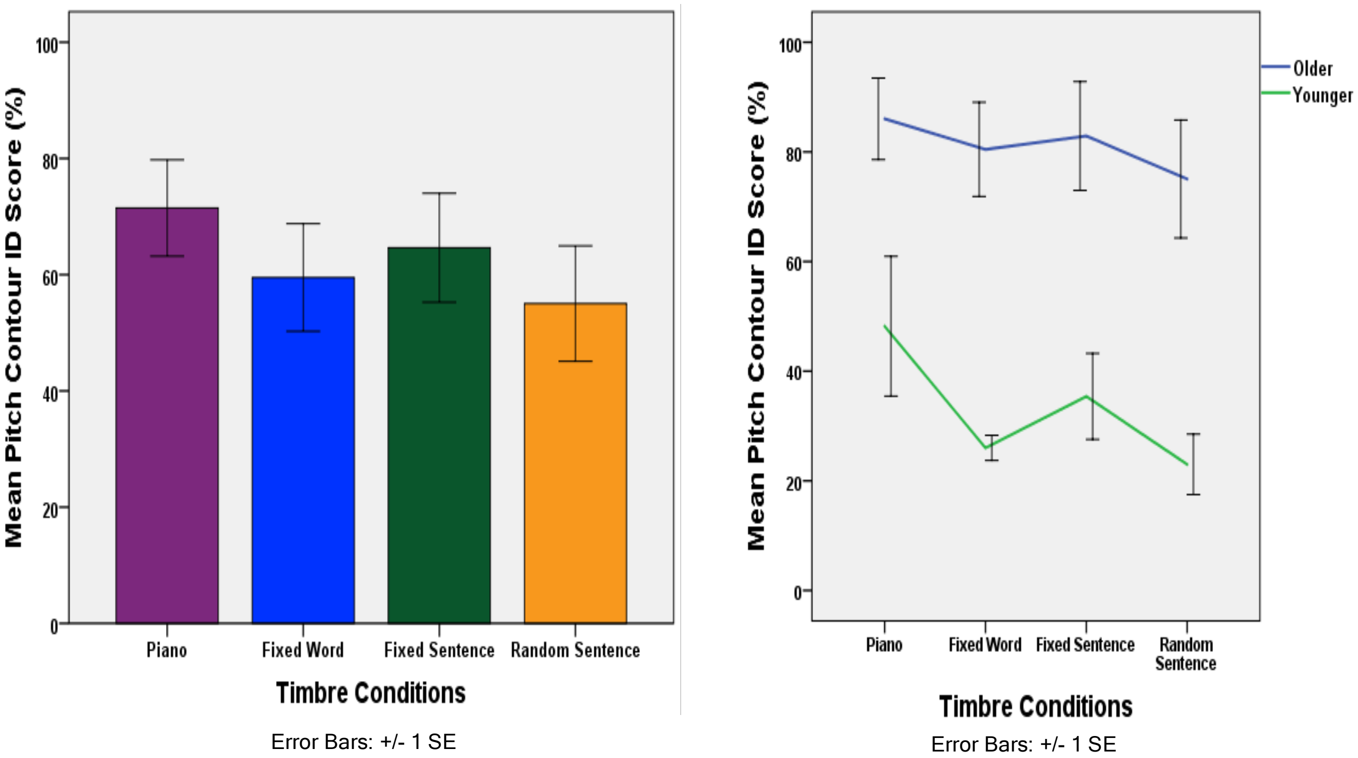


Figure 7. Melodic contour identification scores across the four timbre conditions: Left panel – Overall performance of MCI identification for the four timbre conditions. Right panel – Performance of MCI identification showing the performance difference between the Older and Younger groups.

4. Discussion

4.1 Sentence Recognition in Quiet

The results from the sentence recognition in quiet condition was broken down between the different pitch contour conditions (Random, Flat, Spoken) and there was no significant difference between the performances, although there was a trend that showed that higher scores were achieved for conditions that contained stimuli similar to naturally spoken speech. Crew et al. (2015) showed no significant difference between pitch contour conditions in NH adult listeners, even for the most difficult or unnatural pitch contour condition where they all scored nearly perfect, which appeared to be similar to our study, which showed that the Spoken condition was similar to the other two pitch contour conditions .This finding indicates that pitch variations were not large enough to negatively affect sentence recognition in quiet using the SSC.

The results of this study also showed that the older group scored higher on average than the younger group on all sentence recognition tasks regardless of pitch contour condition. This could be due to a developmental effect on working memory causing the younger group to perform more poorly than the older group. In a study by Linares, Bajo, & Pelegrina (2016) they examined the possible age-related changes throughout childhood and adolescence of working memory. They recruited 96 participants broken up into four different age groups (n=24 per group): 8-9 year olds, 11-12 year olds, 14-15 year olds, and the fourth group was composed of university students. These participants were asked to memorize different parts of the stimuli based on the different conditions which consisted of tasks pertaining to substitution, transformation, and retrieval separately and then combined in some conditions. Results suggested that the younger age groups, 8 year olds specifically, showed lower accuracy than the other age groups throughout the conditions. This study suggests that age-related differences in working memory are present especially when accessing information outside of the focus area for younger children. Thus, a follow-up study may be conducted to investigate the effect of working memory on the identification of the sentence in the SSC.

4.2 Sentence Recognition in Noise

Sentence recognition in noise was tested at three different SNRs (-3, 0, and +3 dB). The results were separated between the different pitch contour conditions (Random, Flat, and Spoken) and revealed a significant effect of sentence recognition in background noise for the different pitch contour conditions. Scores were significantly higher for the 'Spoken' condition than either the 'Random' or 'Flat' condition. This would suggest that once background noise is introduced into the task, natural production of the pitch contour helps participants to be able to correctly identify the correct sentence. This result coincides with the Miller, Schlauch, & Watson (2010) study which showed deviations from a typically intonated F0 contour pattern, like the natural 'Spoken' condition, has a deleterious effect on speech understanding in noise.

An age effect was also seen for the 'O' group over the 'Y' group. There was a natural-over-unnatural benefit, which described less improvement of sentence recognition with pitch contours changing from unnatural conditions to the naturally spoken condition that was smaller for the Y group than for the O group (see Figure 6 - left panel). This difference is largely attributed to the limited natural-over-unnatural benefit at -3 dB SNR for the Y group as opposed to the large natural-over-unnatural benefit at higher SNRs. In contrast, the O group experienced a constantly large natural-over-unnatural benefit at all SNRs (see Figure 6 - left panel). This finding indicates that, when the auditory information is degraded to a given extent, the Y group's ability to use the global pattern, such as pitch contour, for speech recognition may be disproportionally deteriorated.

4.3 Effect of timbre complexity on MCI

For the Melodic Contour Identification (MCI) tasks there was a significant effect of speech timbre complexity on correct identification of these tasks. Comparisons between performances on the four different timbre complexity tasks showed that variations in timbre affected participant's ability for correct MCI identification. These findings are consistent with results from the Crew et al., 2015 study that showed that performance declined as the timbre complexity increased for non-musicians. The 'Piano' condition produced the best results (69.1%), which was significantly different from the 'Fixed Word' (53.5%) and the 'Random Sentence' (49.6%) however did not differ from the 'Fixed Sentence' (60.5%) condition. These results appear to be comparable to what has been found in NH adult listeners in the Crew et al. (2015) study.

Our study did find that there was a significant difference in performance between the two age groups with the 'O' group scoring significantly higher than the 'Y' group. This suggests that there may be an age effect on melodic contour identification as the timbre complexity increases, which was consistent with the Halliday et al. (2008) findings that said children are susceptible to pitch contour changes until they reach around 11 years old and their pitch contour ability becomes more adult like.

4.4. Effect of musical training on sentence recognition and MCI

Benefits of musical training have been shown to help children recognize the sentences in both quiet and noise using the SSC. This finding differs from the results in NH adult listeners (Crew, Galvin, & Fu, 2015), for the quiet condition. The Crew et al (2015) study showed that musical training had no significant effect on performance of sentence recognition compared to no musical training for NH adult listeners. In the current study, the interaction between musical experience and age group was found significant. The older group (O) performed similarly on all conditions, regardless of musical training, which was more adult-like. On the other hand, the younger group (Y), with musical training, performed significantly better, on the sentence recognition tasks in quiet, compared to their non-musically trained peers. This suggests that there may be a developmental effect on timbre recognition that musical training may correct for, until children reach a certain age, in our case 10 years old. After 10 years old it seems that regardless of musical training, children perform similarly for the timbre recognition task.

The benefit of musical training for sentence recognition in the presence of noise is illustrated in the right panel of Figure 6 which shows that, in contrast to the quiet condition, both older and younger groups with musical training experienced such benefit. It appears that musical training facilitated older children at lower SNR's, suggesting its potential benefits for speech recognition in more challenging listening environment. At the highest SNR, the listening environment became less challenging for which the older children without musical training were able to perform comparably to those with musical training. For the younger group, however, children with musical training achieved higher sentence recognition scores than those without such training only at the highest SNR.

This may be due to the nature that listening environment was substantially challenging at lower SNRs and that the amount of musical training that the younger group had received did not reach the level to improve the sentence recognition scores at lower SNRs until the highest SNR.

While the benefit of musical training on MCI was found when piano notes were used as stimuli, such benefit was not found significant when spoken words were used to carry the melodic contour. These results differ from the findings on NH adults in Crew et al. (2015) wherein adult musicians scored significantly higher than adult non-musicians in the MCI task when the timbre complexity was varied across the four conditions (i.e., piano notes, fixed word, fixed sentence, and random sentences). This child-to-adult difference may be attributed to the fact that the musically-trained children have less duration of training than the adult musicians; longer duration of musical experience has been shown to provide a more robust benefit in cognitive tests and speech perception tests (Parbery-Clark, Skoe, Lam, & Kraus, 2009). Thus, while both were musically trained, due to the less musical experience, children may have not fully developed the capability of exploiting cues that adults are able to extract from varying words (i.e., timbre variations) to facilitate pitch contour identification,

4. Conclusion

In this study, music and speech perception were measured in NH children using the SSC. Speech perception was tested while the pitch contours of the sentences were flat or artificially variable across trials, or naturally produced. The music perception was tested using the MCI task while the stimulus sequences were piano notes, fixed words, fixed sentence, and random sentences. Major findings include:

a. Sentence recognition in noise was significantly poorer when speech contour was unnatural, suggesting susceptibility to the atypical speech patterns associated with sung speech.

b. MCI performance was poorer with spoken (word or sentence) stimuli, suggesting interference between timbre and pitch cues for melodic pitch perception

c. MCI and speech performance was significantly poorer for children younger than 10 years of age than for children 10 years of age and older for MCI performance and sentence recognition in both quiet and noise.

d. Children with music training performed significantly better for sentence recognition in quiet and in the MCI task.

e. Pitch and timbre cues were shown to interfere with each other in child listeners, depending on the listening demands. Music training can improve overall speech and pitch perception.

References

Allen, E. J., & Oxenham, A. J. (2014). Symmetric interactions and interference between pitch and timbre. *The Journal of the Acoustical Society of America, 135*(3), 1371-1379. doi: 10.1121/1.4863269

Ardoint, M., Agus, T., Sheft, S., & Lorenzi, C. (2011). Importance of temporal-envelope speech cues in different spectral regions. [Research Support, Non-U.S. Gov't]. *The Journal of the Acoustical Society of America, 130*(2), EL115-121. doi: 10.1121/1.3602462

Assmann, P. F., & Summerfield, Q. (1990). Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies. *The Journal of the Acoustical Society of America, 88*(2), 680-697. doi: http://dx.doi.org/10.1121/1.399772

Binns, C., & Culling, J. F. (2007). The role of fundamental frequency contours in the perception of speech against interfering speech. [Article]. *The Journal of the Acoustical Society of America, 122*(3), 1765-1776. doi: 10.1121/1.2751394

Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. *The Journal of the Acoustical Society of America, 118*(1), 471-482.

Carey, M., Parris, E., Bennett, S., & Lloyd-Thomas, H. (1997). A Comparison Of Model Estimation Techniques For Speaker Verification. 1997 *IEEE International Conference On Acoustics, Speech, And Signal Processing,* 1083-1086. doi:10.1109/ICASSP.1997.596129

Carlson, R., &: Granstrom, B. (1979). Model predictions of vowel dissimilarity. *Kunhl Tekniska Hogskolan: Speech Transmission Laboratories-Quarterly Progress and Status Report,* 3/4, 84-104.

Chatterjee, M., Zion, D. J., Deroche, M. L., Burianek, B. A., Limb, C. J., Goren, A. P., Christensen, J. A. (2015). Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hearing research, 322*, 151-162. doi: http://dx.doi.org/10.1016/j.heares.2014.10.003

Crew, J. D., Galvin, J. J., & Fu, Q.-J. (2015). Melodic contour identification and sentence recognition using sung speech. *The Journal of the Acoustical Society of America, 138*(3), EL347-EL351. doi: 10.1121/1.4929800

Crew, J. D., Galvin, J. J., & Fu, Q.-J. (2016). Perception of Sung Speech in Bimodal Cochlear Implant Users. *Trends in Hearing, 20*, 2331216516669329. doi: 10.1177/2331216516669329

Crew, J. D., John J. Galvin, I., & 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. doi: 10.1121/1.4758770

Darwin, C. J. (2008). Listening to speech in the presence of other sounds. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 363*(1493), 1011-1021. doi: 10.1098/rstb.2007.2156

Delattre, P. C., Liberman, A. M., Cooper, F. S., & Gerstman, L. J. (1952). An experimental study of the acoustic determinants of vowel colour; Observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word*, *8*, 195-210.

Drullman, R., & Bronkhorst, A. W. (2004). Speech perception and talker segregation: Effects of level, pitch, and tactile support with multiple simultaneous talkers. *The Journal of the Acoustical Society of America, 116*(5), 3090-3098. doi: http://dx.doi.org/10.1121/1.1802535

Elliott, T. M., Hamilton, L. S., & Theunissen, F. E. (2013). Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. *The Journal of the Acoustical Society of America, 133*(1), 389-404.

Galvin, J. J., 3rd, Fu, Q. J., & Oba, S. I. (2009). Effect of a competing instrument on melodic contour identification by cochlear implant users. *The Journal of the Acoustical Society of America, 125*(3), EL98-103.

Goswami, U., Fosker, T., Huss, M., Mead, N., & Szűcs, D. (2011). Rise time and formant transition duration in the discrimination of speech sounds: the Ba–Wa distinction in developmental dyslexia. *Developmental science, 14*(1), 34-43.

Grey, J. M. (1975). An exploration of musical timbre: using computer-based techniques for analysis, synthesis and perceptual scaling. Ph.D. dissertation, 1975. Stanford University.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America, 61*(5), 1270-1277.

Halliday, L. F., Taylor, J. L., Edmondson-Jones, A. M., & Moore, D. R. (2008). Frequency discrimination learning in children, *The Journal of the Acoustical Society of America*. 123, 4393–4402.

Jamieson, D. G., Kranjc, G., Yu, K., & Hodgetts, W. E. (2004). Speech intelligibility of young school-aged children in the presence of real-life classroom noise. *Journal of the American Academy of Audiology, 15*(7), 508-517.

Keurs, M., Festen, J. M., & Plomp, R. (1992). Effect of spectral envelope smearing on speech reception. I. *The Journal of the Acoustical Society of America, 91*(5), 2872-2880. doi: 10.1121/1.402950

Klatt, D. H. 1982.‘‘Prediction of perceived phonetic distance from critical-band spectra: A first step,’’ IEEE ICASSP, 1278–1281.

Kong, Y.-Y., Mullangi, A., Marozeau, J., & Epstein, M. (2011). Temporal and Spectral Cues for Musical Timbre Perception in Electric Hearing. *Journal of Speech, Language, and Hearing Research, 54*(3), 981-994. doi: 10.1044/1092-4388(2010/10-0196)

Linares, R., Bajo, M. T., & Pelegrina, S. (2016). Age-related differences in working memory updating components. *Journal of Experimental Child Psychology, 147*, 39–52.

Macherey, O., & Delpierre, A. (2013). Perception of musical timbre by cochlear implant listeners: a multidimensional scaling study. *Ear & Hearing, 34*(4), 426-436.

Massar, M.L., Fickus, M., Bryan, E., Petkie, D., Terzuoli Jr., A. (2011) Fast computation of spectral centroids. Advanced Computational Math 35-83. doi:10.1007/s10444-010-9167-y

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychological Research, 58*(3), 177-192.

McDermott, J. H., & Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology, 18*(4), 452-463. doi: http://dx.doi.org/10.1016/j.conb.2008.09.005

Miller, S. E., Schlauch, R. S., & Watson, P. J. (2010). The effects of fundamental frequency contour manipulations on speech intelligibility in background noisea). *The Journal of the Acoustical Society of America, 128*(1), 435-443. doi:http://dx.doi.org/10.1121/1.3397384

Molis, M. R. (2005). "Evaluating models of vowel perception," *The Journal of the Acoustical Society of America, 118,* 1062‑1071. doi: https://doi.org/10.1121/1.1943907

Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing* (5th ed.) (pp. 270-273). San Diego, CA: Academic Press.

Murray, I.R., & Arnott, J.L. (1993). Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, *93 (2)*, 1097–1108.

Nelson, P. B., & Soli, S. (2000). Acoustical Barriers to LearningChildren at Risk in Every Classroom. *Language, Speech, and Hearing Services in Schools, 31*(4), 356-361.

Neuman, A. C., Wroblewski, M., Hajicek, J., & Rubinstein, A. (2010). Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults. *Ear & Hearing, 31*(3), 336-344. doi: 10.1097/AUD.0b013e3181d3d514

Oxenham, A. J., Bernstein, J. G. W., & Penagos, H. (2004). Correct tonotopic representation is necessary for complex pitch perception. *Proceedings of the National Academy of Sciences of the United States of America, 101*(5), 1421-1425. doi: 10.1073/pnas.0306958101

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear & Hearing, 30*(6), 653-661.

See, R. L., Driscoll, V. D., Gfeller, K., Kliethermes, S., & Oleson, J. (2013). Speech intonation and melodic contour recognition in children with cochlear implants and with normal hearing. *Otology & Neurotology, 34*(3), 490-498.

Stalinski, S. M., Schellenberg, E. G., & Trehub, S. E. (2008). Develop- mental changes in the perception of pitch contour: Distinguishing up from down. *The Journal of the Acoustical Society of America*, 124, Swanepoel, R., Oosthuizen, D. J. J., and Hanekom, J. J. (2012). The relative importance of spectral cues for vowel recognition in severe noise. *The Journal of the Acoustical Society of America, 132*, 2652‑2662.

Tao, D., Deng, R., Jiang, Y., Galvin, J. J., 3rd, Fu, Q. J., & Chen, B. (2015). Melodic pitch perception and lexical tone perception in Mandarin-speaking cochlear implant users. *Ear & Hearing, 36*(1), 102-110.

Vihman, M. M. (2014). *Phonological development : the first two years*: Chichester, England : Wiley-Blackwell, 2014. Second edition.

Appendices

Appendix I. Extended Literature Review

Pitch and Timbre Perception

Perceived pitch in spoken English is identifiable by the fundamental frequency (F0), which allows for speaker identification (Carey, Parris, Lloyd-Thomas & Bennett, 1997), intent (Grant, 1996), and emotion (Murray & Arnott, 1993).

Speaker identification, studied by Carey et al. (1997) revealed the possibility to distinguish individual speakers by using simple parameters such as the mean and the variance of the pitch period in voiced sections of an utterance. The study discovered that gender identification could be indicated correctly 98% of the time by using the mean of the speaker's pitch alone, which led them to hypothesize that the speaker's mean pitch could be helpful in speaker identification. Using an Improved Multiband Excitation (IMBE) speech coder, a pitch estimation algorithm was used to extract values of pitch period for segments of speech marked as vowels by a pattern matching process in the classifier stage. Furthermore, the mean pitch was calculated by using samples of the pitch period found to be within +/- 35% of the initial estimate of the mean pitch value. This process was tested and found successful in speaker identification suggesting that the pitch of a speaker's voice, or the F0, is a useful tool in speaker identification.

Murray and Arnott (1993) discussed the importance of pitch in identifying speaker's emotion. They summarized that pitch is important in emotional expression and will differ in pitch range dependent on the emotion being expressed. Unemotional speech has a narrow pitch range compared to emotional speech which tends to be normally distributed about the average pitch level. Fundamental frequency (F0) and intensity increase in range, most notable at the high end of the range, as the emotional involvement of the speaker increases. Emotion was also classified into two different groups, passive and active, with the passive group having lower pitch compared to the active emotion group, which was characterized by high pitch. The study concluded that pitch contour is the most important parameter in differentiating between basic emotions.

The F0 contour has also been researched extensively showing that this contour can aid in segregating target speech from competing maskers. Darwin (2008) reported that there are two main reasons why a difference in F0 can improve the intelligibility of speech in the presence of competing noise. The first reason is that a difference in F0 improves the definition of the first-formant (F1) frequencies of two speakers compared to when the F0 is the same (Darwin, 2008). For example, two vowels with different F0s can be separated from one another because they create larger differences between the F1s than if the F0s of the vowels were the same. If those two vowels were /i/ with an F0 of 100 Hz and /a/ with an F0 of 140 Hz the definition of F1 would be different between the two vowels and could be identified easier. When the F0s are different the vowels are more discernable, but when they are not more errors are made on correct identification of the vowel. If the F0 difference is too small or summed, the harmonics of the vowels would be too close to be resolved by the cochlea. It is suggested that the F1 difference or a difference in the upper harmonics makes a stronger cue for discerning vowels. Another reason that the F0 can improve intelligibility in noise only applies when there are F0 differences of greater than four semitones. This is because the common harmonic series are grouped together which allows for differentiation of other groups of sounds with different F0s.

Moore (2008) defines timbre as the acoustic correlates that involve the distributions of energy over time and frequency such as features of the spectral or temporal envelope. These features have been widely studied in the musical field and timbre is regularly defined as an attribute that allows listeners distinguish instruments playing the same note with the same loudness and duration (Grey, 1975). Research has also shown that timbre plays a role in speech identification as well.

A study by Swanepoel et al. (2012) investigated the importance of formants in the presence of noise as it increases to severe levels and also to consider how important formants are, as well as the spectral shape, when identifying vowels in noise. Two Afrikaans speakers, one male or one female, produced words with the /p/-vowel-/t/ structure and substituted different vowels in between that were analyzed. Vowel identification was separated and tested in two different situations: when the whole-spectrum of the vowel was present and when only the formants of the vowel were present. This was further broken down by observing the percentage of correct identification when the complete spectrum of the vowel was present, when F1 was suppressed, and when F2 was suppressed. Identification of vowels, in quiet and in the presence of noise, was largely affected by whether the whole-spectrum or only the formants were present. Whole-spectrum representation of vowels was found to be more important as the SNR decreased. For example, at -5 dB SNR there was no significant difference of vowel identification scores in either the whole-spectrum or formant-only conditions. However, at the -10 dB SNR level, when the whole spectrum shape was present participants did significantly better at identifying the vowel than when only the formant information was present. This suggests that speech timbre in the form of the whole spectrum of the target needs to be more complex to be identified correctly in poorer SNR conditions than in quiet. While in quiet, vowel identification was sufficient relying solely on the formant information provided by F1 and F2, when very poor SNR levels are present the auditory system may rely more heavily on more complex speech timbre representation of the signal rather than the formants-only representation.

Goswami et al. (2011) conducted a study to measure children's discrimination of phonetic contrast of Ba/Wa by varying the rate of formant frequency change in the formant transition region or by varying the rate of amplitude change over time (e.g., rise time of the temporal envelope). The study used 106 English speaking children, with normal hearing (<20 dB HL) between 7-12 years of age, who had no learning difficulties. The study showed that the shape of amplitude modulation for particular syllables also contains information important for phonetic discrimination, therefore a rise time—which is a physical property of timbre—deficit can affect consonant identification.

In conclusion, pitch and pitch contours provide suprasemental information and also cues for separation of target speech from competing makers, while timbre, and its acoustic correlates of timbre, are important for speech recognition by providing cues for the identification of segmental elements of speech, such as phonemes.

Pitch and Timbre Perception Variations on the Perception of the Other Attribute

Research has shown that variations in either pitch or timbre may affect the perception of the other attribute. While studies using speech stimuli have mainly examined the effect of variations of pitch contour on the processing of speech timbre (reflected by speech recognition) (e.g., Miller and Schlauch, 2010), using the same set of non-speech stimuli, Allen and Oxenham (2014) systematically studied the effect variations of one attribute on the perception of the other. This section will first review these two studies, and then discuss a very recent study using a same set of speech stimuli carrying various pitch contours to study the effect of variations in one attribute on the perception of the other.

Miller, Schlauch, & Watson (2010) conducted an experiment to investigate how F0 manipulations (four in total) affect speech intelligibility using sentences as stimuli. This study recruited fifteen paid listeners, all native English speakers, with normal hearing in their experiment. Low predictability sentences were used in the presence of background noise and participants were asked to identify their choice by saying it aloud as the examiner wrote their responses on a piece of paper. Results from the study showed that any unnatural F0 contour manipulation decreased speech understanding in background noise. The study also concluded that incorrect or misleading linguistic cues related to intonation have a more deleterious effect on speech understanding than speech comprised of plausible linguistic cues. This can be shown by using speech stimuli produced by an electrolarynx by applying a simple rising or falling intonation contour to speech segments that improve intelligibility in opposition to unnatural monotone speech stimuli.

In the study by Allen & Oxenham (2014) the effects of spectral shape variation on fundamental frequency discrimination and vice-versa were explored using non-speech stimuli. Their goals were to determine whether the interference and interactions between pitch and timbre are symmetric and whether the effects of musical training on subject's ability to ignore these variations when performing a discrimination task. They conducted three experiments with the first measuring basic sensitivity to small changes in either F0 or spectral centroid in the absence of variation in the non-target dimension. The spectral centroid provides a noise-robust estimate of how the dominant frequency of a signal changes over time and is thought of as one of the physical properties of timbre (Massar, Fickus, Bryan, Petkie, and Terzuoli, 2010). Experiment 2 used individual differences limens (DLs) measured in experiment 1, to examine the effects of random variations in either F0 or spectral centroid on listener's ability to discriminate small changes in the other dimension. The third experiment provided a direct test of perceptual symmetry of the two dimensions by measuring performance in both dimensions using stimuli that varied by the same amount in terms of DLs obtained from the individual subjects. The first experiment found that F0 DLs were better in musicians than non-musicians; however the DLs for spectral centroid were not significantly different between the two groups. Results from the second experiment showed that discrimination thresholds in either F0 or spectral centroid were impaired by random variations in the non-target dimension and that the amount of interference was similar for the two dimensions regardless of musical training. The third experiment concluded that individual performance was better when the interference was varied coherently with the target than when varied in the opposite direction. This suggests that listeners sometimes confuse changes across the two dimensions. It was also shown that musicians were no less susceptible to this than non-musicians. The study ultimately suggested that judgments in pitch and timbre (in terms of F0 and spectral centroid, respectively) are similarly affected by random variations in the other dimension, suggesting a relatively symmetric process. Results were similar to the Miller, Schlauch, and Watson (2010) study concluding that changes in pitch and timbre can affect the processing of timbre except the Miller, Schlauch, and Watson (2010) study used speech stimuli instead of non-speech stimuli. Allen and Oxenham (2014) noted that timbre variations could also affect the pitch processing. In addition, the Allen and Oxenham (2014) study showed that there is not a strong musical training effect for interference effects in either dimension.

Most recently, a study by Crew et al. (2015) examined the effect of variations of speech timbre on the identification of pitch contour and vice versa between adult musicians and non-musicians with normal hearing. Their study recruited 16 normal hearing subjects who were divided into two groups – musicians and non-musicians. Musicians were defined as regularly playing a musical instrument at the time of recruitment and non-musicians were defined as never having any formal musical training or never informally learning to play an instrument. The Sung Speech Corpus was used in this study, which consists of 50 sung monosyllabic words produced by a single adult male with the following syntax: "name" "verb" "number" "color" "clothing." Each category contains 10 words, and each word can be sung at 13 different pitches from A2 (110 Hz) to A3 (220 Hz). This allows for a five-word sentence to be constructed with a five-note melody, allowing sentence recognition and melodic contour identification (MCI) to be measured using the same set of stimuli. The SSC also included natural speech utterances for each word to allow for comparisons between the sung speech and the naturally produced speech. Their study concluded that there was no significant musician effect for the sentence recognition tasks, possibly due to ceiling effects; however there was a significant musician effect for the Melodic Contour Identification (MCI) tasks, which became stronger as the tasks became more complex. The musician group performed nearly perfect for all test conditions suggesting that they were better available to extract pitch information despite the changing of timbre (in their case, words). Non-musicians were more affected by changes in timbre compared to their musician counterparts. This study also concluded that when timbre was constant, music notes or words, the non-musicians were better able to extract pitch information, compared to the conditions where timbre was more variable by randomly selecting words to construct a sentence stimulus.

Cochlear Implant User's Pitch and Timbre Identification

Pitch contour effects on timbre processing in normal hearing children and the effects of timbre complexity on identification of pitch contours are little researched. Pediatric CI users' could benefit from this information possibly bolstering the optimization of signal representation of pitch and timbre in the future.

In a study by McDermott & Oxenham (2008) they concluded that normal hearing adult listeners' pitch perception was possible due to robust cues from both temporal fine structure and harmonic resolution. Caclin et al. (2005) summarized that timbre perception is possible due to cues derived by the attack time, from the temporal envelope, and spectral centroid, a component of the spectral envelope.

Cochlear Implant (CI) users are similar to normal hearing (NH) listeners as they both rely on temporal and spectral envelopes for timbre perception. Kong et al. (2011) conducted a study to investigate timbre perception (musical timbre) using a multidimensional scaling technique to derive a timbre space. Their study compared 8 CI users' performances to 15 NH listeners using sixteen stimuli that synthesized western musical instruments. Each listener was asked to judge whether a pair of stimuli presented was similar or dissimilar. Acoustical analyses were performed to characterize the temporal and spectral characteristics of each stimulus in order to examine the psychophysical nature of each perceptual dimension. The study concluded that NH listeners had a timbre space that was best represented in three dimensions compromised of the temporal envelope (log-attack time), the spectral envelope (spectral centroid) and the spectral fine structure (spectral irregularity). However, two dimensions made up the timbre space for CI listeners: temporal envelope and weak signs of the spectral envelope.

This suggested that the temporal envelope was a dominant cue for timbre perception in CI users. The study also suggested that compared with NH listeners, CI users showed reduced reliance on both the spectral envelope and the spectral fine structure for timbre perception.

Pitch perception for CI users depends heavily on the spectral envelope, which was concluded by Crew, Galvin, & Fu (2012). Their goal was to investigate the effect of channel interaction on melodic pitch perception. In this study, twenty normal hearing subjects were asked to identify melodic contours that were made up of five musical notes. There were nine possible options for the melodic contours: “rising,” “falling,” “flat,” “rising-flat,” “falling-flat,” “rising-falling,” “falling-rising,” “flat-rising,” and “flat-falling.” There were two different conditions present: the first was unprocessed natural sounding speech. The second was vocoded CI simulations using sinewave carriers that simulated different amounts of channel interaction. Each subject was familiarized with the unprocessed stimuli for familiarization. Results showed that all subjects scored above 90% on the unprocessed stimuli tasks; however when the vocoded conditions were tested performance fell in relationship as the amount of channel interaction was increased. This suggests that the greater the amount of channel interaction the worse the melodic pitch perception will be. It was also shown that the amount of channel interaction and the CI signal processing itself weakens spectral envelope cues. This suggests that increasing the number of channels in the cochlear implant may not enhance spectral contrasts and in fact lead to more channel interaction causing weakened variance in the spectral envelope.

In short, while the spectral envelope plays a large role for timbre perception in NH listeners, CI users appear to not use this cue for timbre perception but rely more heavily than NH listeners on it for pitch perception. For CI listeners' pitch perception, such disproportionally higher reliance on the same cue (i.e., spectral envelopes) for timbre perception may generate interference of pitch and timbre perception.

Age effect for Pitch Identification

The effect of pitch and timbre variations on identification has been discussed at length for adults, but how does it relate to children? It is known that children can identify two different pitches as being the same or different by the age of 6 years old (Cooper, 1994), but identifying the pitch contour tends to be more difficult.

Stalinski, Schellenberg, & Trehub (2008) conducted two experiments. They first studied 26, five year old, normal hearing children with no history of musical experience to investigate their ability to identify pitch direction as well as to investigate the age-related changes in their ability to identify directional changes in pitch. 11 synthesized tones were presented, similar to piano timbre, with a fundamental frequency of 880 Hz. There were five higher and five lower tones around the F0 which were displaced in pitch by 4, 2, 0.5, and 0.3 semitones. Participants were asked to judge whether the second sound in a series of three sounds went up or down. When a visual cue was present the participants were nearly perfect in identifying the pitch changes of up from down. During the trials, the five years old were able to identify directional changes in pitch, after a few minutes of training.

Experiment 2, conducted by Stalinski, Schellenberg, & Trehub (2008) they included three different age groups of children: 29 six year olds, 30 eight year olds, 30 eleven year olds and 29 young adults. In this experiment no participants had musical experience. The results showed a significant age effect such that the 6 year olds performed significantly poorer than the other age groups. When the 6 year old group was excluded from the statistical analysis the significant effect of age disappeared. Consistent with other literature, this study showed that an 8 year olds' performance did not differ from that of adults suggesting that pitch resolution has reached adult like maturity by this age.

In a study by Halliday, Taylor, Edmondson-Jones, and Moore (2008) the pitch discrimination abilities of high versus low pitch was investigated between children and adults. They broke their subjects into four different groups: 6-7 year olds, 8-9 year olds, 10-11 year olds, and adults. All participants were trained on the task and screened to ensure that they could differentiate between 1 and 1.5 kHz. Non-verbal IQ was taken into account during statistical analysis as it is known to be associated with pitch discrimination. The results of this study showed that all child groups performed significantly worse than the adult group and that the youngest group (6-7 year olds) performed more poorly than the oldest child group tested (10-11 year olds). Their study concluded that pitch discrimination abilities continue to develop into childhood and generally will not reach an adult-like level until after 11 years old.

Two different studies concluding two different ages that pitch discrimination and pitch contour identification reaches adult like levels. One of the goals of the our current study will be to observe the possibility of the age-effect on the pitch and timbre perception as the other attribute is varied by separating our participants into two separate groups 7-9 years old and 10-16 years old.

Appendix II. Musical Experience Questionnaire

Please fill out this questionnaire to the best of your ability. If you have any questions feel free to ask for assistance. If a question does not pertain to you please answer with N/A.

Do you have musical experience?

What type of musical experience do you have? (Composing, playing an instrument, singing, etc.)

How many years of musical experience do you have?

At what age did you begin practicing and honing your musical ability?

Is there a family history of musical experience? If so, are those family members immediate of extended?

Have you ever taken music lessons? Private or through school? How long?

Were you classical trained as a musician or self-taught?

How often did/do you practice your musical skills? (daily, weekly, monthly, etc.) How many hours per practice session on average?

If you do play an instrument – what instrument do you play?

What genre of music do you prefer to listen to, perform, or compose?

Are there certain environments you practice in or listen to music that you enjoy more?

Can you sight read?

On a scale of 1 to 10 (1 being not confident; 10 being very confident) rate your musical ability.

On a scale from 1-10 (1 being not confident and 10 being very confident) rank your ability on discriminating pitches of tones in music.