Music and Speech Perception in Children Using Sung Speech

Yingjiu Nie[1], John J. Galvin III[2], Michael Morikawa[1], Victoria André[1], Harley Wheeler[1], and Qian-Jie Fu[3]

Abstract

This study examined music and speech perception in normal-hearing children with some or no musical training. Thirty children (mean age = 11.3 years), 15 with and 15 without formal music training participated in the study. Music perception was measured using a melodic contour identification (MCI) task; stimuli were a piano sample or sung speech with a fixed timbre (same word for each note) or a mixed timbre (different words for each note). Speech perception was measured in quiet and in steady noise using a matrix-styled sentence recognition task; stimuli were naturally intonated speech or sung speech with a fixed pitch (same note for each word) or a mixed pitch (different notes for each word). Significant musician advantages were observed for MCI and speech in noise but not for speech in quiet. MCI performance was significantly poorer with the mixed timbre stimuli. Speech performance in noise was significantly poorer with the fixed or mixed pitch stimuli than with spoken speech. Across all subjects, age at testing and MCI performance were significantly correlated with speech performance in noise. MCI and speech performance in quiet was significantly poorer for children than for adults from a related study using the same stimuli and tasks; speech performance in noise was significantly poorer for young than for older children. Long-term music training appeared to benefit melodic pitch perception and speech understanding in noise in these pediatric listeners.

Introduction

Pitch and timbre are important cues for both speech and music perception. For speech, pitch information can be used to segregate competing sound sources (Assmann & Summerfield, 1990; Brokx & Nooteboom, 1982; Darwin, 2008; Drullman & Bronkhorst, 2004), identify talkers (Carey, Parris, Lloyd-Thomas, & Bennett, 1996), perceive vocal emotion (Murray & Arnott, 1993), and understand lexical tones (Deutsch, Henthorn, & Dolson, 2004; Lin, 1988). Timbre information is especially important for perception of phonemes (Carlson, Granstrom, & Klatt, 1979; Goswami, Fosker, Huss, Mead, & Szűcs, 2011; Molis, 2005; Swanepoel, Oosthuizen, & Hanekom, 2012). For music, pitch is important for melody perception and segregation of multiple instruments. Timbre is important for instrument identification and for instrument segregation (Grey, 1977, 1978; McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995). Pitch is generally associated

with fundamental frequency (F0) and harmonic information, while timbre is associated with spectral and temporal envelope information. While pitch and timbre cues can be considered as independent (Marozeau, de Cheveigné, McAdams, & Winsberg, 2003; Marozeau & de Cheveigné, 2007), they are inexorably combined in many auditory objects. Distortions to pitch and timbre cues can interfere with auditory object perception

[1]Department of Communication Sciences and Disorders, James Madison University, Harrisonburg, VA, USA
[2]House Ear Institute, Los Angeles, CA, USA
[3]Department of Head and Neck Surgery, University of California—Los Angeles, CA, USA

Corresponding author:
Yingjiu Nie, Department of Communication Sciences and Disorders, James Madison University, 235 Martin Luther King Jr Way, MSC 4304, Harrisonburg, VA 22801, USA.
Email: nieyx@jmu.edu

(Crew, Galvin, & Fu, 2012, 2015; Fuller, Galvin, Maat, Free, & Başkent, 2014). Music training has been shown to provide advantages in both music and speech perception in adults (Kraus & Chandrasekaran, 2010; Parbery-Clark, Skoe, Lam, & Kraus, 2009; Schön, Magne, & Besson, 2004). Presumably, musicians are better able to perceive the cue of concern (pitch or timbre) in the presence of a possibly conflicting cue (Crew et al., 2015) or degraded listening conditions (Fuller, Galvin, Maat, et al., 2014; Fuller, Galvin, Free, & Başkent, 2014). In contrast, Allen and Oxenham (2014) found that musicians and nonmusicians may be similarly susceptible to interference between pitch and timbre cues. Other studies have shown only weak, inconsistent, or nonexistent musician effects for speech perception (Boebinger et al., 2015; Deroche, Limb, Chatterjee, & Gracco, 2017; Madsen, Whiteford, & Oxenham, 2017; Ruggles, Freyman, & Oxenham, 2014).

Most previous studies that have played pitch against timbre cues have focused on discrimination of F0 and spectral shape (Allen & Oxenham, 2014) or have used similarity ratings and multidimensional scaling to characterize the pitch and timbre space (Kong, Mullangi, Marozeau, & Epstein, 2011; Marozeau & de Cheveigné, 2007; Marozeau et al., 2003). Listeners in these studies were typically asked to compare short, simple stimuli (e.g., single notes), which may overestimate sensitivity to pitch or timbre cues relative to longer exemplars of running music or speech, as is encountered in everyday listening. In a melodic contour identification (MCI) task (Galvin, Fu, & Nogaki, 2007; Galvin, Qian-Jie, & Shannon, 2009), listeners must attend to changes in pitch direction over several notes, rather than focus on pitch differences only between two notes. In sentence recognition, listeners attend to the sequence of words strung together, rather than focusing only on differences in spectral and temporal envelopes associated with single phonemes. Thus, discrimination between relatively short stimuli may not predict listeners' ability to use pitch or timbre cues over a longer period as occurs in everyday listening (e.g., a piece of music, running speech). As speech perception tasks become more difficult (e.g., under conditions of noise, spectral-temporal degradation), musicians may also exhibit some advantage over nonmusicians (Fuller, Galvin, Maat, et al., 2014; Fuller, Galvin, Free, & Başkent, 2014; Kraus & Chandrasekaran, 2010; Parbery-Clark et al., 2009; Schön et al., 2004).

Crew et al. (2015) introduced the Sung Speech Corpus as a tool with which to measure MCI and sentence recognition while manipulating the pitch and timbre cues within either task. For example, MCI can be measured using a single word (fixed timbre) or different words (mixed timbre). Similarly, sentence recognition can be measured using a single F0 across words (fixed pitch) or different F0s across words (mixed pitch). With adult normal-hearing (NH) listeners, Crew et al. (2015) found a musician advantage for MCI performance but not for sentence recognition. Different from musicians, MCI performance for nonmusicians was significantly poorer in the mixed timbre condition, suggesting greater difficulty in extracting pitch information from dynamically changing spectral envelopes.

All of the above-cited studies involved adult listeners. Depending on age, children may exhibit different sensitivity to pitch and timbre cues than observed with adults. Using pure-tone stimuli, developmental effects on frequency discrimination have been extensively studied (Cooper, 1994; Halliday, Taylor, Edmondson-Jones, & Moore, 2008; Jensen, 1993; Maxon & Hochberg, 1982; Thompson, Cranford, & Hoyer, 1999). Halliday et al. (2008) reported that pure-tone frequency discrimination for 11-year-olds was similar to adults, with poorer discrimination for younger children. In contrast, Stalinski, Schellenberg, and Trehub (2008) reported no significant difference between adults and children older than 8 years for F0 discrimination for three-note contours played by a piano sample, in which only the middle note was changed in frequency. Deroche, Zion, Schurman, and Chatterjee (2012) also found no signficant effect of age at testing in children between 6 and 16 years old for modulation frequency discrimination with noise or harmonic complex carriers. Different from frequency discrimination, pitch contour identification involves perception of the global pattern over the duration of the stimulus, requiring cortical network processes (Johnsrude, Penhune, & Zatorre, 2000; Lee, Janata, Frost, Hanke, & Granger, 2011; Tramo, Cariani, Koh, Makris, & Braida, 2005). Soderquist and Moore (1970) showed a training effect on frequency discrimination in 9-year-old children and suggested immature pitch contour identification at this age. Thus, while 8-year-olds may reach adult-like levels of pitch ranking or frequency discrimination (Stalinski et al., 2008), 8- to 9-year-olds may still exhibit immature pitch contour identification.

While children may lag behind adults in pitch contour identification, little is known about the effect of timbre on children's pitch perception and vice-versa. When perceiving a melodic contour, do children rely more strongly on F0 or the spectral envelope (centroid or edge)? Do changes in pitch affect timbre perception (words in sentences)? Does early musical training affect interactions between pitch and timbre cues for music and speech perception? Does the interplay between pitch and timbre cues affect children differently from adults? To answer these questions, we measured music and speech perception in NH children aged between 8 and 16 years old using stimuli and methods similar to Crew et al. (2015). Music perception was measured using an MCI task; stimuli consisted of piano samples and sung

speech with fixed timbre (same word across notes) or mixed timbre (different words across notes). Speech perception was measured using a closed-set matrix sentence recognition task; stimuli consisted of naturally spoken speech and sung speech with fixed pitch (same F0 across words) or mixed pitch (different F0 across words). Note that the speech perception task was intended to observe the effects of changes in F0 on perception of the dynamic spectral envelope cues in the five-word *sentences*. This is a simplification of the complexity associated with speech understanding in quiet and in noise, but the study was designed to measure the effects of pitch and timbre cues using the same stimuli, which could not be done by, for example, comparing familiar melody recognition to vocal emotion recognition.

Methods

Subjects

Thirty children (13 males and 17 females) aged between 8.1 and 16.9 years old (mean age at testing = 11.3 ± 2.6 years old) participated in this study. This age range was comparable to Deroche et al. (2012; 6–16 years old), who measured sensitivity to pitch cues in school-aged children. All participants had pure-tone hearing thresholds ⩽ 15 dB HL at audiometric frequencies between 250 and 8,000 Hz and normal type A tympanometry (consistent with normal middle ear function). Subjects were paid for their participation in the study.

Table 1 shows demographic information for the subjects and responses to some items from the music questionnaire; the questionnaire is included in Appendix A. Fifteen subjects were considered musicians, and 15 were considered nonmusicians. For the musicians, the mean age at testing = 12.1 ± 2.8 years old, the mean amount of training was 5.8 ± 2.1 years, the mean age that music training begun was 6.1 ± 1.6 years, and the mean time spent training each day was 15.6 ± 8.8 min. Thirteen of the musicians had family members that were musicians. The mean self-rated musical confidence (on a scale from 1 to 10 representing the range from lowest to highest confidence) was 6.6 ± 1.8, and the mean self-rated pitch confidence (on a scale from 1 to 10) was 6.6 ± 2.7. For the nonmusicians, the mean age at testing = 10.5 ± 2.1 years old. Five of the nonmusicians had family members that were musicians. The mean self-rated musical confidence was 2.8 ± 2.4, and the mean self-rated pitch confidence was 3.1 ± 2.9.

Music Stimuli and Testing

Music perception was tested using an MCI task (Crew et al., 2015; Galvin et al., 2007, 2009). Melodic contours

consisted of five-note sequences that represented nine changes in pitch: rising, rising-flat, rising-falling, flat, flat rising, flat falling, falling, falling-flat, and falling-rising. The spacing between consecutive notes (depending on the target contour) was 1, 2, or 3 semitones. The *instrument* playing the notes was a MIDI piano sample (Galvin, Fu, & Oba, 2008) or sung speech (Crew et al., 2015; Crew, Galvin, & Fu, 2016). For the piano sample, the lowest note in any contour was A3 (220 Hz), the highest note was A4 (440 Hz), the duration of each note was 250 ms, and the time between successive notes was 50 ms. For sung speech, each of the 50 words was produced by an adult male for each semitone between A2 (110 Hz) and A3 (220 Hz). For more details regarding the sung speech stimuli, see Crew et al. (2015, 2016). The lowest note in any contour was A2 (110 Hz), the highest note was A3 (220 Hz), and the duration of each note was 500 ms. Two sung speech conditions were tested: (a) *fixed timbre*, in which the same word *Bob* was used for each note ("Bob-Bob-Bob-Bob-Bob") and (b) *mixed timbre*, in which words were randomly selected (for each trial) from within each category (name, verb, number, color, and clothing) and used for each note (e.g., "Bob-sells-three-blue-ties"). For piano and sung speech, all stimuli were normalized to have the same long-term RMS power (60 dB). For the mixed timbre condition, new words were selected for each trial. For each test block, there were 27 stimuli (9 contours × 3 semitone spacings).

Custom software (Angel Sound™; http://angelsound.emilyfufoundation.org) was used for testing. During each trial of a test, a contour was randomly selected (without replacement) from among the 27 stimuli and presented only once to the subject, who responded by clicking on one of the nine response boxes shown onscreen. No trial-by-trial feedback was provided, and stimuli were presented one time (no repeating of stimuli). The order of test conditions was randomized within and across subjects. Stimuli were presented at 60 dBA to the right channel of circumaural headphones (Sennheiser HDA 200) connected to a headphone amplifier (Tucker-Davis Technologies HB6).

Speech Stimuli and Testing

Sentence recognition was measured in quiet and in steady noise using a matrix-styled procedure (Crew et al., 2015, 2016; Hagerman, 1982; Kollmeier et al., 2015). Stimuli were the same as in Crew et al. (2015, 2016) and consisted of 50 words produced by an adult male for each semitone between A2 (110 Hz) and A3 (220 Hz), as well as spoken with a natural intonation. Three stimuli conditions were tested: (a) spoken speech, (b) sung speech with *fixed pitch*, in which each word of the sentence was sung at the same pitch (D#3, or

Table 1. Demographic Information, Music Experience, and Responses to Questionnaire Items for the Pediatric Subjects.

Subject	Age at test (years)	Gender	Music experience (years)	Type of music experience	Age at train (years)	Musician in family?	Instrument	Time train (min/day)	Musical confidence (self-rated)	Pitch confidence (self-rated)
Music 1	8.9	F	5	C, P, S	4	Y	Vi, Pi	25	10	10
Music 2	9.3	F	5	P	4	Y	Vi	15	5	7
Music 3	9.5	M	5	P, S	4	Y	Pi, Ch	20	7	9
Music 4	9.6	F	4	P	6	Y	Pi	12	6	5
Music 5	9.8	M	4	P, S	6	Y	Pi, Uk	14	6	7
Music 6	10.3	F	4	P	5	Y	Vi, Pi	11	5	6
Music 7	10.7	M	4	P, S	6	Y	Tr	N/A	8	10
Music 8	11.5	F	6	P	6	Y	Pi, Ch	17	10	5
Music 9	11.8	F	4	P	8	Y	Pi	N/A	3	1
Music 10	12.8	F	5	P	7	N	Pi	8	6	2
Music 11	13.9	F	8	P	6	Y	Pi, Fh	38	6	8
Music 12	14.9	F	7	P	7	N	Pi, Sa	1	7	6
Music 13	15.8	F	6	C, P, S	10	Y	Sa	13	8	7
Music 14	16.4	F	9	P	7	Y	Pi	15	6	6
Music 15	16.9	F	11	P, S	6	Y	Pi, Fl	13	6	10
NM1	8.1	M	0	NM	NA	Y	NA	0	5	5
NM2	8.2	F	0	NM	NA	N	NA	0	2	1
NM3	8.3	M	0	NM	NA	Y	NA	0	5	5
NM4	9.0	F	0	NM	NA	N	NA	0	1	1
NM5	9.2	F	0	NM	NA	Y	NA	0	1	1
NM6	9.3	M	0	NM	NA	N	NA	0	3	3
NM7	9.7	M	0	NM	NA	Y	NA	0	3	2
NM8	10.1	M	0	NM	NA	Y	NA	0	6	7
NM9	10.3	M	0	NM	NA	Y	NA	0	1	1
NM10	10.4	M	0	NM	NA	N	NA	0	1	1
NM11	11.2	M	0	NM	NA	N	NA	0	9	9
NM12	12.3	F	0	NM	NA	N	NA	0	1	1
NM13	13.1	M	0	NM	NA	N	NA	0	1	1
NM14	14.1	F	0	NM	NA	N	NA	0	2	8
NM15	14.6	M	0	NM	NA	N	NA	0	1	1

Note. For subject, Music = musician, NM = nonmusician. F = female, M = male. For type of music experience (exp), C = compose music, P = play instrument, S = sing. For instrument, Vi = violin, Pi = piano, Ch = choir, Uk = ukelele, Tr = trombone, Fh = french horn, Sa = saxophone, Fl = flute.

155 Hz), and (c) sung speech with *mixed pitch*, in which each sentence was paired with 1 of the 27 contours used for MCI testing. For each stimulus condition, there were 27 stimuli in each test (same as for the MCI task). For testing in noise, steady noise was matched to the spectrum of all words produced by the male talker; the signal-to-noise ratio (SNR) was 0 dB, calculated according to the long-term RMS of the sentence. The onset and offset of the noise were 500 ms before and after the sentence, respectively.

During testing, a sentence was randomly generated by selecting one word from each of the five categories (name, verb, number, color, and clothing). The subject responded by clicking on the words which they identified best matched what they heard in the sentence. Subjects were instructed to make a choice for each of the words in the sentence, and to guess if they were unsure. Once all five response words were selected, the subject pressed the *Next* button, and a new sentence was randomly generated. For more details about the speech testing, see Crew et al. (2015, 2016). The stimulus and noise conditions were randomized within and across subjects. Stimuli were presented at 60 dBA via the right channel of circumaural headphones (Sennheiser HDA 200) connected to a headphone amplifier (Tucker-Davis Technologies HB6). As with the MCI testing, no feedback was provided and stimuli were presented one time (no repeats).

Results

Music Perception

Figure 1 shows boxplots of overall MCI performance, as well as performance with 1, 2, and 3 semitone spacings for the piano, fixed timbre, and mixed timbre stimulus conditions, for the musician and nonmusician groups. The mean performance difference (across semitone spacings) between musicians and nonmusicians was 36.0, 24.7, and 30.9 percentage points for the piano, fixed timbre, and mixed timbre conditions, respectively. A split-plot repeated measures analysis of variance (RM ANOVA) with timbre (piano, fixed, and mixed) and semitone spacing (1, 2, and 3) as within-subject factors and group (musicians and nonmusicians) as the between-subject factor was performed on the MCI data; an arcsine transform (Studebaker, 1985) was performed on the data before analysis to reduce floor and ceiling effects. Results are shown in Table 2. There were significant effects for subject group ($p < .001$), semitone spacing ($p = .010$), and timbre ($p < .001$). Post hoc Bonferroni pairwise comparisons showed that performance was significantly better with three semitones than with one semitone ($p = .010$); there were no significant differences between the remaining spacing conditions ($p > .05$ in both cases). Post hoc Bonferroni pairwise comparisons showed that performance was significantly poorer with the mixed timbre than with the fixed timbre ($p < .001$) or piano conditions ($p < .001$). Although there was no significant interaction between timbre and group
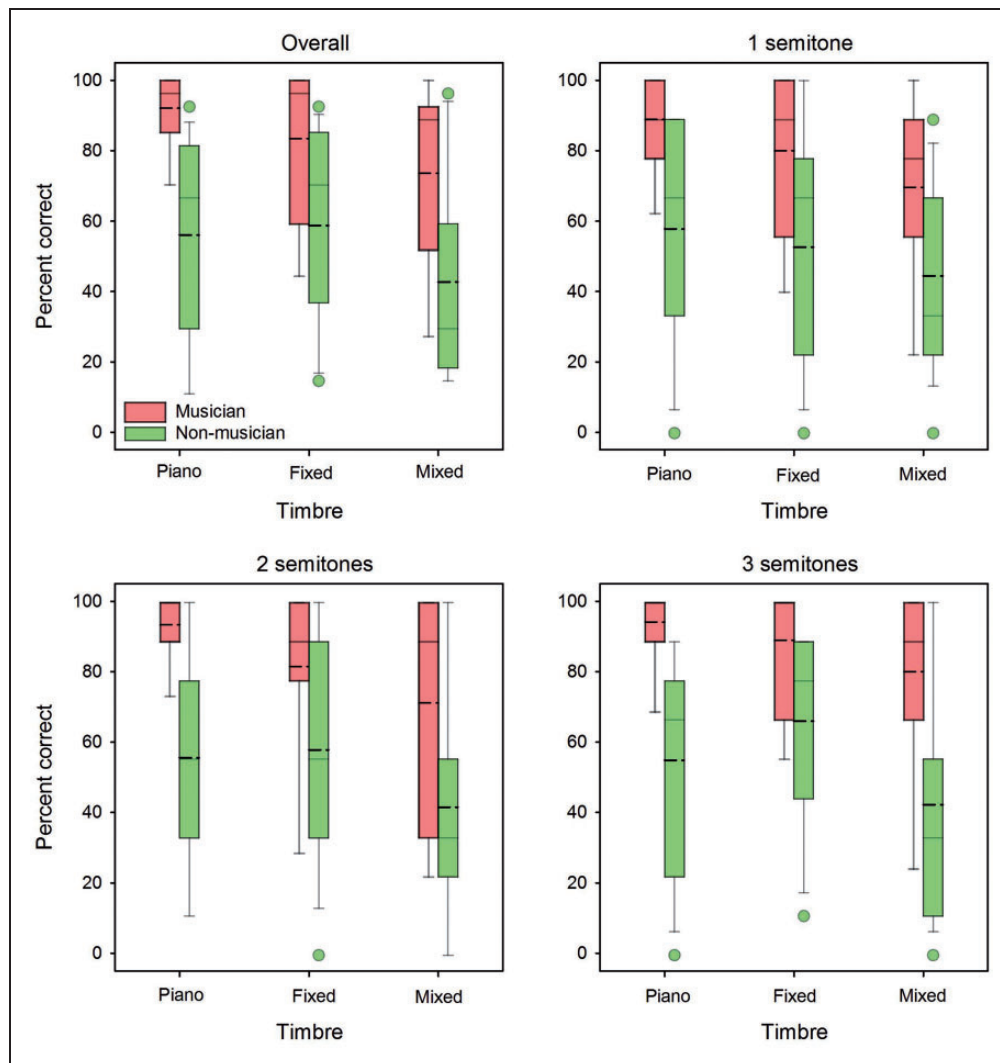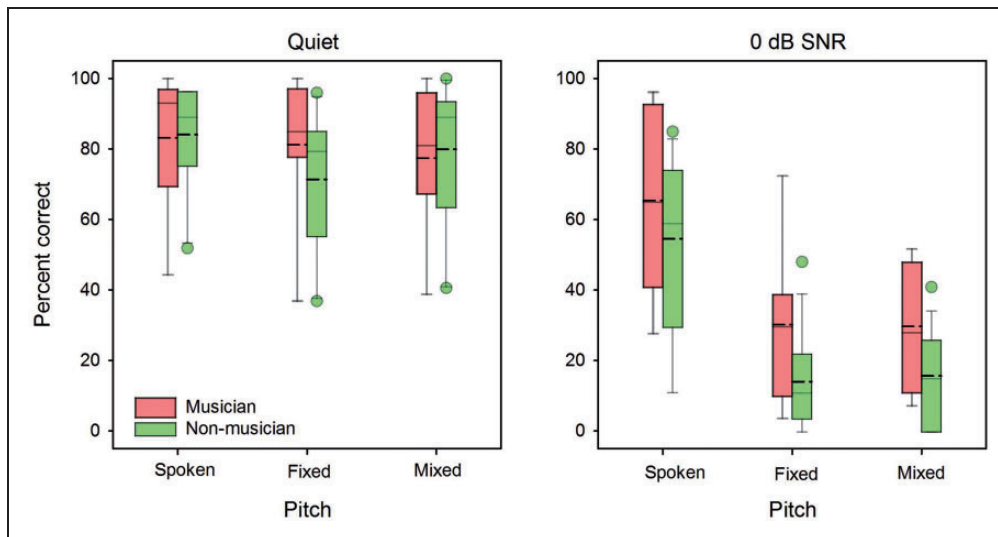

Figure 1. Boxplots for pediatric musician and nonmusician MCI scores with different stimulus types. The boxes show the 25th to 75th percentiles, the error bars show the 5th and 95th percentiles, the circles show outliers, the solid horizontal lines show median performance, and the dashed horizontal lines show mean performance. Clockwise from the top left, data are shown for overall MCI performance and MCI performance with 1-, 2-, or 3-semitone spacing.

Table 2. Results of Split-Plot RM ANOVAs on Data Shown in Figures 1 and 2. The Asterisks and Italics Indicate Significant Effects. For Post Hoc Pairwise Comparisons, Bonferroni Correction Was Applied.

	df, res	*F*	*p*	η^2	Post hoc ($p < .05$)
MCI					
Timbre	2, 56	17.3	*<.001**	0.38	Piano, Fixed > Mixed
Spacing	2, 56	5.0	*.010**	0.15	3 > 2, 1
Group	1, 28	15.5	*<.001**	0.36	Musician > nonmusician
Timbre × Spacing	4, 112	1.0	.438	0.03	
Timbre × Group	2, 56	2.8	.071	0.09	Musician: Piano, fixed > mixed
Spacing × Group	2, 56	2.1	.128	0.07	
Timbre × Spacing × Group	4, 112	0.5	.726	0.02	
Speech in quiet					
Pitch	2, 44	5.4	*.008**	0.20	Spoken > fixed
Group	1, 22	0.2	.634	0.01	Musician > nonmusician
Pitch × Group	2, 44	4.6	*.015**	0.17	Nonmusician: Spoken > fixed
Speech in noise					
Pitch	2, 54	86.9	*<.001**	0.76	Spoken > fixed, mixed
Group	1, 27	5.5	*.026**	0.17	Musician > nonmusician
Pitch × Group	2, 54	0.6	.538	0.02	

Note. MCI = melodic contour identification.

Figure 2. Boxplots for pediatric musician and nonmusician sentence recognition scores in quiet (left panel) and in noise (right panel) with different stimulus types. The boxes show the 25th to 75th percentiles, the error bars show the 5th and 95th percentiles, the circles show outliers, the solid horizontal lines show median performance, and the dashed horizontal lines shows mean performance.

($p = .071$), post hoc Bonferroni pairwise comparisons (two tailed) showed that musician performance was significantly poorer with the mixed timbre than with the fixed timbre ($p < .001$) or piano conditions ($p = .012$); there was no significant timbre effect for nonmusicians ($p > .05$ in all cases). Large effect sizes were observed for timbre, spacing, and group ($\eta^2 > 0.14$ in all cases). Mann–Whitney rank sum tests also showed significantly

higher musician self-ratings for musical confidence ($p < .001$) and pitch confidence ($p = .004$).

Speech Perception

Figure 2 shows boxplots of sentence recognition performance in quiet and noise with spoken speech and sung speech with fixed or mixed pitch. Note that because

of time constraints, musician data were unavailable for one subject for speech in quiet and in noise, and non-musician data were unavailable for five subjects for speech in quiet. In quiet, the mean performance difference between musicians and nonmusicians was −1.0, 9.8, and −2.5 percentage points for the spoken, fixed pitch, and mixed pitch conditions, respectively. In noise, the mean performance difference between musicians and nonmusicians was 10.9, 16.2, and 14.0 percentage points for the spoken, fixed pitch, and mixed pitch conditions, respectively.

Split-plot RM ANOVAs with pitch (spoken, fixed, and mixed) and group (musicians and nonmusicians) as the between-subject factor was performed on the speech data in quiet and in noise; an arcsine transform (Studebaker, 1985) was performed on the data before analysis to reduce floor and ceiling effects. Results are shown in Table 2. There were significant pitch effects for speech in quiet ($p = .008$) and speech in noise ($p < .001$). There was a significant musician advantage for speech in noise ($p = .026$) but not for speech in quiet ($p = .634$). There was a significant interaction between pitch and group for speech in quiet ($p = .015$). For speech in quiet, post hoc Bonferroni pairwise comparisons (two tailed) showed that nonmusician performance was significantly poorer with the fixed pitch than with spoken speech ($p = .012$), with no significant differences among the remaining conditions ($p > .05$ in both cases); there was no significant pitch effect for musicians ($p > .05$ in all cases). For speech in noise, post hoc Bonferroni pairwise comparisons (two tailed) showed that performance was significantly better with spoken speech than with fixed ($p < .001$) or mixed pitch ($p < .001$). For speech in quiet and in noise, there was a large effect size for pitch ($\eta^2 > 0.14$ in both cases). There was a small effect size of group for speech in quiet ($\eta^2 = 0.01$), but a large effect size for speech in noise ($\eta^2 = 0.17$).

Correlation Analyses

Various demographic factors (age at testing, the number of years of music training, the time spent training, the age when music training began, self-reported music confidence, and self-reported pitch discrimination confidence) were compared with subjects' music and speech performance; an arcsine transform (Studebaker, 1985) was performed on the music and speech data before analysis. Speech and music data were first analyzed for covariance. MCI data were first averaged across the three semitone spacing conditions; performance among the three timbre conditions was highly correlated ($p < .001$ in all cases), so scores were collapsed across timbre conditions to give a single MCI score for analysis. Speech data in quiet and at 0 dB SNR were both highly correlated across the three pitch conditions ($p < .001$ in

all cases), so scores were collapsed across pitch conditions within each SNR condition. Because different numbers of subjects completed the speech in quiet ($n = 24$) and speech in noise ($n = 29$) conditions, correlations were performed separately for the speech data. Self-reported music and pitch confidence scores were highly correlated ($p < .001$), so data were averaged to give a single confidence score. For musicians, age at testing and music experience were highly correlated, so data were collapsed into a single age or experience variable. The results of Pearson correlations are shown in Table 3.

After Bonferroni correction for multiple comparisons, across all subjects, significant correlations were observed between speech in noise and age at testing, between MCI and confidence, between MCI and speech in noise, and between speech in quiet and speech in noise. Among musicians, significant correlations were observed between speech in noise and age or experience, between speech in noise and age at training, between MCI and confidence, between MCI and speech in noise, and between speech in quiet and speech in noise. Among nonmusicians, significant correlations were observed only between speech in noise and age at testing.

While the Bonferroni-adjusted p values failed to meet significance level, medium and large effect sizes ($r > 0.30$) were observed between age and MCI and speech in quiet, between confidence and speech in noise, and between MCI versus speech in quiet, across all subjects. Among musicians, medium and large effect sizes were observed between MCI and age or experience, age at training, time spent training, and speech in quiet and between speech in quiet and age or experience, age at training, and time spent training. Among nonmusicians, medium and large effect sizes were observed between confidence and MCI, between MCI and speech in quiet, and between speech in noise and speech in quiet.

Age Effects

In this study, there was a broad range for age at testing which appeared to significantly contribute to speech performance. It is possible that speech and music performance may differ greatly between relatively young and old children. To further examine age effects, musicians and nonmusicians were grouped according to age range: 8 to 9 years (young) versus 10 to 16 years (older). Adult musician and nonmusician data from Crew et al. (2015) were added to the comparison (age range: 24–47 years old). The 10-year age cutoff was selected in consideration of data from Halliday et al. (2008) who reported that frequency discrimination was poorer for young children, but was comparable to adults by age 11 years. The 10-year age cutoff was also selected to establish comparable number of subjects for the nonmusicians where age effects might be better observed, as age at testing was

Table 3. Results of Pearson Correlations.

		MCI			Speech in quiet			Speech in noise	
	df	r	p	df	r	p	df	r	p
All									
Age at test	28	0.37	.042	22	0.45	.029	27	0.73	<.001*
Confidence	28	0.55	.002*	22	−0.01	.993	27	0.34	.068
MCI				22	0.48	.018	27	0.55	.002*
Speech in quiet							27	0.65	<.001*
Musician									
Age/experience	13	0.41	.132	12	0.52	.055	12	0.69	.008*
Age at train	13	0.46	.088	12	0.50	.070	12	0.68	.007*
Time train	13	0.32	.292	12	0.31	.305	12	0.23	.458
Confidence	13	0.17	.536	12	−0.06	.819	12	0.02	.945
MCI				12	0.50	.067	12	0.68	.007*
Speech quiet							12	0.79	<.001*
Nonmusician									
Age at test	13	−0.04	.894	8	0.23	.530	13	0.66	.008*
Confidence	13	0.31	.263	8	−0.14	.669	13	0.17	.546
MCI				8	0.53	.115	13	0.18	.514
Speech in quiet							8	0.41	.241

Note. MCI = melodic contour identification.
MCI data were collapsed across the three timbre and three semitone spacing conditions. Speech data were collapsed across the three pitch conditions. Self-reported music and pitch discrimination data were collapsed into a single *confidence* variable due to covariance. For musicians, age at testing and musical experience were collapsed into the age or experience variable due to covariance. Unadjusted *p* values are shown for all correlations. The asterisks and italics represent significant correlations after Bonferroni adjustment for multiple pairwise comparisons.

significantly correlated with music training experience, creating a possible confound for observing age effects in musicians. Figure 3 shows overall MCI performance (averaged across semitone spacings; top row), speech in quiet (middle row), and speech in noise (bottom row) for young, older, and adult musicians (left column), nonmusicians (middle column), and all subjects (combined musician and nonmusician; right column). Note that there were no adult data for speech in noise.
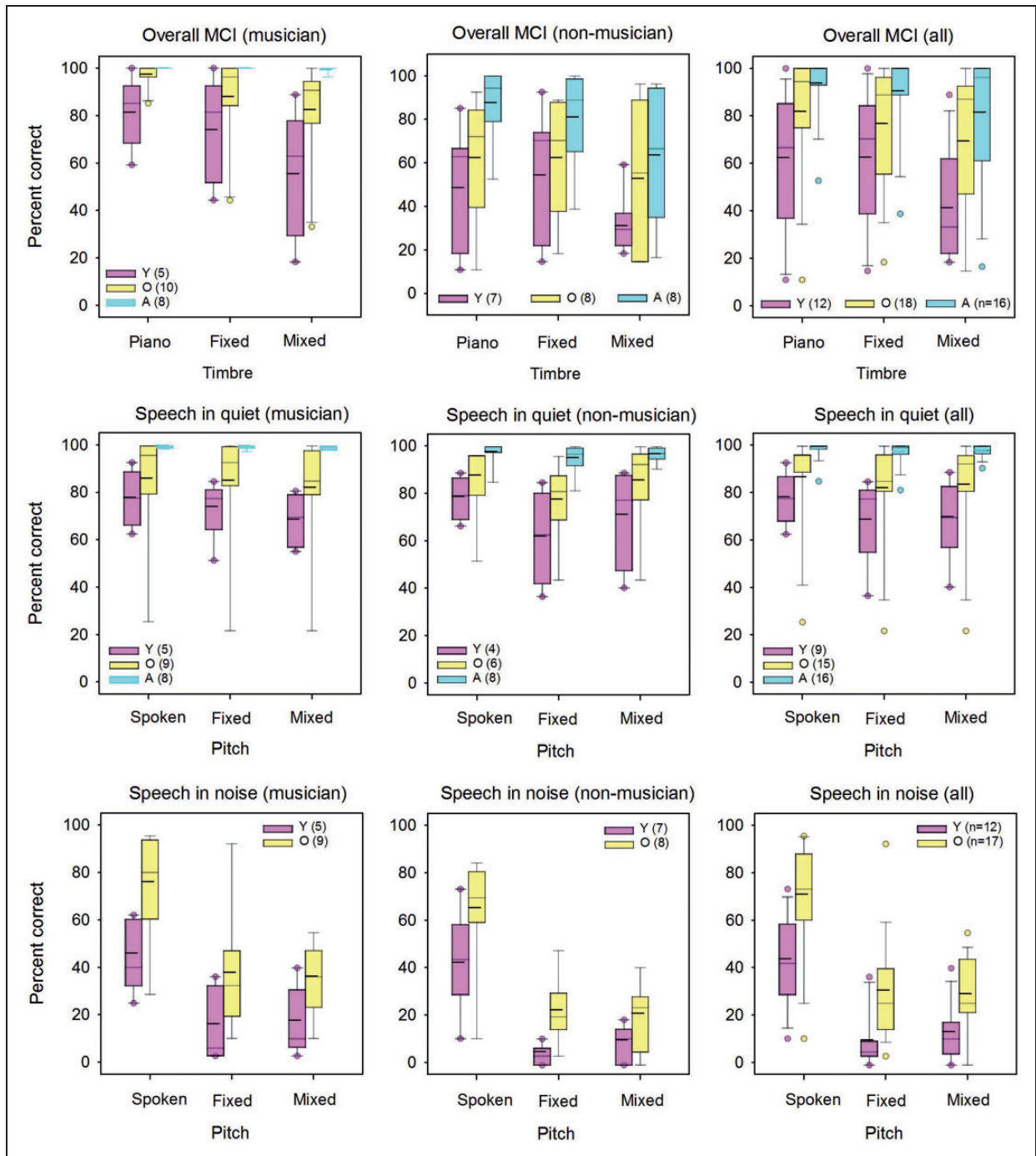
For musicians (left column in Figure 3), MCI performance generally worsened from piano to fixed timbre to mixed timbre for the young and older groups; adult performance was near-perfect across timbre conditions. MCI performance was generally better for adults than for young and older children and better for older than for young children. For speech in quiet, there were no strong differences among the three pitch conditions. Again, performance was generally better for adults than for young and older children and better for older than for younger children. For speech in noise, performance was generally better with spoken speech than with fixed or mixed pitch. Performance was generally better for older than for younger children.

For nonmusicians (middle column of Figure 3), MCI performance was markedly poorer with mixed timbre compared with the piano for young children and adults; for older children, performance was generally similar across timbre conditions. Performance was generally best for adults and better for older than for young children; in the mixed timbre condition, performance (although highly variable) was markedly better for adults and older children than for young children. Similar to musicians, nonmusician performance for speech in quiet was generally similar across pitch conditions. Performance was generally better for adults than for young and older children and better for older than for young children. For speech in noise, the fixed and mixed pitch conditions seemed to more strongly affect nonmusician than musician performance. Similar to musicians, nonmusician performance was generally better for older than for younger children.

Because of the limited number of musician and nonmusician subjects in each age-group and the large variability in performance, statistical analyses within and among these groups are not likely to be meaningful. Instead, musicians and nonmusicians were combined within each age-group (right column of Figure 3). Split-plot RM ANOVAs were performed on the MCI, speech in quiet, and speech in noise using the combined musician and nonmusician data, with timbre (or pitch) as

Figure 3. Boxplots of overall MCI (top row), sentence recognition in quiet (middle row), and sentence recognition in noise scores (bottom row) for musicians (left column), nonmusicians (middle column), and all subjects (musicians and nonmusicians together; right column); the adult data are from Crew et al., 2015). Y (n) = young child subjects (8–9 years old); O (n) = older child subjects (10–16 years old); A (n) = adult subjects (24–47 years old). The boxes show the 25th to 75th percentiles, the error bars show the 5th and 95th percentiles, the circles show outliers, the solid horizontal lines show median performance, and the dashed horizontal lines show mean performance.

the within-subject factor and age-group (young, older, and adult) as the between-subject factor; data were subjected to an arcsine transform (Studebaker, 1985) before analysis. Results are shown in Table 4. Significant age

effects were observed for MCI ($p = .001$), speech in quiet ($p < .001$), and speech in noise ($p = .001$). For MCI, post hoc Bonferroni pairwise comparisons (two tailed) showed that performance was significantly better for

Table 4. Results of Split-Plot RM ANOVAs on Data Shown in Figure 3; Musician and Nonmusician Data Were Combined Within Each Age-Group. The Asterisks and Italics Indicate Significant Effects. For Post Hoc Pairwise Comparisons, Bonferroni Correction Was Applied.

	df, res	*F*	*p*	η^2	Post hoc (*p* < .05)
MCI					
Timbre	2, 86	28.2	*<.001**	0.40	Piano, fixed > mixed
Group	2, 86	8.5	*.001**	0.28	Adult > young
Timbre × Group	4, 86	1.2	.328	0.05	
Speech in quiet					
Pitch	2, 74	6.2	*.003**	0.14	Spoken > fixed, mixed
Group	2, 74	17.2	*<.001**	0.48	Adult > young, older; Older > younger
Pitch × Group	4, 74	0.3	.866	0.02	
Speech in noise					
Pitch	2, 54	84.8	*<.001**	0.76	Spoken > fixed, mixed
Group	1, 54	13.7	*.001**	0.34	Older > young
Pitch × Group	2, 54	1.1	.336	0.04	

Note. MCI = melodic contour identification.

adults than for younger children ($p < .001$) but not significantly different between adults and older children ($p = .142$) or between younger and older children ($p = .071$). For speech in quiet, post hoc Bonferroni pairwise comparisons (two tailed) showed that performance was significantly better for adults than for young ($p < .001$) or older children ($p = .003$) and significantly better for older than young children ($p = .046$). For speech in noise, post hoc Bonferroni pairwise comparisons (two tailed) showed that performance was significantly better for older than young children ($p = .001$). Large effect sizes were observed for timbre for MCI ($\eta^2 = 0.40$) and pitch for speech in quiet ($\eta^2 = 0.14$) and in noise ($\eta^2 = 0.76$), and for group across all music and speech measures ($\eta^2 > 0.25$ in all cases). Note that there were no significant interactions between timbre and age-group for MCI ($p = .328$) and between pitch and age-group for speech in quiet ($p = .866$) or speech in noise ($p = .366$). This suggests that the effects of interfering pitch or timbre cues did not significantly differ across age-groups.

Discussion

Significant musician effects were observed for MCI and sentence recognition in noise but not for sentence recognition in quiet. MCI performance was significantly affected by timbre only in musicians. For musicians and nonmusicians, sentence recognition in noise was significantly poorer with sung speech (with fixed or mixed pitch) than with naturally produced spoken speech. Speech and music perception was generally poorer in children than in adults using the same listening tasks and stimuli. Later, we discuss the results in greater detail.

Music Perception

MCI performance was significantly better in musicians than nonmusicians. Musician performance was significantly better with the piano and fixed timbre than with the mixed timbre; nonmusician performance was not significantly affected by instrument timbre. Note that piano stimuli were shorter (250 ms/note) than the fixed or mixed timbre stimuli (500 ms/note). However, performance was similar between the piano and fixed timbre stimuli and better with the piano than with the mixed timbre stimuli, suggesting that 250 ms/note may have been sufficiently long to extract pitch for this listening task.

In general, MCI performance was poorer when the timbre cues varied across notes in the contour. Semantic complexity, due to the words changing across notes, may have also contributed to the poorer performance in the mixed timbre condition. For the piano or fixed timbre conditions, each note varied only in the dimension of pitch; the timbre (instrument or word) was unchanged within the contour. For the mixed timbre condition, both the timbre and the word changed. While sentence recognition in quiet was generally good, it was not near-perfect, as for the adults in Crew et al. (2015). This suggests that the present children may have had some difficulty with the speech materials or sentence constructs from the Sung Speech Corpus. Note that in Crew et al. (2015), MCI performance was poorer with the mixed timbre for nonmusician adults, who would presumably be less susceptible than children to changes in semantic information across notes. It would be interesting to observe whether MCI performance would be similarly affected by nonlinguistic stimuli with dynamic timbre (e.g., spectral bands, vowels, etc.). The mixed timbre sung speech used in this study and in Crew

et al. (2015) is roughly analogous to sung lyrics, as are often encountered in everyday music listening. Sequential notes in musical melodies are not typically played by different instruments (e.g., except in gamelan music).

In this study, timbre cues significantly affected only musician MCI performance. Allen and Oxenham (2014) found no significant difference between musicians and nonmusicians in terms of the capacity of timbre cues to interfere with pitch perception. This discrepancy may be due to differences in stimuli and experimental design. In Allen and Oxenham (2014), F0 discrimination was measured using single 500 ms stimuli. In the present MCI task with the fixed or mixed timbre conditions, listeners were required to identify five-note melodic contours over a longer duration (500 ms × 5 note = 2500 ms total stimulus duration). Also, the MCI task required listeners to track changes in pitch direction, rather than the relatively simple discrimination in Allen and Oxenham (2014). It is unclear how such *focused* attention on differences between stimuli may relate to listeners' ability to perceive changes in pitch with dynamic timbre cues over a longer duration, as occurs with music (especially vocal music). It is possible that when short-term memory demands are increased (as in the MCI task, relative to F0 discrimination), long-term music training may provide an advantage.

Mean MCI performance with the piano for the present children was approximately 20 percentage points poorer than observed for adults (Crew et al., 2015; Galvin et al., 2008). With the mixed timbre stimuli, MCI performance was significantly poorer than with the piano or fixed timbre stimuli, even with the relatively wide three-semitone spacing; similar patterns of results were observed in adults in Crew et al. (2015). When musician and nonmusician data were combined, MCI performance was significantly better with the piano and fixed timbre than with the mixed timbre, with no interaction between age-group (young, older, and adult) and timbre. Thus, while MCI performance was poorer in children than in adults, age did not seem to be a factor in terms of timbre effects when musicians and nonmusicians were combined in each age-group. Both adults and children seemed better able to extract melodic pitch information when the spectral envelope was relatively stable across notes. When only musicians were considered (top left panel of Figure 3), adults were not sensitive to timbre, but MCI performance progressively worsened from piano to fixed timbre to mixed timbre for the young and older children. The young musicians may have attended more strongly to spectral envelope cues than did older musicians, who may have been better able to extract pitch information according to F0. Previous studies have shown that cortical networks are required for processing of complex stimuli

(Heffner & Whitfield, 1976; Tramo et al., 2005) and for pattern recognition (Johnsrude et al., 2000; Lee et al., 2011; Tramo et al., 2005). Moore and Linthicum (2007) reported that the auditory cortex does not mature until adolescence. When nonmusicians were considered, performance with the mixed timbre condition was markedly poorer than with the piano or fixed timbre for young and adult listeners (with no clear timbre effect for older children). Taken together, there is somewhat of a mixed result for age effects on MCI performance, with a clear advantage for adults, and a less-clear interaction with the present timbre conditions.

Across all subjects, MCI performance was not significantly correlated with age at testing. Note that before Bonferroni adjustment for multiple comparisons, there was a significant correlation between age at testing and MCI performance ($r = 0.37$, $p = .042$). The lack of correlation may be due to insufficient power (0.31 when $\alpha = 0.025$) or to the fairly broad age range of children (8–16 years). If subjects 12 years or younger were considered ($n = 20$), there was a significant correlation between age at testing and MCI performance ($r = 0.66$, $p = .002$). There was no significant correlation for subjects older than 12 years ($n = 10$, $r = 0.42$, $p = .225$), suggesting that age effects that might have contributed to MCI performance may have been resolved by age 12 years. The data in Figure 3 show that performance was generally poorer for young (8–9 years old) than older (10–16 years old) children (especially for musicians), with adult performance much better than that of young or older children. While previous studies have shown that simple pitch discrimination performance in children becomes adultlike by age 10 years (Cooper, 1994; Maxon & Hochberg, 1982; Stalinski et al., 2008), the present data suggest that age may be a factor when listeners are required to use pitch information in a contour identification task, especially for children younger than 12 years old. Interestingly, while Deroche et al. (2012) found no significant age effects for F0 discrimination among children age 6 to 16 years old, Deroche, Kulkarni, Christensen, Limb, and Chatterjee (2016) found that perception of F0 sweeps was significantly poorer for children than for adults. The present correlation data are somewhat in agreement with Deroche et al. (2012), but a more detailed observation of the present age effects is in agreement with Deroche et al. (2016), where perception of changes in F0 sweeps may be more analogous to perception of melodic contours.

While long-term music training appeared to enhance MCI performance, there were no significant correlations between MCI performance and age or music experience, age that training was begun, or the amount of daily practice. It could be that the musicians had sufficient experience (all had more than 4 years of training) to perform well in the MCI task. Nonetheless, musicians were

affected by stimulus type, with significantly poorer performance with the mixed timbre stimuli. It is possible that music training with a single instrument may have provided some advantage for the piano and fixed timbre stimuli; note that 11 of the 15 musicians took piano lessons. According to the questionnaire data, musicians reported significantly better confidence in pitch perception than did nonmusicians. While self-reported music or pitch discrimination confidence was not significantly correlated with MCI performance within musicians and nonmusicians, confidence was significantly correlated with MCI performance when musicians and nonmusicians were pooled together. This suggests some relationship between behavioral and subjective reports of musical pitch perception when musician status was not considered.

Interestingly, when data were collapsed across the timbre and pitch conditions, MCI performance was significantly correlated with speech perception in noise across all subjects (Table 3). This correlation appears to be highly driven by the musician data, as there was no significant correlation between MCI and speech in noise for nonmusicians. This correlation appeared to be driven by the fixed timbre or pitch ($r = 0.50$, $p = .005$) and mixed timbre or pitch conditions ($r = 0.56$, $p = .002$). Thus, relationships between speech and music perception may more strongly emerge with challenging stimuli and listening tasks.

Speech Perception

Across all subjects, mean performance in noise with spoken speech was 38 and 37 percentage points better than with sung speech with a fixed or mixed timbre, respectively. In quiet, mean performance with spoken speech was six points better than with fixed pitch and five points better than with mixed pitch. In quiet, there was a significant difference between the spoken and fixed speech conditions only for nonmusicians ($p = .012$). In noise, performance was significantly better with spoken speech than with the fixed or mixed pitch for musicians and nonmusicians ($p < .05$ in all cases). The 38-point deficit associated with the fixed pitch stimuli (relative to spoken speech) was much larger than reported in previous studies with adult listeners in which the natural pitch variation in running speech was flattened (Miller, Schlauch, & Watson, 2010) or distorted (Binns & Culling, 2007). Taken together, the production of sung speech in this study (and in Crew et al., 2015) appears to have had a greater effect on speech perception than F0 variations.

Speech performance in quiet was significantly poorer for young and older children than for the adults in Crew et al. (2015). Speech performance in quiet and in noise was significantly poorer for young than for older children. Thus, there appear to be general age effects for the present matrix-styled sentence test, similar to that observed with other sentence test materials (Myhrum, Tvete, Heldahl, Moen, & Soli, 2016; Ng, Meston, Scollie, & Seewald, 2011). Across all children, age at testing was significantly correlated with speech performance in quiet and in noise. This is consistent with findings that speech understanding in noise continues to improve until adolescence (Jamieson, Kranjc, Yu, & Hodgetts, 2004; Nelson & Soli, 2000; Neuman, Wroblewski, Hajicek, & Rubinstein, 2010).

There was a significant musician effect for speech performance in noise (but not in quiet). Among musicians, the amount of music training and the age that training was begun was significantly correlated with speech perception in noise. With NH adults, previous studies have reported significant musician effects in noise, but the musician advantage was often quite small (Parbery-Clark et al., 2009). Fuller, Galvin, Maat, et al. (2014) and Fuller, Galvin, Free, et al. (2014) observed musician effects only for pitch-mediated speech tasks such as vocal emotion and voice gender recognition, with no effect for sentence recognition in noise. Similarly, Ruggles et al. (2014) showed no musician effect for sentence recognition in noise when pitch cues were preserved or removed. A number of recent studies have suggested that musician advantages for speech perception may not be due to better perception of voice pitch cues (Baskent & Gaudrain, 2016; Deroche et al., 2017; Madsen et al., 2017). In this study, musicians may have been less susceptible to the atypical sung speech patterns, with or without variations in pitch.

Why Sung Speech in Children?

While many previous studies have examined various aspects of pitch and timbre perception, relatively few have explored interactions between pitch and timbre cues outside of simple discrimination or ranking tasks. The advantage of using sung speech is that listeners must extract useful pitch or timbre cues from longer stimuli, as may be encountered in everyday listening. However, the melodic contours and the matrix sentences in this study and in Crew et al. (2015, 2016) are quite different than would occur in everyday listening; duration and amplitude cues are held constant across notes or words, pitch information is quantized according to semitone steps, there are no natural transitions between subsequent notes or words, and so on. Still, the construction of the contours and sentences allows for direct manipulation of pitch and timbre information that is usually not possible with running speech or excerpts of music. This can provide important information about how pitch and timbre cues may contribute to speech and music perception, as well as how these cues may interfere with each other.

MCI and speech performance in quiet was significantly better in adults than in children, and there appeared to be broad age effects among children (see Figure 3). While both the MCI and speech tests used closed-set paradigms, there may have been a substantial cognitive load for these tasks, especially for the sentence recognition task where subjects had to select from 10 choices in each of five columns. For younger children, it may be preferable to include fewer stimuli in the MCI and sentence recognition tests. The poorer performance in children compared with adults may also have been partly associated with slight procedural differences between groups. In Crew et al. (2015), the adult subjects were allowed to repeat stimuli up to a maximum of three times; in the present study, stimuli were only presented one time, with no repeats. While allowing repetitions of stimuli may have improved performance in children, it would have also generated longer test blocks that may have adversely affected attentiveness, possibly reducing performance.

One issue that relates to musician advantages in children (as shown in this study) are covarying factors of which music training is only one of many influential variables. Parents who wish and can afford to invest in music lessons may provide a rich learning environment that can advantage their children in many ways. Corrigall, Schellenberg, and Misura (2013) showed significant correlations in children between the duration of music training and socioeconomic status (SES), duration of nonmusical extra-curricular activities, IQ, school performance, conscientiousness, and openness-to-experience. Swaminathan, Schellenberg, and Khalil (2017) also found a significant relationship between the amount of music training and SES, as well as with nonverbal intelligence, melody perception, and rhythm perception. The authors also found a significant relationship between music aptitude and intelligence after controlling for SES and music training and suggested that high-functioning children may be more likely to pursue music lessons because of fundamentally better music perception (apart from music training). In general, parents' education, musical experience, and income would seem to be limiting factors in children's participation in music training. In this study, 13 of the 15 musicians also had parents who were musicians, while only 6 of 15 nonmusicians had parents who were musicians. However, parents' musicianship need not be a limiting factor in children's musical training. Although it was less available now than in previous generations, music education was once a core activity in public and private schools, where there might be greater access to instruments and music lessons. A recent study by Ilari, Keller, Damasio, and Habibi (2016) investigated the benefits of 1 year of music training in school in underprivileged Latino children. Results showed better pitch perception and production in students who participated in the program

and declining performance in those who did not. This suggests that music lessons in school (or otherwise funded) may benefit low SES children who may not otherwise have access to music training. The source of benefit of music training for speech perception in children remains unclear. Unfortunately, data related to families' SES or to children's cognitive function were not collected in this study but would be valuable for future music studies in children.

Conclusions

In this study, music and speech perception were measured in children; music perception was measured using an MCI task, and speech perception in quiet and in noise was measured using a matrix-styled sentence recognition task. Stimuli for MCI testing included a piano sample and sung speech with either fixed timbre or mixed timbre. Stimuli for speech testing included naturally spoken speech and sung speech with either fixed pitch or mixed pitch. Major findings included the following:

1. Significant musician advantages were observed for MCI and speech in noise but not for speech in quiet. The results with children were consistent with those from the related study with adults from Crew et al. (2015).
2. MCI performance was significantly poorer with mixed timbre stimuli, suggesting possible interference between timbre and pitch cues for melodic pitch perception or possible influence of the semantic information contained within the mixed timbre stimuli.
3. Sentence recognition in noise was significantly poorer with the fixed or mixed pitch stimuli than with spoken speech, suggesting susceptibility to the atypical speech patterns associated with sung speech.
4. MCI and speech in quiet were significantly better for adults than for children. Among musicians, music and speech performance was generally better for older (10–16 years old) than for young musicians (8–9 years old); age effects were less clear among child nonmusicians. Age at testing was significantly correlated with speech performance in children, suggesting that development may have contributed to the pattern of speech results.

Appendix A

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 classically trained as a musician or self-taught?

How often did or 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 to 10 (1 *being not confident* and 10 *being very confident*) rank your ability on discriminating pitches of tones in music.

Acknowledgments

Declaration of Conflicting Interests

Funding

Supplemental Material

Supplementary material for this article is available online.

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.

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: 10.1121/1.399772.

Baskent, D., & Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *The Journal of the Acoustical Society of America*, *139*, EL51–EL56. doi: 10.1121/1.4942628.

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

Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *Journal of the Acoustical Society of America*, *137*(1), 378–387. doi: 10.1121/1.4904537.

Brokx, J. P. L., & Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics*, *10*(1), 23–36.

Carey, M. J., Parris, E. S., Lloyd-Thomas, H., & Bennett, S. (1996). Robust prosodic features for speaker identification. In *Proceeding of Fourth International Conference on Spoken Language Processing* (pp. 1800–1803). Washington, DC: IEEE.

Carlson, R., Granstrom, B., & Klatt, D. H. (1979). Vowel perception: The relative perceptual salience of selected acoustic manipulations. *Speech transmission laboratory quarterly progress and status* (Report no. STL-QPSR 3-4; pp. 73–83). Stockholm, Sweden: Royal Institute of Technology.

Cooper, N. (1994). An exploratory study in the measurement of children's pitch discrimination ability. *Psychology of Music*, *22*(1), 56–62. doi: 10.1177/0305735694221005.

Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music training, cognition, and personality. *Frontiers in Psychology*, *4*(222). doi:10.3389/fpsyg.2013.00222.

Crew, J. D., Galvin, J. J. III, & 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.

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*, 1–15. doi:10.1177/2331216516669329.

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

Deroche, M. L., Zion, D. J., Schurman, J. R., & Chatterjee, M. (2012). Sensitivity of school-aged children to pitch-related cues. *The Journal of the Acoustical Society of America*, *131*, 2938–2947. doi: 10.1121/1.3692230.

Deroche, M. L. D., Kulkarni, A. M., Christensen, J. A., Limb, C. J., & Chatterjee, M. (2016). Deficits in the sensitivity to pitch sweeps by school-aged children wearing cochlear implants. *Frontiers in Neuroscience*, *10*. doi:10.3389/fnins.2016.00073.

Deroche, M. L. D., Limb, C. J., Chatterjee, M., & Gracco, V. L. (2017). Similar abilities of musicians and non-musicians to segregate voices by fundamental frequency. *The Journal of the Acoustical Society of America*, *142*(4), 1739–1755. doi: 10.1121/1.5005496.

Deutsch, D., Henthorn, T., & Dolson, M. (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception: An Interdisciplinary Journal*, *21*(3), 339–356. doi: 10.1525/mp.2004.21.3.339.

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: 10.1121/1.1802535.

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. doi:10.3389/fnins.2014.00179.

Fuller, C. D., Galvin, J. J. III, Free, R. H., & Başkent, D. (2014). Musician effect in cochlear implant simulated gender categorization. *The Journal of the Acoustical Society of America*, *135*(3), 1664–1664. doi: 10.1121/1.4865263.

Galvin, J. J., Fu, Q.-J., & Nogaki, G. (2007). Melodic contour identification by cochlear implant listeners. *Ear and Hearing*, *28*(3), 302–319. doi: 10.1097/01.aud.0000261689.35445.20.

Galvin, J. J., Fu, Q.-J., & Oba, S. (2008). Effect of instrument timbre on melodic contour identification by cochlear implant users. *The Journal of the Acoustical Society of America*, *124*(4), EL189–EL195. doi: 10.1121/1.2961171.

Galvin, J. J., Qian-Jie, F., & 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. doi: 10.1111/j.1749-6632.2009.04551.x.

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. doi: 10.1111/j.1467-7687.2010.00955.x.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*, *61*(5), 1270–1277. doi: 10.1121/1.381428.

Grey, J. M. (1978). Timbre discrimination in musical patterns. *The Journal of the Acoustical Society of America*, *64*(2), 467–472.

Hagerman, B. (1982). Sentences for testing speech intelligibility in noise. *Scandanavian Audiology*, *11*(2), 79–87. doi: 10.3109/01050398209076203.

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*(6), 4393–4402. doi: 10.1121/1.2890749.

Heffner, H., & Whitfield, I. C. (1976). Perception of the missing fundamental by cats. *The Journal of the Acoustical Society of America*, *59*(4), 915–919. doi: 10.1121/1.380951.

Ilari, B. S., Keller, P., Damasio, H., & Habibi, A. (2016). The development of musical skills of underprivileged children over the course of 1 year: A study in the context of an el sistema-inspired program. *Frontiers in Psychology*, *2*(7), 62. doi:10.3389/fpsyg.2016.00062.

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. doi: 10.3766/jaaa.15.7.5.

Jensen, J. K. (1993). Development of basic auditory discrimination in preschool children. *Psychological Science*, *4*(2), 104–107. doi: 10.1111/j.1467-9280.1993.tb00469.x.

Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, *123*(1), 155–163. doi: 10.1093/brain/123.1.155.

Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M. A., Uslar, V., Brand, T., & Wagener, K. C. (2015). The multilingual matrix test: Principles, applications, and comparison across languages: A review. *International Journal of Audiology*, *54*, 3–16. doi: 10.3109/14992027.2015.1020971.

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).

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605. doi:10.1038/nrn2882.

Lee, Y. S., Janata, P., Frost, C., Hanke, M., & Granger, R. (2011). Investigation of melodic contour processing in the brain using multivariate pattern-based fMRI. *Neuroimage*, *57*(1), 293–300. doi: 10.1016/j.neuroimage.2011.02.006.

Lin, M. C. (1988). The acoustic characteristics and perceptual cues of tones in standard Chinese. *Chinese Yuwen*, *205*, 182–193.

Madsen, S. M. K., Whiteford, K. L., & Oxenham, A. J. (2017). Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds. *Scientific Reports*, *7*(1), 12624. doi:10.1038/s41598-017-12937-9.

Marozeau, J., & de Cheveigné, A. (2007). The effect of fundamental frequency on the brightness dimension of timbre. *The Journal of the Acoustical Society of America*, *121*(1), 383–387. doi: 10.1121/1.2384910.

Marozeau, J., de Cheveigné, A., McAdams, S., & Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *The Journal of the Acoustical Society of America*, *114*(5), 2946–2957. doi: 10.1121/1.1618239.

Maxon, A. B., & Hochberg, I. (1982). Development of psychoacoustic behavior: Sensitivity and discrimination. *Ear and Hearing*, *3*(6), 301–308.

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. doi: 10.1007/BF00419633.

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

Molis, M. R. (2005). Evaluating models of vowel perception. *The Journal of the Acoustical Society of America*, *118*(2), 1062–1071. doi: 10.1121/1.1943907.

Moore, J. K., & Linthicum, F. H. Jr. (2007). The human auditory system: A timeline of development. *International Journal of Audiology*, *46*(9), 460–478. doi: 10.1080/14992020701383019.

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. doi: 10.1121/1.405558.

Myhrum, M., Tvete, O. E., Heldahl, M. G., Moen, I., & Soli, S. D. (2016). The Norwegian hearing in noise test for children. *Ear & Hearing*, *37*(1), 80–92. doi: 10.1097/AUD.0000000000000224.

Nelson, P. B., & Soli, S. (2000). Acoustical barriers to learning-children at risk in every classroom. *Language, Speech, and Hearing Services in Schools*, *31*(4), 356–361. doi: 10.1044/0161-1461.3104.356.

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.

Ng, S. L., Meston, C. N., Scollie, S. D., & Seewald, R. C. (2011). Adaptation of the BKB-SIN test for use as a pediatric aided outcome measure. *Journal of the American Academy of Audiology*, *22*(6), 375–386. doi: 10.3766/jaaa.22.6.6.

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing*, *30*(6), 653–661. doi: 10.1097/AUD.0b013e3181b412e9.

Ruggles, D. R., Freyman, R. L., & Oxenham, A. J. (2014). Influence of musical training on understanding voiced and whispered speech in noise. *PloS One*, *9*(1), e86980. doi:10.1371/journal.pone.0086980.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*(3), 341–349. doi: 10.1111/1469-8986.00172.x.

Soderquist, D. R., & Moore, M. J. (1970). Effect of training on frequency discrimination in primary school children. *The Journal of Auditory Research*, *10*(3), 185–192.

Stalinski, S. M., Schellenberg, E. G., & Trehub, S. E. (2008). Developmental changes in the perception of pitch contour: Distinguishing up from down. *The Journal of the Acoustical Society of America*, *124*(3), 1759–1763. doi: 10.1121/1.2956470.

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *Journal of Speech and Hearing Research*, *28*(3), 455–462. doi: 10.1044/jshr.2803.455.

Swaminathan, S., Schellenberg, E. G., & Khalil, S. (2017). Revisiting the association between music lessons and intelligence: Training effects or music aptitude? *Intelligence*, *62*, 119–124. doi: 10.1016/j.intell.2017.03.005.

Swanepoel, R., Oosthuizen, D. J. J., & 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*(4), 2652–2662.

Thompson, N. C., Cranford, J. L., & Hoyer, E. (1999). Brief-tone frequency discrimination by children. *Journal of Speech, Language & Hearing Research*, *42*(5), 1061–1068. doi: 10.1044/jslhr.4205.1061.

Tramo, M. J., Cariani, P. A., Koh, C. K., Makris, N., & Braida, L. D. (2005). Neurophysiology and neuroanatomy of pitch perception. *Annals of the New York Academy of Sciences*, *1060*, 148–174. doi: 10.1196/annals.1360.011.