University of Reading

Individuals with autism spectrum disorder are impaired in absolute but not relative pitch and duration matching in speech and song imitation

Article

Published Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1002/aur.2569

Publisher: John Wiley

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

RESEARCH ARTICLE

Individuals with autism spectrum disorder are impaired in absolute but not relative pitch and duration matching in speech and song imitation

Li Wang[1] | Peter Q. Pfordresher[2] | Cunmei Jiang[3] | Fang Liu[1]

[1]School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK

[2]Department of Psychology, University at Buffalo, State University of New York, Buffalo, New York, USA

[3]Music College, Shanghai Normal University, Shanghai, China

Correspondence
Fang Liu, School of Psychology & Clinical Language Sciences, University of Reading, Earley Gate, Reading RG6 6AL, UK.
Email: f.liu@reading.ac.uk

Abstract

Individuals with autism spectrum disorder (ASD) often exhibit atypical imitation. However, few studies have identified clear quantitative characteristics of vocal imitation in ASD. This study investigated imitation of speech and song in English-speaking individuals with and without ASD and its modulation by age. Participants consisted of 25 autistic children and 19 autistic adults, who were compared to 25 children and 19 adults with typical development matched on age, gender, musical training, and cognitive abilities. The task required participants to imitate speech and song stimuli with varying pitch and duration patterns. Acoustic analyses of the imitation performance suggested that individuals with ASD were worse than controls on absolute pitch and duration matching for both speech and song imitation, although they performed as well as controls on relative pitch and duration matching. Furthermore, the two groups produced similar numbers of pitch contour, pitch interval-, and time errors. Across both groups, sung pitch was imitated more accurately than spoken pitch, whereas spoken duration was imitated more accurately than sung duration. Children imitated spoken pitch more accurately than adults when it came to speech stimuli, whereas age showed no significant relationship to song imitation. These results reveal a vocal imitation deficit across speech and music domains in ASD that is specific to absolute pitch and duration matching. This finding provides evidence for shared mechanisms between speech and song imitation, which involves independent implementation of relative versus absolute features.

Lay Summary

Individuals with autism spectrum disorder (ASD) often exhibit atypical imitation of actions and gestures. Characteristics of vocal imitation in ASD remain unclear. By comparing speech and song imitation, this study shows that individuals with ASD have a vocal imitative deficit that is specific to absolute pitch and duration matching, while performing as well as controls on relative pitch and duration matching, across speech and music domains.

KEYWORDS
ASD, duration, pitch, song, speech, vocal imitation

INTRODUCTION

Imitation is a fundamental skill that emerges early in typical human development (Meltzoff, 2017). It is essential for learning of complex constructs, including language (McEwen et al., 2007; Rose et al., 2009) and social interaction (Kuhl, 2007; Masur, 2006; Vivanti & Hamilton, 2014). In particular, by imitating others or being imitated, individuals gradually become aware of the physical world, such as cause-effect relations

(Meltzoff & Williamson, 2013); the mental states of other people, such as their intentions and feelings (Meltzoff & Keith Moore, 1994); and the sounds around them, such as languages (Charman et al., 2000; Young et al., 2011).

Imitation in individuals with autism spectrum disorder (ASD) is often described as atypical (Williams et al., 2001). Deficits in imitation skills in ASD have been reported for a variety of tasks, including action imitation, which involves using body and hands to act (Ham et al., 2011; Young et al., 2011); object-directed action imitation where actions involve objects (Cossu et al., 2012; Vivanti et al., 2014); facial imitation (Bernier et al., 2007) and vocal imitation (McCann et al., 2007). Specifically, when individuals with ASD are instructed to imitate an action or utterance, they imitate with lower levels of accuracy and do so less frequently than typically developing (TD) counterparts (Edwards, 2014; Turan & Okcun Akcamus, 2012; Vivanti & Hamilton, 2014). Imitation deficits in ASD mainly manifest in high fidelity imitation of form, rather than in emulation of function or end points (Edwards, 2014). Functional magnetic resonance imaging studies suggest dysfunction of the mirror neuron system during action imitation in ASD (Yang & Hofmann, 2016).

Compared with other areas of imitation (e.g., action, object, and face), research on vocal imitation in ASD is relatively scarce and has only focused on either the speech or music domain. Although several studies have addressed vocal imitation of speech in ASD, results to date are mixed regarding whether and to what extent (pitch, duration, and/or the balance between the two) individuals with ASD are associated with speech imitation deficits. Some differences across studies may be due to the use of acoustic analyses versus the use of perceptual ratings. For instance, based on ratings by speech and language therapists, children with ASD had impaired imitation of various prosodic forms, including affect, intonation, chunking, and focus (McCann et al., 2007; McCann & Peppé, 2003; Peppé et al., 2007, 2011). By contrast, acoustic analyses of pitch range showed no difference across groups for imitation of stress, despite the fact that ASD participants received lower perceptual ratings of accuracy than TD controls (Paul et al., 2008). Thus, Van Santen et al. (2010) called attention to the unreliability and bias of clinicians' perceptual ratings (not strictly blind to participants' diagnostic status) and advocated the advantages and objectivity of instrumental methods. However, studies employing acoustic measures to assess imitation performance also produced divergent findings, with one study showing a group difference in duration only but not in mean pitch (Diehl & Paul, 2012) and other studies reporting both pitch and duration differences between groups (Fosnot & Jun, 1999; Hubbard & Trauner, 2007).

In contrast to speech, music has been seen as an area of exceptional skills in ASD (Molnar-Szakacs & Heaton, 2012; Ouimet et al., 2012). However, only one study has examined music imitation in ASD, and the results suggested that children with ASD showed comparable or better performance than controls when imitating pitch, rhythm, and duration of musical tones based on independent observers' judgment (Applebaum et al., 1979). Thus, regarding vocal imitation, ASD seems associated with atypical speech imitation but normal to superior music imitation. Given that vocal imitation is crucial for language acquisition (Kuhl, 2000; Kuhl & Meltzoff, 1996) and successful imitation requires sensorimotor, cognitive, and social skills (Fridland & Moore, 2015; Heyes, 2001; Nguyen & Delvaux, 2015; Over & Carpenter, 2013; Pagliarini et al., 2020), a potential impairment in vocal imitation may be related to landmark deficits of ASD including social and communicative difficulties (American Psychiatric Association, 2013; Diehl et al., 2015; McCann & Peppé, 2003). Previous studies have suggested that musical training benefits speech processing (Patel, 2011, 2012) and similar acoustic cues are used in emotional communication in music and speech (Juslin & Laukka, 2003). In addition, vocal imitation mechanisms are likely shared between speech and song production in adults with typical development (Mantell & Pfordresher, 2013; Wisniewski et al., 2013). Thus, the intimate link between music and speech begs the question as to whether vocal imitation impairment in ASD is indeed domain specific, especially when there has only been one study examining music imitation in ASD (Applebaum et al., 1979).

The domain specificity or generality of vocal imitation impairment in ASD is particularly relevant to a longstanding debate about whether speech and music share the same underlying processing systems (Albouy et al., 2020; Norman-Haignere et al., 2015; Zatorre & Gandour, 2008). The modular or domain-specific framework proposes that speech and music may involve distinct modules or mechanisms that deal with a particular aspect of the input and its output representation, either exclusively or more effectively than any other mechanisms (Fodor, 1983, 2001; Peretz, 2009; Peretz et al., 2015; Peretz & Coltheart, 2003; Peretz & Zatorre, 2005). While speaking and singing involve multiple processing components, musical abilities depend, in part, on modular processes such as tonal encoding of pitch, which is music-specific and independent of spoken pitch processing (Peretz, 2009; Peretz & Coltheart, 2003). In contrast to this view, others have suggested that speech and music systems may not be entirely modular or independent (Kunert & Slevc, 2015; Patel, 2013). Rather, there are shared or domain-general mechanisms underlying the processing of information across both domains (Koelsch, 2011; Koelsch & Siebel, 2005; Patel, 2008; Sammler et al., 2009). Numerous studies have provided evidence in support of either the domain-specific or domain-general view (Kunert & Slevc, 2015; Peretz et al., 2015). In addition to comparing music with language processing in typical development (Slevc

et al., 2009; Slevc & Miyake, 2006), neurodevelopmental disorders such as congenital amusia (Liu et al., 2010, 2013) and ASD (DePriest et al., 2017; Jiang et al., 2015) could offer special insight into this debate, particularly regarding whether deficits are only present in one domain (e.g., music), but not in the other (e.g., speech).

Specifically, as a functional output representation, vocal imitation of speech and song could inform the domain-specific versus domain-general debate from a production perspective (Peretz, 2009). Using matched speech and song stimuli, Liu et al. (2013) directly compared speech with music imitation in individuals with and without congenital amusia, a disorder of music processing (Ayotte et al., 2002). Individuals with congenital amusia demonstrated impaired pitch and duration matching of speech and song in terms of both absolute and relative measures. These findings suggest that vocal imitation mechanisms are likely shared between speech and song production even in congenital amusia (Liu et al., 2013). Although prior findings on vocal imitation in individuals with ASD seem to favor the domain-specific model, there have been too few published studies especially in music imitation in ASD to draw valid conclusions to inform the theoretical debate of music and language processing in this clinical population. Also, no studies have directly compared imitation abilities in speech versus music in ASD using matched linguistic and musical tasks to address the question of whether impaired speech imitation but spared/enhanced music imitation would be present in the same sample of participants. Furthermore, the absolute measures on pitch and duration matching used in Liu et al. (2013) required higher fidelity imitation than the relative measures. It remains to be determined whether individuals with ASD would show worse performance on absolute feature matching than relative feature matching during vocal imitation, similar to other areas of imitation in ASD (Edwards, 2014).

Studies of speech imitation in TD children and adults suggest that speech imitation ability is influenced by age (Kent & Forner, 1980; Loeb & Allen, 1993; Snow, 1998). Specifically, young children (aged 4) tended to imitate speech segments with longer duration and greater variability than older children (aged 6 and 12) and adults (Kent & Forner, 1980). While 3- and 4-year-olds showed more difficulty in imitating rising intonation contours in questions than falling intonation contours in statements, 5-year-olds were able to imitate both types of contours (Loeb & Allen, 1993; Snow, 1998). Studies of music imitation have examined pitch matching of tones or melodies in children (Cooper, 1995; Geringer, 1983) and adults (Amir et al., 2003; Pfordresher & Brown, 2007). Among children, pitch matching accuracy increases with age, with fourth-graders (9–10 years old) performing significantly better than third-graders (8–9 years old) (Cooper, 1995). While more than half of fourth-graders could match pitch within 50 cents (0.5 semitones), preschool children (4–5 years old) produced a median

deviation of 2.5 semitones (Geringer, 1983). Results on tempo matching or rhythm reproduction during music imitation suggest that, at 6 years, both musicians and non-musicians were able to reproduce rhythmic patterns embedded in a string of syllables (Gérard & Auxiette, 1992; Reifinger, 2006), and rhythmic response ability increased with age from Grade 1 to Grade 3 students (Schleuter & Schleuter, 1985). In addition, adults outperformed 5- to 7-year-old children on rhythm repetition, melody repetition, prosody repetition, as well as a range of other language and music tasks (Cohrdes et al., 2016). Thus, similar to speech imitation, music imitation ability is also influenced by age in typical development. In ASD, the evidence from studies on non-vocal imitation (e.g., action, object, and face) suggests that, although imitative abilities increase over time, impairments in imitation continue throughout the lifespan in ASD (Biscaldi et al., 2014; Vivanti & Hamilton, 2014; Young et al., 2011). However, how age influences speech and music imitation in ASD has not been systematically studied.

In the current study, we examined vocal imitation abilities in children and adults with and without ASD using matched speech and song stimuli, addressing three research questions: (1) Do imitation abilities of individuals with ASD differ from controls in terms of pitch and duration matching across speech and music domains? (2) Do individuals with ASD differ from controls with respect to relative and absolute feature matching in vocal imitation? (3) Do vocal imitation abilities in ASD and TD vary with age? Based on previous findings, we hypothesized that: (1) Participants with ASD would show impaired pitch and duration imitation in speech but not in song compared to controls; (2) Participants with ASD would show poorer performance on absolute feature matching than on relative feature matching as compared to controls; and (3) Across both groups, the adult cohort would perform better than the child cohort overall.

METHOD

Participants

A group of 44 individuals with ASD and 44 matched controls were recruited via a variety of methods including email lists, local social media advertisements, and local experimental participant databases. All were native British English speakers with no speech or hearing problems, and reported no history of other neurological or psychiatric disorders. Participants in the ASD group had a formal diagnosis of ASD by clinicians. Participants in the control group were included using the cut-off scores of 32 (adults), 30 (adolescents) or 76 (children) on the Autism-Spectrum Quotient (AQ) (Auyeung et al., 2008; Baron-Cohen et al., 2001, 2006). All participants had normal hearing in both ears, with pure-tone air

conduction thresholds of 25 dB HL or better at frequencies of 0.5, 1, 2, and 4 kHz. Participants' nonverbal IQ was estimated using the Raven's Standard Progressive Matrices Test (Raven et al., 1998), and verbal IQ was estimated by the Receptive One Word Picture Vocabulary Test IV (ROWPVT-IV) (Martin & Brownell, 2011). The Corsi block-tapping task was used to assess participants' nonverbal short-term memory span (Kessels et al., 2000), and the forward digit span task was used to assess verbal short-term memory (Wechsler, 2003). Participants were further divided into two age cohorts, children (<16) and adults (≥16), based on the age cut-off of 16 years, following the definition of adults in the AQ (Baron-Cohen et al., 2001). The reason that we used a two-way rather than a three-way split of age cohorts (children, adolescents, and adults) was to ensure that there were enough participants in each cohort. The age range of the child cohort was between 7.39 and 15.75 years and that of the adult cohort was between 16 and 56.75 years. Children's music perception skills were assessed using the Montreal Battery of Evaluation of Musical Abilities (MBEMA), which consists of five subtests with 20 trials each measuring the perception of scale, contour, interval, rhythm and recognition memory of musical melodies (Peretz et al., 2013). Adults were assessed using the Montreal Battery of Evaluation of Amusia (MBEA), which contains six subtests with 30 trials each measuring the perception of scale, contour, interval, rhythm, meter, and recognition memory of musical melodies (Peretz et al., 2003). All participants also completed a questionnaire about their musical, language, and medical background, where they were also asked to report whether they possess absolute pitch or perfect pitch, the ability to identify a musical note without a reference tone (Deutsch, 2013). As can be seen in Table 1, the ASD and control groups were comparable on all background measures. The study was approved by the University of Reading Research Ethics Committee. Written informed consent/assent was obtained from the participants and/or their parents prior to the experiment.

Stimuli

The target stimuli were 12 sentences either spoken or sung as statements or questions from Mantell and Pfordresher (2013), yielding 48 sentences with three to five syllables each. The speech stimuli were naturally spoken, and the pitch-time trajectory did not correspond to any diatonic scales. In order to create contour variation in the sequences, statements were produced with a falling contour and questions with a rising contour. The song stimuli comprised pitches from a major diatonic scale that approximated the global melodic contours of the speech stimuli. Each sung syllable had a roughly identical duration so as to invoke a metrical beat, resulting in the song stimuli being longer than the speech stimuli. Three

versions of the speech/song stimuli were used for different age and gender groups. The adult male and female versions were taken from Mantell and Pfordresher (2013) and used for male/female participants ≥12 years old. For child participants <12 years old, a child version was created by a child (female, 11-year-old, with 5 years of musical training) imitating the female version but in her own pitch range (see Figure 1; for more details, see Table S1).

Procedure

The presentation of the target stimuli and the recording of the imitations were both done using Praat (Boersma & Weenink, 2001). Participants were seated in a soundproof booth and presented with four practice trials (with items different from those in experimental trials: two speech vs. two song) to familiarize themselves with the task and the recording environment. Following the practice session, participants were presented with each of the 48 speech/song sentences one at a time in a pseudorandom order to ensure that different experimental conditions would alternate in an unpredictable manner and that long runs of the same condition (possible with true randomization) would not occur. Participants were instructed to imitate exactly the pitch and timing patterns of the sentences to the best of their ability, while their voices were recorded. Each sentence was played once and only replayed when participants failed to catch the words, and not when they wanted to listen to it again so they could imitate it better.

Data analysis

Recordings were analyzed in Praat using ProsodyPro (Xu, 2013) to extract the pitch and duration of each syllable rhyme. The rhyme was defined as the vowel portion of the syllable plus any final voiced consonant (e.g., **car**, b**oo**k), which was done by the first author (a phonetician). Octave errors in pitch imitation were corrected, that is, when imitated pitch was more than 6 semitones (half octave) apart from the model pitch, the value was adjusted as 12 – imitated pitch. In total, less than 3% of the data samples needed to be adjusted and most of these errors were caused by creaky voices, resulting in decreased fundamental frequency, F0 (Johnson, 2011). For accurate acoustic analysis of the data, we used ProsodyPro to manually add these missed vocal pulse marks for F0 based on the waveforms and spectrograms, to avoid having erroneous outliers misleading imitation results. Trials were not excluded when participants repeated the sentences slightly incorrectly but with the correct rhyme, for example, substituting "he" for "she" or "brought" for "bought." In the literature, pitch accuracy in singing and imitation has been analyzed using a variety of measures, such as using median F0 (Dalla Bella et al., 2007; Dalla Bella

TABLE 1 Characteristics of the ASD (n = 44) and control groups (n = 44)

Age cohort	Background measures	ASD	Control	W	p	Rank-biserial correlation
Children (n = 50)	Gender (F:M)	4:21	4:21			
	Age	11.41 (2.64)	11.17 (2.63)	335	0.67	0.07
	Musical training	2.18 (2.71)	1.76 (2.26)	335	0.66	0.07
	NVIQ	69.00 (26.14)	79.40 (23.06)	225	0.08	0.28
	VIQ	122.12 (15.43)	126.20 (14.96)	273	0.45	0.13
	Corsi	5.12 (1.30)	5.60 (1.38)	267.5	0.37	0.14
	Digit span	5.48 (1.05)	5.80 (0.91)	246.5	0.18	0.21
	Self-reported absolute pitch	**n = 2**	**n = 2**			
	Scale	16.64 (2.20)	17.29 (17.29)	232	0.17	0.23
	Contour	15.04 (3.41)	16.63 (2.70)	218.5	0.10	0.27
	Interval	15.52 (3.38)	16.50 (3.60)	239	0.22	0.20
	Rhythm	16.16 (3.38)	17.46 (3.05)	232.5	0.18	0.23
	Memory	16.48 (3.08)	17.75 (2.42)	227.5	0.14	0.24
	Pitch composite	47.20 (8.11)	50.42 (8.11)	224.5	0.13	0.25
	MBEMA global	79.84 (13.63)	85.62 (12.62)	216	0.10	0.28
Adults (n = 38)	Gender (F:M)	10:9	10:9			
	Age	34.51 (13.55)	33.74 (12.73)	184.5	0.92	0.02
	Musical training	4.26 (6.40)	5.50 (6.72)	165	0.65	0.09
	NVIQ	53.95 (28.07)	47.37 (31.11)	202.5	0.52	0.12
	VIQ	110.72 (13.75)	111.53 (14.09)	153	0.59	0.11
	Corsi	5.74 (1.45)	6.26 (1.15)	140.5	0.22	0.22
	Digit span	7.05 (1.62)	7.00 (1.11)	180	1.00	0.003
	Self-reported absolute pitch	**n = 2**				
	Scale	25.79 (3.41)	26.16 (2.57)	172	0.81	0.05
	Contour	25.11 (3.43)	25.53 (2.95)	177.5	0.94	0.02
	Interval	24.21 (3.17)	25.53 (2.95)	141	0.25	0.22
	Rhythm	26.00 (3.87)	26.63 (2.11)	186.5	0.87	0.03
	Meter	22.11 (9.38)	25.26 (6.98)	145.5	0.31	0.19
	Memory	27.11 (1.88)	26.42 (2.39)	205.5	0.47	0.14
	Pitch composite	75.11 (9.04)	77.21 (6.02)	163.5	0.63	0.09
	MBEA global	83.5 (10.5)	86.4 (7.3)	154	0.45	0.15

Note: Musical training: years of musical training; NVIQ: percentile point of Raven's Standard Progressive Matrices Test; VIQ: standard score of Receptive One Word Picture Vocabulary Test; Corsi: raw score of nonverbal short-term memory; Digit span: raw score of verbal short-term memory; The child cohort used the MBEMA with five subtests (Scale, Contour, Interval, Rhythm, and Memory) with 20 trials each, the pitch composite is the sum of the scale, contour, and interval scores, and the MBEMA global score is the percentage of correct responses out of the total 100 trials; the adult cohort used the MBEA with six subtests (Scale, Contour, Interval, Rhythm, Meter, Memory) with 30 trials each, the pitch composite is the sum of the scale, contour, and interval scores, and the MBEA global score is the percentage of correct responses out of the total 180 trials; 2-tailed Mann–Whitney-Wilcoxon Test results were used to compare group difference and effect size was given by the rank biserial correlation in the Mann–Whitney test.

Abbreviations: ASD, autism spectrum disorder; MBEA, Montreal battery of evaluation of amusia; MBEMA, Montreal battery of evaluation of musical abilities.

et al., 2009; Liu et al., 2013) or mean F0 (Hutchins & Peretz, 2012) of the vowel or vocalic group to indicate pitch height of each note/syllable, or calculating mean absolute pitch error and pitch correlation across the entire pitch trajectories of the model and imitated sequences (Mantell & Pfordresher, 2013). For timing accuracy, either subjective ratings, for example, 0 = "incorrect," 0.5 = "partly correct," and 1 = "correct" (Cohrdes et al., 2016), or objective acoustic analyses, for example, number of time errors as determined by a 25% time

deviation (Dalla Bella et al., 2007; Dalla Bella et al., 2009; Liu et al., 2013; Tremblay-Champoux et al., 2010) have been used. The pros and cons of these different methods and measures have been discussed (Dalla Bella, 2015). Since the ability to imitate/produce absolute versus relative features and pitch versus timing variables can dissociate in different "phenotypes" of poor singing (Berkowska & Dalla Bella, 2013; Dalla Bella & Berkowska, 2009), it is recommended that these dimensions be examined separately (Dalla Bella, 2015). Compared to mean F0, median
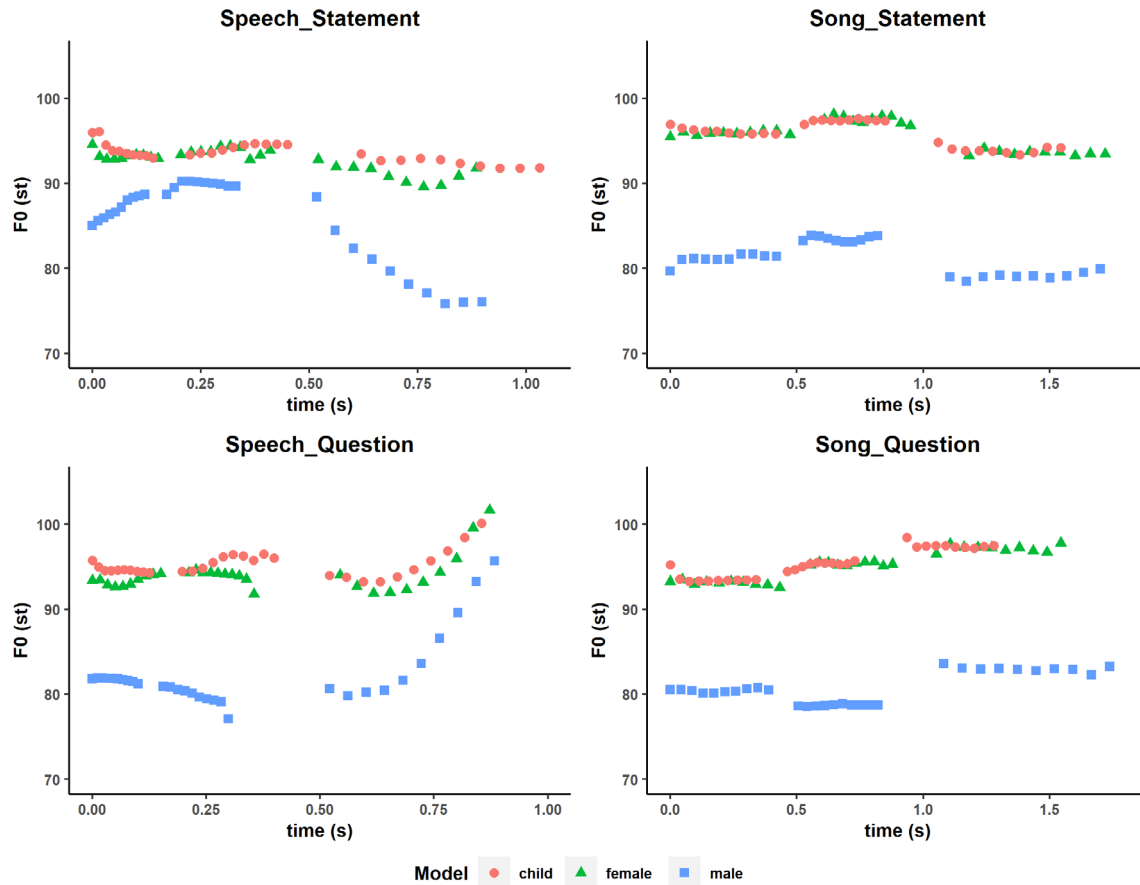
FIGURE 1 The pitch-time trajectory of the sentence "They went home" under different conditions by child/female/male target speakers

F0 is a preferable measure of pitch height, since it is less affected by extreme or erroneous variation of F0 due to creaky voice (Dalla Bella, 2015). In contrast to the whole trajectory analysis of each sequence (Mantell & Pfordresher, 2013), measuring the median F0 of each note/syllable rhyme (or vowel group) makes the calculation of pitch interval and pitch contour (two critical components in memory for melodies) between consecutive notes/syllables possible (Dowling & Fujitani, 1971). Most importantly, similar to music, there are pitch targets in speech across tone and intonation languages, such as high, low, rising, and falling, and they are realized based on linguistic functions and articulatory constraints (Xu, 2005; Xu & Prom-on, 2014; Xu & Wang, 2001). With a tonal perception model, speech prosody can be transcribed using a stylization of pitch levels and movements coupled with vocalic segments (Mertens, 2004), enabling the comparison of spoken and musical rhythm and melody (Patel et al., 2006). Thus, taking a comparative approach to studying music and language (Patel et al., 2006) and balancing the advantages and disadvantages of different methods (Dalla Bella, 2015), we adapted the following absolute and relative pitch and time measures from earlier studies (Dalla Bella et al., 2009; Dalla Bella, Deutsch, et al., 2007a; Liu et al., 2013, 2016; Tremblay-Champoux

et al., 2010) to compare imitation accuracy between music and speech in the current study.

The absolute pitch deviation (in cents): Median F0 was extracted from each syllable rhyme and then subtracted from that of their matched model to find the pitch deviation (in absolute value) for each imitated rhyme. The deviations were averaged over all syllables/notes in each utterance/melody, in order to control for the non-independence of the data points within each utterance/melody (Mcdonald, 2014). The bigger the value, the less accurate the imitation in terms of absolute pitch matching.

The relative pitch deviation (in cents): Pitch interval was calculated as the absolute difference in median F0 between two consecutive syllables/notes, and then subtracted from their matched model speaker's pitch interval (in absolute value). The deviations were averaged over all intervals in each utterance/melody and the bigger the value, the less accurate the imitation in terms of relative pitch matching.

The number of contour errors: Contour errors were defined as imitated pitch intervals that differed from the corresponding target pitch intervals in regard to pitch directions (up, down, or level). Pitch direction was considered to be up or down if the difference in pitch interval

was higher or lower by 50 cents (100 cents = 1 semitone) or more; otherwise (the difference was within 50 cents), the pitch intervals were considered to form a level/flat pitch direction. The number of contour errors was summed over each utterance/melody.

The number of pitch interval errors: Pitch interval errors were defined as imitated pitch intervals that were larger or smaller than the corresponding target pitch intervals by 100 cents without considering the pitch direction. Specifically, imitated and target pitch intervals were compared using absolute values. The number of pitch interval errors was summed over each utterance/melody.

The absolute duration difference (in milliseconds): Duration was extracted from each syllable rhyme and then subtracted from their matched model speaker's production to find the absolute difference for each rhyme. The differences were averaged over all rhymes in each utterance/melody and the larger the value, the less accurate the imitation in terms of absolute duration matching.

The relative duration difference (in milliseconds): Interonset interval (IOI) was calculated as the difference between two consecutive syllables/notes, and then subtracted from their matched model speaker's IOI (in absolute value). The differences were averaged over all IOIs in each utterance/melody and the larger the value, the less accurate the imitation in terms of relative duration matching.

The number of time errors: Time errors were defined as imitated syllables/notes that were more than 25% longer or shorter than the corresponding target syllables/notes (Dalla Bella et al., 2007; Dalla Bella et al., 2009; Prince & Pfordresher, 2012). This measure takes into account that in Western tonal music, event durations constitute simple integer ratio relationships, for example, sixteenth notes (1/4 a beat), eighth notes (1/2 a beat), quarter notes (1 beat), and so forth, and counting time errors this way will capture the violation to the time signature (Drake & Palmer, 2000). Similarly, in stress-timed languages such as English, speech rhythm can also be measured in relative terms, making the comparison of spoken and musical rhythm possible (Patel et al., 2006; Patel & Daniele, 2003). The number of time errors was summed over each utterance/melody.

Statistical analyses were conducted using Rstudio (RStudio Team, 2020). We performed linear mixed effects analysis using the *lme4* (Bates et al., 2015; Brauer & Curtin, 2018) package with the above-mentioned pitch and time variables as the dependent variable and Diagnostic Group (effect-coded: Control vs. ASD), Age cohorts (effect-coded: Child vs. Adult), and Condition (effect-coded: Speech vs. Music) as well as all possible interactions as fixed effects. All models were fit using the maximal random effects structure that converged with two random factors (subject vs. file) (Barr, 2013; Barr et al., 2013). When the maximal model failed to converge, the random correlations were removed

first. If the model still failed to converge, the random effect with the least variance was iteratively removed until the model converged. Statistical significance of the fixed effects was estimated using the *summary( )* function of the *lmerTest* package (Kuznetsova et al., 2017), which provided *p* values for the corresponding *t* tests. Subsequent post-hoc comparisons, if any, were conducted using the *emmeans* package (Lenth et al., 2018).

RESULTS

Absolute pitch deviation

Figure 2(a) shows boxplots of the absolute pitch deviations for the ASD and control groups. Results from the linear mixed-effects model (Table 2) revealed a significant main effect of Diagnostic Group, as the ASD group (M (SD) = 156.17(116.08)) produced significantly larger absolute pitch deviations than did the Control group (M (SD) = 124.48(97.45)). The main effect of Condition was also significant, with both groups showing better absolute pitch matching for the Music condition (M(SD) = 94.03 (84.96)) than for the Speech condition (M(SD) = 186.62 (109.39)). There was also a significant Condition by Age interaction and Post-hoc analyses with Bonferroni correction suggested that the child cohorts (M(SD) = 168.46 (98.4)) showed better absolute pitch matching than the adult cohorts (M(SD) = 210.52(118.23)) in the speech condition ($t(144) = 2.77$, $p = 0.006$), whereas comparable performance (Child: M(SD) = 104.3(87.68); Adult: M(SD) = 80.54(79.29)) was observed in the music condition ($t(144) = 1.56$, $p = 0.12$). No other remaining main effects and interactions were significant.

Additionally, to evaluate the performance of those participants who self-reported possessing absolute pitch, we closely inspected the results of these participants (Table 3). Given that the ASD group showed impaired imitation of absolute pitch, we took the values from the control group as the "standard" (M(SD) = 124.48(97.45), Range: 63.36–298.26) and only two of them performed better than the average level.

Relative pitch deviation

Figure 2(b) shows boxplots of the relative pitch deviations for the ASD and control groups. Results from the linear mixed-effects model revealed a significant main effect of Condition, with both groups showing better relative pitch matching for the Music condition (M (SD) = 86.5(67.1)) than for the Speech condition (M (SD) = 163.43(115.27)). The interaction between Condition and Age was also significant, and Post-hoc analyses with Bonferroni correction suggested that the child cohorts (M(SD) = 148.75(103.56)) showed better relative pitch matching than the adult cohorts (M(SD) = 182.8
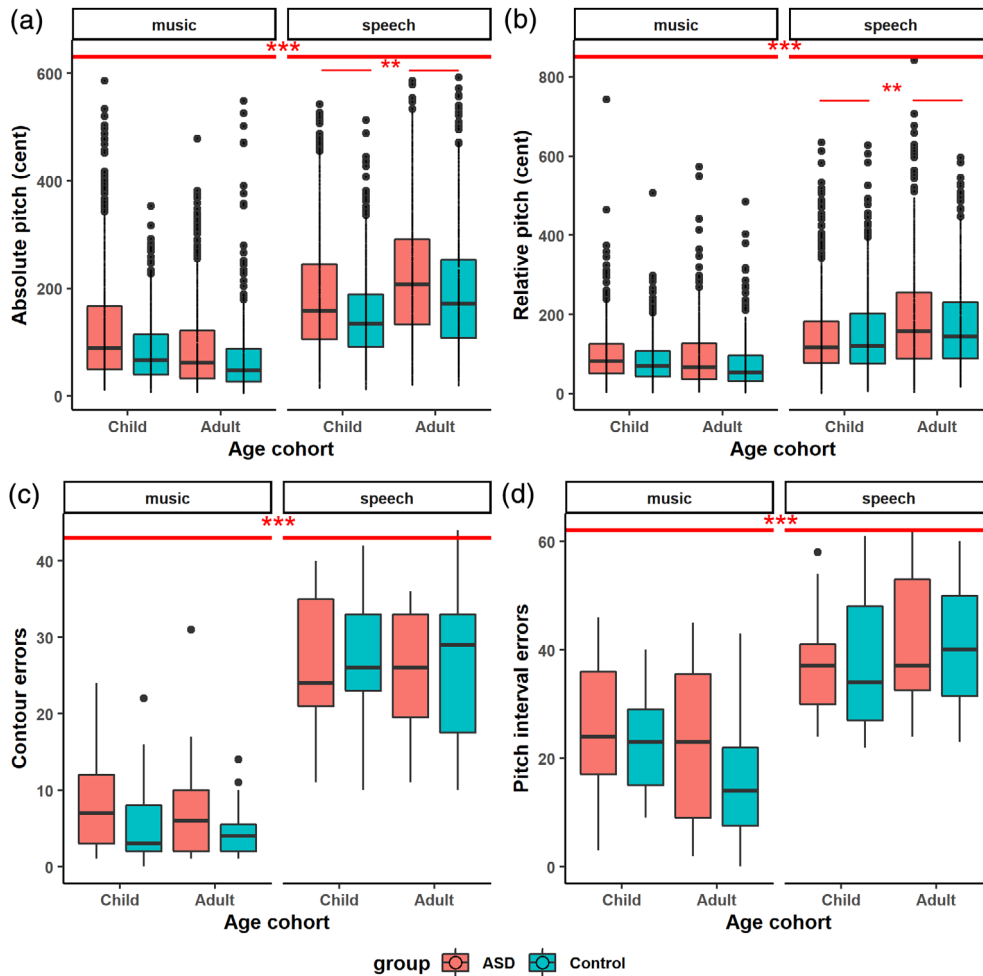
FIGURE 2 Boxplots of pitch-related measures for the ASD and control groups. (a) the absolute pitch deviation; (b) the relative pitch deviation; (c) the number of contour errors; (d) the number of pitch interval errors (asterisks represent *p*-values between variables with *$p < 0.05$, **$p < 0.01$ and ***$p < 0.001$). ASD, autism spectrum disorder

(126.6)) in the speech condition ($t(129) = 2.82$, $p = 0.006$), while comparable performance (Child: M(SD) = 90.47(64.14); Adult: M(SD) = 81.27(70.49)) was observed in the music condition ($t(129) = 0.76$, $p = 0.45$). No other remaining main effects and interactions were significant (see Table 2).

Number of contour errors

Figure 2(c) shows boxplots of the number of contour errors for the ASD and control groups. Results from the linear mixed-effects model revealed, as shown in Table 2, a significant main effect of Condition, with both groups showing fewer contour errors with the Music condition (M(SD) = 6.49(5.84)) than the Speech condition (M(SD) = 25.83(8.95)). No other remaining main effects and interactions were significant.

Number of pitch interval errors

Figure 2(d) shows boxplots of the number of pitch interval errors for the ASD and control groups. As shown in

Table 2, the linear mixed-effects model revealed a significant main effect of Condition, as both groups showed fewer pitch interval errors in the Music condition (M(SD) = 22.56(13.46)) than in the Speech condition (M(SD) = 39.28(11.2)). The interaction between Age and Condition was also significant, although Post-hoc analyses with Bonferroni correction revealed no significant difference between the child cohorts and adult cohorts in either condition (Speech: $t(138) = -1.54$, $p = 0.13$; Music: $t(138) = 1.92$, $p = 0.06$). No other remaining main effects and interactions were significant.

Absolute duration difference

Figure 3(a) shows boxplots of the absolute duration differences for the ASD and control groups. The linear mixed-effects model revealed, as shown in Table 4, a significant main effect of Group, as the ASD group (M(SD) = 57.92(40.09)) produced significantly larger absolute duration differences than did the Control group (M(SD) = 51.48(32.55)). The main effect of Condition was also significant, with both groups showing larger absolute duration differences in the Music condition (M

TABLE 2 Results from the linear mixed-effects models for the pitch-related measures.

	Estimate	Std. error	*df*	*t*	*p*
The absolute pitch deviations model					
Diagnostic group	**15.57**	**6.44**	**84.00**	**2.42**	**0.02***
Age	4.60	6.44	84.00	0.71	0.48
Condition	**−48.53**	**4.16**	**94.87**	**−11.67**	**<0.001***
Group: Age	−2.37	6.44	84.00	−0.37	0.71
Group: Condition	−0.39	3.80	84.00	−0.10	0.92
Age: Condition	**−16.49**	**4.09**	**94.06**	**−4.04**	**<0.001***
Group:Age:Condition	−1.41	3.80	84.00	−0.37	0.71
The relative pitch deviations model					
Diagnostic group	5.96	5.36	84.00	1.11	0.27
Age	6.30	5.36	84.00	1.17	0.24
Condition	**−39.97**	**3.74**	**67.65**	**−10.69**	**<0.001***
Group: Age	3.95	5.36	84.00	0.74	0.46
Group: Condition	2.72	2.90	84.02	0.94	0.35
Age: Condition	**−10.91**	**2.90**	**84.02**	**−3.76**	**<0.001***
Group:Age:Condition	−2.30	2.90	84.02	−0.79	0.43
The contour errors model					
Diagnostic group	0.03	0.03	84.08	1.12	0.27
Age	−0.03	0.04	64.06	−0.74	0.46
Condition	**−0.42**	**0.04**	**31.98**	**−9.63**	**<0.001***
Group: Age	0.007	0.03	84.08	0.24	0.81
Group: Condition	0.02	0.02	84.17	0.91	0.36
Age: Condition	0.008	0.02	84.18	0.42	0.67
Group:Age:Condition	−0.004	0.02	84.17	−0.18	0.86
The pitch interval errors model					
Diagnostic group	0.07	0.05	84.03	1.51	0.13
Age	−0.01	0.05	84.20	−0.23	0.82
Condition	**−0.38**	**0.03**	**52.38**	**−11.15**	**<0.001***
Group: Age	0.02	0.05	84.03	0.37	0.71
Group: Condition	0.04	0.03	83.95	1.76	0.08
Age: Condition	**−0.10**	**0.03**	**57.66**	**−3.04**	**0.004***
Group:Age:Condition	0.001	0.03	83.95	0.05	0.96

Note: *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

TABLE 3 The results of absolute pitch deviation for participants who self-reported possessing absolute pitch

ID	Diagnostic group	Age cohort	Absolute pitch deviation
Participant 1	ASD	Child	143.33
Participant 2	ASD	Child	76.70
Participant 3	ASD	Adult	173.24
Participant 4	ASD	Adult	156.85
Participant 5	Control	Child	78.43
Participant 6	Control	Child	180.60

Abbreviation: ASD, autism spectrum disorder.

(SD) = 71.73(41.31)) than in the Speech condition (M (SD) = 37.67(20.01)). No other remaining main effects and interactions were significant.

Relative duration difference

Figure 3(b) shows boxplots of the relative duration differences for the ASD and control groups. The linear mixed-effects model revealed a significant main effect of Condition, with both groups showing bigger relative duration differences in the Music condition (M(SD) = 55.64 (37.05)) than in the Speech condition (M(SD) = 39.75 (27.2)). No other remaining main effects and interactions were significant (see Table 4).

Number of time errors

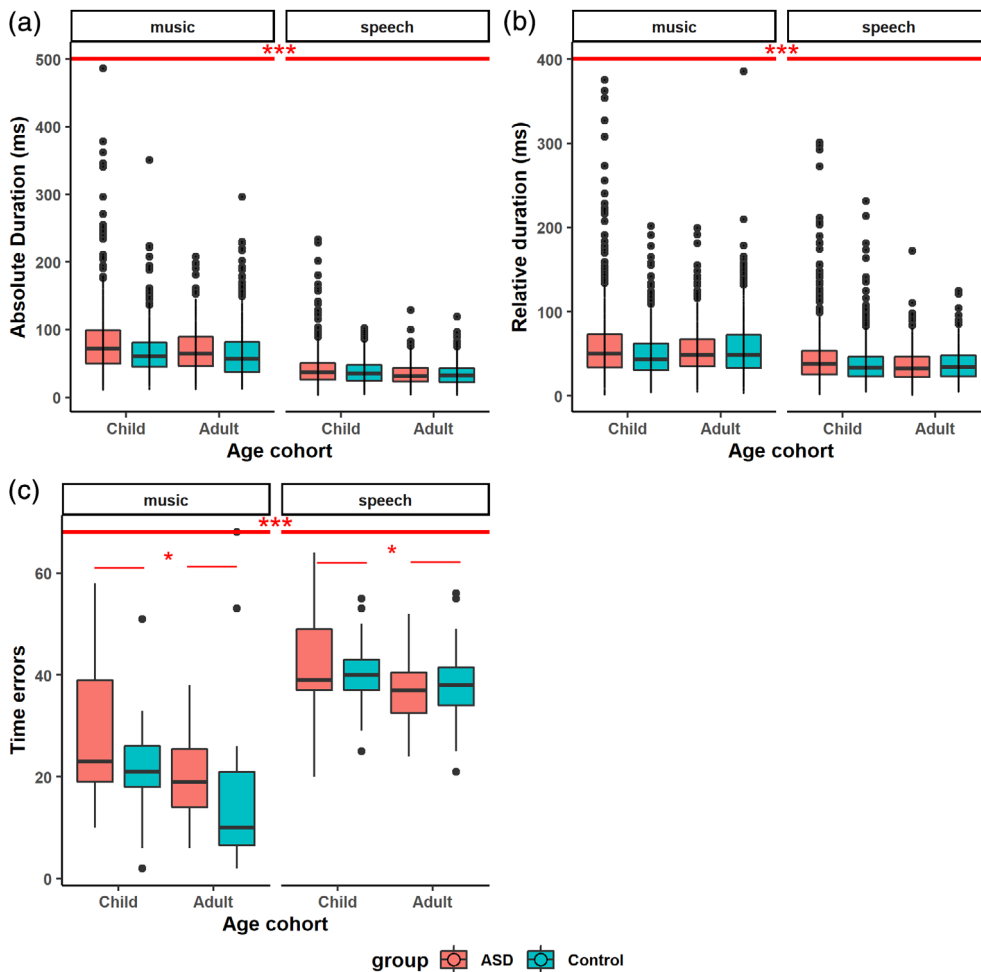Figure 3(c) shows boxplots of the number of time errors for the ASD and control groups. The linear mixed-

FIGURE 3 Boxplots of duration-related measures for the ASD and control groups. (a) the absolute duration difference; (b) the relative duration difference; (c) the number of time errors (asterisks represent *p*-values between variables with *$p < 0.05$, **$p < 0.01$ and ***$p < 0.001$). ASD, autism spectrum disorder

effects model revealed a significant main effect of Age, as the adult cohorts (M(SD) = 28.49(14.97)) produced fewer time errors than did the child cohorts (M(SD) = 33.19 (12.98)). There was a main effect of Condition, as both groups made fewer time errors with the Music condition (M(SD) = 22.39(13.1)) than with the Speech condition (M(SD) = 39.93(8.26)). No other remaining main effects and interactions were significant (see Table 4).

DISCUSSION

The present study investigated imitation of speech and song in English-speaking individuals with and without ASD and its modulation by age using absolute and relative pitch and duration measures. The main results showed that individuals with ASD were worse than controls on absolute pitch and duration matching, while performing as well as controls on relative pitch and duration matching in both speech and song imitation. In addition, the two groups produced similar numbers of pitch contour errors, pitch interval errors, and time errors. Furthermore, like the controls, individuals with ASD

imitated sung pitch more accurately than spoken pitch, whereas spoken duration was imitated more accurately than sung duration. Across both groups, children tended to imitate pitch more accurately than adults when it came to speech stimuli rather than song stimuli, whereas adults made fewer time errors than did children in both stimulus types.

In terms of absolute feature matching during vocal imitation, we discovered impaired performance in the ASD group for both pitch and duration across both speech and song conditions as compared to the control group. This finding is in line with previous results showing impaired imitation of form in ASD (Edwards, 2014). A few previous studies also showed impaired pitch and duration imitation for speech in ASD (Fosnot & Jun, 1999; Hubbard & Trauner, 2007). However, other studies indicated that speech imitation deficits in ASD only manifested in duration (Diehl & Paul, 2012; Paul et al., 2008). The discrepancy may be related to the different methods used to measure imitation performance across the studies. While we compared group differences in imitation by measuring how well participants in each group matched the pitch and duration features of the model utterances, previous studies ignored the model but

TABLE 4 Results from the linear mixed-effects model for the duration-related measures.

	Estimate	Std. error	df	t	p
The absolute duration differences model					
Diagnostic group	**2.96**	**1.48**	**84.01**	**2.01**	**0.048***
Age	−2.73	1.51	87.62	−1.80	0.07
Condition	**17.00**	**1.46**	**62.19**	**11.64**	**<0.001*****
Group: Age	−1.94	1.48	84.01	−1.31	0.19
Group: Condition	2.03	1.07	84.00	1.90	0.06
Age: Condition	−0.26	1.07	84.00	−0.25	0.81
Group:Age:Condition	−0.95	1.07	84.00	−0.89	0.38
The relative duration differences model					
Diagnostic group	2.07	1.59	85.63	1.30	0.20
Age	−1.36	1.58	84.15	−0.86	0.39
Condition	**8.17**	**1.31**	**64.56**	**6.22**	**<0.001*****
Group: Age	−2.95	1.57	83.87	−1.88	0.06
Group: Condition	0.28	1.04	78.87	0.27	0.79
Age: Condition	1.62	1.08	80.44	1.50	0.14
Group:Age:Condition	−0.93	1.01	83.79	−0.92	0.36
The time errors model					
Diagnostic group	0.07	0.04	83.96	1.66	0.10
Age	**−0.10**	**0.05**	**87.60**	**−2.30**	**0.02***
Condition	**−0.39**	**0.04**	**48.47**	**−10.40**	**<0.001*****
Group: Age	−0.02	0.04	83.96	−0.53	0.60
Group: Condition	0.05	0.03	84.07	1.97	0.052
Age: Condition	−0.04	0.03	57.96	−1.09	0.28
Group:Age:Condition	0.0009	0.03	84.07	0.03	0.97

Note: *p < 0.05, **p < 0.01, and ***p < 0.001.

compared the pitch and duration patterns of the produced utterances across groups (Diehl & Paul, 2012; Paul et al., 2008). Thus, as in previous vocal imitation studies (Liu et al., 2013; Mantell & Pfordresher, 2013), we measured imitation abilities by comparing acoustic parameters between the model and imitated utterances, and the smaller the difference, the more accurate the imitation. Using this method, we were able to reveal differences in absolute feature matching during imitation between groups. However, previous studies only showed the differences in characteristics between the produced utterances of the two groups (Diehl & Paul, 2012; Paul et al., 2008), thus measuring speech production, rather than imitation accuracy.

In contrast to the intact musical imitation abilities reported in a previous study on children with ASD (Applebaum et al., 1979), our finding demonstrated that both children and adults with ASD were impaired in absolute pitch and duration matching for song imitation. One explanation for this discrepancy may be related to how the accuracy of imitation was calculated. Specifically, Applebaum et al. (1979) relied on subjective perceptual ratings of imitation accuracy by two independent observers, whereas the current study employed objective acoustic analyses. A second possible explanation relates

to the difference in sample size. While 88 participants (44 per group) were involved in the present study, only six individuals participated in Applebaum et al.'s (1979) study (three per group). Thus, the current results may be more reliable given the objective acoustic analyses and a larger sample size.

Despite impaired absolute pitch and duration matching, individuals with ASD showed comparable performance to controls on relative pitch and duration matching, as well as on other measures of relative-feature matching (e.g., number of pitch contour, pitch interval, and time errors). Our results are consistent with previous findings on poor singers (Berkowska & Dalla Bella, 2009; Dalla Bella & Berkowska, 2009). For instance, Dalla Bella and Berkowska (2009) examined occasional singers' pitch and duration accuracy in terms of both absolute and relative features when spontaneously producing well-known melodies, as well as when imitating these melodies with a metronome at a slower tempo. They found that poor singers performed less accurately in the absolute measures than the relative measures and suggested that the production of relative and absolute pitch and time features may be independent in the music domain. Our results extend those of Dalla Bella and Berkowska (2009), showing that the dissociation between relative and

absolute pitch and duration matching is also the case for impaired vocal imitation in ASD and that the dissociation exists not only in music but also in speech.

However, to the best of our knowledge, no previous studies in ASD have examined relative versus absolute feature or relative feature matching alone in either speech or music imitation in ASD, which makes it difficult to find evidence to explain why individuals with ASD showed preserved relative but impaired absolute pitch and duration matching during vocal imitation. We propose two possibilities for the divergent results of absolute versus relative feature matching in ASD below, which would require further investigations by future studies. First, one possibility might relate to the differential requirement for fidelity of imitation between absolute and relative features. There has been extensive evidence from non-vocal studies (e.g., action, objects, and face) suggesting that individuals with ASD manifest impaired imitation ability in tasks that require high fidelity imitation, such as reproducing precisely both the form and the end result of a model (Edwards, 2014). However, tasks requiring lower fidelity, such as emulation that only requires reproducing the final result/goal without considering the forms needed to achieve the final goal, generally fail to observe deficits in the ASD group (Hamilton, 2008; Edwards, 2014). In our study, the absolute measures required higher fidelity imitation compared to the relative measures. In particular, absolute measures examined the exact matching of pitch and duration features for each syllable/note, while relative measures assessed the matching of the relative pitch and timing relationship between two consecutive syllables/notes. Thus, our current results indicate for the first time that, consistent with non-vocal imitation studies (Hamilton, 2008; Edwards, 2014), individuals with ASD show impaired vocal imitation ability in tasks requiring high fidelity (i.e., absolute feature matching), but not in tasks requiring lower fidelity (i.e., relative feature matching). Second, it is possible that the imitation mode (relative vs. absolute) that participants were experiencing during vocal imitation may account for the dissociation. Specifically, evidence from perception research in TD indicates that, as children mature from 3 to 6 years, there is a general developmental shift from an absolute to a relative mode in pitch perception (Crozier, 1997; Saffran & Griepentrog, 2001; Sergeant & Roche, 1973; Takeuchi & Hulse, 1993). Studies also found that while adults relied primarily on relative pitch cues, they were able to access absolute cues under certain conditions (Saffran & Griepentrog, 2001), and both children and adults demonstrated absolute memory of familiar melodies (Levitin, 1994; Schellenberg & Trehub, 2003, 2008). Taking these findings together, it is possible that different participants may depend on different perception modes when imitating speech and song. The reason that the two groups did not differ significantly in relative pitch and duration matching may be because participants all

tended to the relative cues. While controls also accessed absolute cues during the process, participants with ASD did not or were less capable of doing so. Future studies are required to test this possibility by examining the relationship between perception and production during vocal imitation.

When using acoustically matched speech and song stimuli testing the same sample of participants, we observed impairments (i.e., absolute pitch and duration matching) as well as preserved skills (i.e., relative pitch and duration matching) in ASD not only in speech but also in music. Hence, compatible with the findings of vocal imitation in people with typical development (Mantell & Pfordresher, 2013) and those with congenital amusia (Liu et al., 2013), vocal imitation also constitutes domain-general mechanisms in individuals with ASD. These findings provide support for using music therapy to improve speech for those individuals with ASD who manifest deficits in language (James et al., 2015). In addition, successful imitation requires sensorimotor, cognitive, and social skills (Fridland & Moore, 2015; Heyes, 2001; Nguyen & Delvaux, 2015; Over & Carpenter, 2013; Pagliarini et al., 2020). Thus, the benefit of music imitation may extend to improving cognitive and social skills in ASD (Boster et al., 2020).

It has been reported that absolute pitch (AP) ability is more common among individuals with ASD than in nonclinical populations (Heaton et al., 1998; Mottron et al., 1999; Stanutz et al., 2014). However, the present imitative results were not in line with these findings. Rather, we found that individuals with ASD showed impaired absolute pitch and duration matching. While we did not test our participants' receptive AP ability in the current study, we did ask whether they have absolute pitch (or perfect pitch, the ability to identify a musical note without a reference tone) in a questionnaire. According to the self-reports, two children with ASD (out of 25) and two adults with ASD (out of 19), as well as two control children (out of 25) possessed AP. However, they did not perform exceptionally when imitating absolute pitch, which suggests that receptive AP may not transfer to expressive AP in imitation. This finding is consistent with the dual-route model, which posits that vocal stimuli are processed for motor-relevant features and conscious, symbolic representations along two different, independent pathways (Hutchins & Moreno, 2013). Thus, vocal perception and production abilities could be uncorrelated, and each can be intact or impaired without affecting the other (Griffiths, 2008; Hutchins & Moreno, 2013; Loui, 2015). Notably, our findings are based on self-reports rather than experimental testing of AP. Further studies are needed to clarify the nature of receptive and productive AP in ASD.

Moreover, the current study examined whether speech and song imitation abilities vary with age in ASD and controls. Across both groups, adults made fewer time errors in both speech and song imitation relative to the

child cohort. Time errors were defined as deviation from the target duration by 25%, and this is the only measure where the two age cohorts differed significantly, but not in other timing matching measures (e.g., absolute and relative duration matching). These results suggest that while both children and adults can imitate the duration of speech and song segments comparably, children may have greater duration variability than adults when errors were measured relative to the duration of each segment. The findings are in agreement with previous research indicating that there is a developmental decrease in duration variability (Kent & Forner, 1980; Munson, 2004; Smith, 1978). Indeed, children possess less refined neuromotor capabilities than adults (Smith, 1978), and they are unable to exert adult-like control of speech production mechanisms. Hence, children's output reflects greater variability of phonetic segments compared to adults (Kent & Forner, 1980; Koenig et al., 2008; Munson, 2004; Smith et al., 1996).

Conversely to what was observed in timing matching, across both groups, children tended to imitate absolute and relative pitch more accurately than adults when it came to speech stimuli rather than song stimuli. This result may be due to children attending to speech pitches more readily than adults. Speech imitation is based on intentional understanding (Over & Gattis, 2010). Individuals thus tend to imitate the functional goal (e.g., statements with falling pitch contours vs. questions with rising pitch contours) rather than copying the exact form of the utterances (Liu et al., 2010, 2013). In the present study, this tendency appeared more pronounced in adults than in children. Indeed, we did not find any differences between the child and adult cohorts in the pitch contour imitation, as they all preferred to and were able to imitate the functional goals (rising vs. falling). However, adults neglected form-related information in speech more saliently, resulting in poorer performance than children on exact matching of absolute and relative pitch. On the other hand, the results could also mean that children do not make as strong distinctions between speech and song as adults do. Studies have shown that, unlike musical communication, speech comprehension is remarkably robust to lack of detail in pitch variation (Liu et al., 2015; Patel, 2011; Patel et al., 2010). This is because the need for pitch precision in speech can be relaxed by integrating multiple context-based cues (including the voice onset time, vowel length, fundamental frequency, and first and second formant patterns) and knowledge sources (including semantics, syntax, and pragmatics) (Mattys et al., 2005; Toscano et al., 2010; Toscano & McMurray, 2010). However, since the integrating abilities in children are not as mature as adults (McCreery & Stelmachowicz, 2011; Stelmachowicz et al., 2000), they may still mainly rely on pitch cues in speech imitation as they do in music imitation.

Generally speaking, we did not observe the developmental increase in imitative abilities that has been suggested by previous studies in speech (Kent & Forner, 1980; Loeb & Allen, 1993; Snow, 1998) and music (Cohrdes et al., 2016; Cooper, 1995; Geringer, 1983), except in the duration variability. One explanation for this discrepancy may be related to differences in age of the participants among the studies. The youngest child participant in the present study was 7.39 years old, whilst several previous studies examined the development from 3 to 5 years (Loeb & Allen, 1993; Snow, 1998) or 5–7 years (Cohrdes et al., 2016). Thus, it is possible that the present task was too simple to reveal the developmental change for participants beyond 7 years old, since 5-year-olds were already able to imitate falling versus rising contours (Loeb & Allen, 1993; Snow, 1998). In addition, the different grouping of age cohorts between studies might also account for the discrepancy in findings. Specifically, we grouped participants below 16 into the child cohort and those above 16 into the adult cohort, and age-related differences were then examined by comparing these two age cohorts. However, previous studies compared age-related differences at year-level (Cooper, 1995; Geringer, 1983; Kent & Forner, 1980; Loeb & Allen, 1993; Snow, 1998), for example, comparing 5 years with 4 years (Loeb & Allen, 1993; Snow, 1998). Thus, subtle developments over time may be masked in the present study, given the wide age range within each age cohort. Across our pre-defined age cohorts, however, there was no significant age × group interaction on any of the absolute or relative pitch or duration measures we examined. This suggests that age (≥16 or <16 years) influences speech and music imitation similarly across ASD and TD. Thus, our results on vocal imitation corroborate previous findings of persistent impairments in other areas of imitation across the lifespan in ASD (Biscaldi et al., 2014; Vivanti & Hamilton, 2014; Young et al., 2011).

Consistent with previous studies (Liu et al., 2013; Mantell & Pfordresher, 2013; Wisniewski et al., 2013), both the ASD and control groups imitated song more accurately than did speech across all pitch-related measures. Several possibilities may explain this result. First, the reason for the enhanced pitch imitation in songs may be because, in order to achieve adequate communication, a higher degree of pitch precision is required for conveying musical meaning than speech meaning (Patel, 2008, 2011). Indeed, even individuals with congenital amusia imitated musical pitch better than linguistic pitch, since music is form-driven and speech is function-driven (Liu et al., 2013). Studies of intonation imitation among TD adults also suggest that English speakers tended to imitate the phonological structure (e.g., pitch accent, intonational phrase boundary), rather than the phonetic details (e.g., pause duration, irregular pitch periods), of intonation (Cole & Shattuck-Hufnagel, 2011). Thus, the worse pitch matching in speech imitation compared to song imitation may be because people tend to imitate the functional goal, rather than the exact form, of speech utterances (Liu et al., 2010). Secondly, one may argue that the slower tempo in songs might have positively

affected pitch imitation, since singing accuracy improves considerably when people sing at slower as opposed to faster tempos (Dalla Bella et al., 2007). However, even when durations were equated across speech and song stimuli, the pitch imitation advantage for singing remained (Mantell & Pfordresher, 2013). Thus, the enhanced sung pitch imitation cannot simply be attributed to differences in the rate of speech versus song stimuli in the present study.

Interestingly, our results on duration matching across speech and song imitation indicate that the effect of domain is not equivalent across different duration measures. Whereas both groups achieved better absolute and relative duration matching in speech than in song, they made fewer time errors in song than in speech. The fewer time errors in song than in speech may not necessarily imply that duration matching during song imitation was superior to speech imitation for both groups. This is because the more time errors in speech than in song may be caused by the higher time precision required in speech. Specifically, given that our time errors were defined using deviation from 25% of the target duration and that song stimuli contained longer target durations than speech stimuli (see Figure 1), the accuracy requirement was higher for speech than for song imitation. The results of more accurate imitation for both absolute and relative duration in speech than in song are consistent with previous findings (Albouy et al., 2020; Liu et al., 2013; Mantell & Pfordresher, 2013). Overall, both groups tended to be particularly sensitive to the appropriateness of duration in speech compared to that in song, and were more sensitive to pitch in song than in speech, suggesting that pitch imitation is independent from the imitation of duration (Dalla Bella et al., 2007; Dalla Bella et al., 2009; Drake & Palmer, 2000; Mantell & Pfordresher, 2013).

A caveat about the design of the current study is the stimuli we used for participants to imitate in our experiment. Taken from Mantell and Pfordresher (2013), the stimuli for the adult male and female versions were produced in American English with a midland dialect (male speaker) and an inland North dialect (female speaker). The child version was created by a child imitating the female model in British English. Since all our participants were British English speakers, one may wonder whether or to what extent the different dialects have affected imitative performance. However, research has shown that speakers can imitate detailed intonational patterns of a different variety of their language (D'Imperio et al., 2014). Similar to controls, individuals with ASD who have good language abilities can perceive acoustic differences in dialects as well as use these cues to group the dialects into the areas they come from (Clopper et al., 2012). Consistent with these findings, children, female and male adults with ASD and their TD counterparts in our study performed comparably in imitation of relative pitch and duration, suggesting that different dialects may not have affected imitation performance of our participants. In addition, impairments of absolute pitch and duration matching in ASD were observed not only in the adult cohort but also in the child cohort who imitated British English. Taken together, while the effect of dialect is unlikely to have influenced the current results, further studies are required to examine this specific hypothesis.

CONCLUSION

Using sentences and melodies that shared critical features, this study revealed for the first time that vocal imitative skills in individuals with ASD are impaired in absolute pitch and duration matching but intact in relative pitch and duration matching across speech and music domains. From children to adults, vocal imitation showed an improvement in the number of time errors across speech and song, but a decrease in pitch imitation accuracy in the speech condition only. These findings support the idea that speech and song imitation may involve shared cognitive and motor mechanisms, which may have implications for the development of language in individuals with ASD (Stone & Yoder, 2001).

ORCID
Li Wang https://orcid.org/0000-0001-5318-2408
Peter Q. Pfordresher https://orcid.org/0000-0002-7024-0785
Cunmei Jiang https://orcid.org/0000-0002-0264-5924
Fang Liu https://orcid.org/0000-0002-7776-0222

REFERENCES
Albouy, P., Benjamin, L., Morillon, B., & Zatorre, R. J. (2020). Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science*, *367*(6481), 1043–1047. https://doi.org/10.1126/science.aaz3468

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596

Amir, O., Amir, N., & Kishon-Rabin, L. (2003). The effect of superior auditory skills on vocal accuracy. *The Journal of the Acoustical Society of America*, *113*(2), 1102–1108. https://doi.org/10.1121/1.1536632

Applebaum, E., Egel, A. L., Koegel, R. L., & Imhoff, B. (1979). Measuring musical abilities of autistic children. *Journal of Autism and Developmental Disorders*, *9*(3), 279–285. https://doi.org/10.1007/BF01531742

Auyeung, B., Baron-Cohen, S., Wheelwright, S., & Allison, C. (2008). The autism Spectrum quotient: Children's version (AQ-child).

Journal of Autism and Developmental Disorders, *38*(7), 1230–1240. https://doi.org/10.1007/s10803-007-0504-z

Ayotte, J., Peretz, I., & Hyde, K. L. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain: A Journal of Neurology*, *125*(Pt 2), 238–251. https://doi.org/10.1093/brain/awf028.

Baron-Cohen, S., Hoekstra, R. A., Knickmeyer, R., & Wheelwright, S. (2006). The autism-spectrum quotient (AQ)—Adolescent version. *Journal of Autism and Developmental Disorders*, *36*(3), 343–350. https://doi.org/10.1007/s10803-006-0073-6

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-Spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, Malesand females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/A:1005653411471

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 328. https://doi.org/10.3389/fpsyg.2013.00328

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. http://dx.doi.org/10.18637/jss.v067.i01.

Berkowska, M., & Dalla Bella, S. (2009). Acquired and congenital disorders of sung performance: A review. *Advances in Cognitive Psychology*, *5*, 69–83. https://doi.org/10.2478/v10053-008-0068-2

Berkowska, M., & Dalla Bella, S. (2013). Uncovering phenotypes of poor-pitch singing: The sung performance battery (SPB). *Frontiers in Psychology*, *4*, 714. https://doi.org/10.3389/fpsyg.2013.00714

Bernier, R., Dawson, G., Webb, S., & Murias, M. (2007). EEG mu rhythm and imitation impairments in individuals with autism spectrum disorder. *Brain and Cognition*, *64*(3), 228–237. https://doi.org/10.1016/j.bandc.2007.03.004

Biscaldi, M., Rauh, R., Irion, L., Jung, N. H., Mall, V., Fleischhaker, C., & Klein, C. (2014). Deficits in motor abilities and developmental fractionation of imitation performance in high-functioning autism spectrum disorders. *European Child & Adolescent Psychiatry*, *23*(7), 599–610. https://doi.org/10.1007/s00787-013-0475-x

Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*, 341–345.

Boster, J. B., Spitzley, A. M., Castle, T. W., Jewell, A. R., Corso, C. L., & McCarthy, J. W. (2020). Music improves social and participation outcomes for individuals with communication disorders: A systematic review. *Journal of Music Therapy*, *58*, 12–42. https://doi.org/10.1093/jmt/thaa015

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*(3), 389–411. https://doi.org/10.1037/met0000159

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., & Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive Development*, *15*(4), 481–498. https://doi.org/10.1016/S0885-2014(01)00037-5

Clopper, C. G., Rohrbeck, K. L., & Wagner, L. (2012). Perception of dialect variation by Young adults with high-functioning autism. *Journal of Autism and Developmental Disorders*, *42*(5), 740–754. https://doi.org/10.1007/s10803-011-1305-y

Cohrdes, C., Grolig, L., & Schroeder, S. (2016). Relating language and music skills in Young children: A first approach to systemize and compare distinct competencies on different levels. *Frontiers in Psychology*, *7*, 1616. https://doi.org/10.3389/fpsyg.2016.01616

Cole, J., & Shattuck-Hufnagel, S. (2011). The phonology and phonetics of perceived prosody: What do listeners imitate? *Proceedings of Interspeech*, *2011*, 969–972.

Cooper, N. A. (1995). Children's singing accuracy as a function of grade level, gender, and individual versus unison singing. *Journal of Research in Music Education*, *43*(3), 222–231. https://doi.org/10.2307/3345637

Cossu, G., Boria, S., Copioli, C., Bracceschi, R., Giuberti, V., Santelli, E., & Gallese, V. (2012). Motor representation of actions in children with autism. *PLoS One*, *7*(9), e44779. https://doi.org/10.1371/journal.pone.0044779

Crozier, J. B. (1997). Absolute pitch: Practice makes perfect, the earlier the better. *Psychology of Music*, *25*(2), 110–119.

Dalla Bella, S. (2015). Defining poor-pitch singing: A problem of measurement and sensitivity. *Music Perception*, *32*(3), 272–282. https://doi.org/10.1525/mp.2015.32.3.272

Dalla Bella, S., & Berkowska, M. (2009). Singing proficiency in the majority: Normality and "phenotypes" of poor singing. *Annals of the New York Academy of Sciences*, *1169*, 99–107. https://doi.org/10.1111/j.1749-6632.2009.04558.x

Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2009). Singing in congenital amusia. *The Journal of the Acoustical Society of America*, *126*(1), 414–424. https://doi.org/10.1121/1.3132504

Dalla Bella, S., Giguère, J.-F., Peretz, I., & Deutsch, D. (2007). Singing proficiency in the general population. *The Journal of the Acoustical Society of America*, *121*(2), 1182–1189. https://doi.org/10.1121/1.2427111

DePriest, J., Glushko, A., Steinhauer, K., & Koelsch, S. (2017). Language and music phrase boundary processing in autism Spectrum disorder: An ERP study. *Scientific Reports*, *7*(1), 14465. https://doi.org/10.1038/s41598-017-14538-y

Deutsch, D. (2013). Absolute Pitch. In *The psychology of music* (pp. 141–182). Elsevier. https://doi.org/10.1016/B978-0-12-381460-9.00005-5

Diehl, J. J., Friedberg, C., Paul, R., & Snedeker, J. (2015). The use of prosody during syntactic processing in children and adolescents with autism spectrum disorders. *Development and Psychopathology*, *27*(3), 867–884. https://doi.org/10.1017/S0954579414000741

Diehl, J. J., & Paul, R. (2012). Acoustic differences in the imitation of prosodic patterns in children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *6*(1), 123–134. https://doi.org/10.1016/j.rasd.2011.03.012

D'Imperio, M., Cavone, R., & Petrone, C. (2014). Phonetic and phonological imitation of intonation in two varieties of Italian. *Frontiers in Psychology*, *5*, 1226. https://doi.org/10.3389/fpsyg.2014.01226

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *The Journal of the Acoustical Society of America*, *49*(2B), 524–531. https://doi.org/10.1121/1.1912382

Drake, C., & Palmer, C. (2000). Skill acquisition in music performance: Relations between planning and temporal control. *Cognition*, *74*(1), 1–32. https://doi.org/10.1016/S0010-0277(99)00061-X

Edwards, L. A. (2014). A meta-analysis of imitation abilities in individuals with autism Spectrum disorders. *Autism Research*, *7*(3), 363–380. https://doi.org/10.1002/aur.1379

Fodor, J. A. (1983). *The modularity of mind*. MIT Press.

Fodor, J. A. (2001). *The mind doesn't work that way: The scope and limits of computational psychology*. MIT Press.

Fosnot, S. M., & Jun, S.-A. (1999). Prosodic characteristics in children with stuttering or autism during reading and imitation. Proceedings of the 14th International Congress of Phonetic Sciences, 1925-1928.

Fridland, E., & Moore, R. (2015). Imitation reconsidered. *Philosophical Psychology*, *28*(6), 856–880. https://doi.org/10.1080/09515089.2014.942896

Gérard, C., & Auxiette, C. (1992). The processing of musical prosody by musical and nonmusical children. *Music Perception*, *10*(1), 93–125. https://doi.org/10.2307/40285541

Geringer, J. M. (1983). The relationship of pitch-matching and pitch-discrimination abilities of preschool and fourth-grade students. *Journal of Research in Music Education*, *31*(2), 93–99. https://doi.org/10.2307/3345213

Griffiths, T. D. (2008). Sensory systems: Auditory action streams? *Current Biology: CB*, *18*(9), R387–R388. https://doi.org/10.1016/j.cub.2008.03.007

Ham, H. S., Bartolo, A., Corley, M., Rajendran, G., Szabo, A., & Swanson, S. (2011). Exploring the relationship between gestural recognition and imitation: Evidence of dyspraxia in autism Spectrum disorders. *Journal of Autism and Developmental Disorders*, *41*(1), 1–12. https://doi.org/10.1007/s10803-010-1011-1

Hamilton, A. F. D. C. (2008). Emulation and mimicry for social interaction: A theoretical approach to imitation in autism. *Quarterly Journal of Experimental Psychology*, *61*(1), 101–115. http://dx.doi.org/10.1080/17470210701508798.

Heaton, P., Hermelin, B., & Pring, L. (1998). Autism and pitch processing: A precursor for savant musical ability? *Music Perception: An Interdisciplinary Journal*, *15*(3), 291–305. https://doi.org/10.2307/40285769

Heyes, C. (2001). Causes and consequences of imitation. *Trends in Cognitive Sciences*, *5*(6), 253–261. https://doi.org/10.1016/S1364-6613(00)01661-2

Hubbard, K., & Trauner, D. A. (2007). Intonation and emotion in autistic Spectrum disorders. *Journal of Psycholinguistic Research*, *36*(2), 159–173. https://doi.org/10.1007/s10936-006-9037-4

Hutchins, S., & Moreno, S. (2013). The linked dual representation model of vocal perception and production. *Frontiers in Psychology*, *4*, 825. https://doi.org/10.3389/fpsyg.2013.00825

Hutchins, S., & Peretz, I. (2012). Amusics can imitate what they cannot discriminate. *Brain and Language*, *123*(3), 234–239. https://doi.org/10.1016/j.bandl.2012.09.011

James, R., Sigafoos, J., Green, V. A., Lancioni, G. E., O'Reilly, M. F., Lang, R., Davis, T., Carnett, A., Achmadi, D., Gevarter, C., & Marschik, P. B. (2015). Music therapy for individuals with autism Spectrum disorder: A systematic review. *Review Journal of Autism and Developmental Disorders*, *2*(1), 39–54. https://doi.org/10.1007/s40489-014-0035-4

Jiang, J., Liu, F., Wan, X., & Jiang, C. (2015). Perception of melodic contour and intonation in autism Spectrum disorder: Evidence from mandarin speakers. *Journal of Autism and Developmental Disorders*, *45*(7), 2067–2075. https://doi.org/10.1007/s10803-015-2370-4

Johnson, K. (2011). *Acoustic and auditory phonetics*. John Wiley & Sons.

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

Kent, R. D., & Forner, L. L. (1980). Speech segment durations in sentence recitations by children and adults. *Journal of Phonetics*, *8*(2), 157–168. https://doi.org/10.1016/S0095-4470(19)31460-3

Kessels, R. P. C., van Zandvoort, M. J. E., Postma, A., Kappelle, L. J., & de Haan, E. H. F. (2000). The Corsi block-tapping task: Standardization and normative data. *Applied Neuropsychology*, *7*(4), 252–258. https://doi.org/10.1207/S15324826AN0704_8

Koelsch, S. (2011). Toward a neural basis of music perception – A review and updated model. *Frontier in Psychology*, *2*, 110. https://doi.org/10.3389/fpsyg.2011.00110

Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences*, *9*(12), 578–584. https://doi.org/10.1016/j.tics.2005.10.001

Koenig, L. L., Lucero, J. C., & Perlman, E. (2008). Speech production variability in fricatives of children and adults: Results of functional data analysis. *The Journal of the Acoustical Society of America*, *124*(5), 3158–3170. https://doi.org/10.1121/1.2981639

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, *97*(22), 11850–11857. https://doi.org/10.1073/pnas.97.22.11850

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Developmental Science*, *10*(1), 110–120. https://doi.org/10.1111/j.1467-7687.2007.00572.x

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, *100*(4), 2425–2438. https://doi.org/10.1121/1.417951

Kunert, R., & Slevc, L. R. (2015). A commentary on: "Neural overlap in processing music and speech.". *Frontiers in Human Neuroscience*, *9*, 330. https://doi.org/10.3389/fnhum.2015.00330

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 13. https://doi.org/10.18637/jss.v082.i13

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). *Emmeans: Estimated marginal means, aka least-squares means* (1(2)) [Computer software]. R package version. https://github.com/rvlenth/emmeans

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics*, *56*(4), 414–423. https://doi.org/10.3758/BF03206733

Liu, F., Chan, A. H. D., Ciocca, V., Roquet, C., Peretz, I., & Wong, P. C. M. (2016). Pitch perception and production in congenital amusia: Evidence from Cantonese speakers. *The Journal of the Acoustical Society of America*, *140*(1), 563–575. https://doi.org/10.1121/1.4955182

Liu, F., Jiang, C., Pfordresher, P. Q., Mantell, J. T., Xu, Y., Yang, Y., & Stewart, L. (2013). Individuals with congenital amusia imitate pitches more accurately in singing than in speaking: Implications for music and language processing. *Attention, Perception, & Psychophysics*, *75*(8), 1783–1798. https://doi.org/10.3758/s13414-013-0506-1

Liu, F., Jiang, C., Wang, B., Xu, Y., & Patel, A. D. (2015). A music perception disorder (congenital amusia) influences speech comprehension. *Neuropsychologia*, *66*, 111–118. https://doi.org/10.1016/j.neuropsychologia.2014.11.001

Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: Discrimination, identification and imitation. *Brain*, *133*(6), 1682–1693. https://doi.org/10.1093/brain/awq089

Loeb, D. F., & Allen, G. D. (1993). Preschoolers' imitation of intonation contours. *Journal of Speech, Language, and Hearing Research*, *36*(1), 4–13. https://doi.org/10.1044/jshr.3601.04

Loui, P. (2015). A dual-stream neuroanatomy of singing. *Music Perception*, *32*(3), 232–241. https://doi.org/10.1525/mp.2015.32.3.232

Mantell, J. T., & Pfordresher, P. Q. (2013). Vocal imitation of song and speech. *Cognition*, *127*(2), 177–202. https://doi.org/10.1016/j.cognition.2012.12.008

Martin, N. A., & Brownell, R. (2011). *Receptive One-Word Picture Vocabulary Test, 4th Edition (ROWPVT-4)*. Academic Therapy Publications.

Masur, E. F. (2006). Vocal and action imitation by infants and toddlers during dyadic interactions: Development, causes, and consequences. In *Imitation and the social mind: Autism and typical development* (pp. 27–47). The Guilford Press.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, *134*(4), 477–500.

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language & Communication Disorders*, *38*(4), 325–350. https://doi.org/10.1080/1368282031000154204

McCann, J., Peppé, S., Gibbon, F. E., O'Hare, A., & Rutherford, M. (2007). Prosody and its relationship to language in school-aged children with high-functioning autism. *International Journal of Language & Communication Disorders*, *42*(6), 682–702. https://doi.org/10.1080/13682820601170102

McCreery, R. W., & Stelmachowicz, P. G. (2011). Audibility-based predictions of speech recognition for children and adults with normal hearing. *The Journal of the Acoustical Society of America*, *130*(6), 4070–4081. https://doi.org/10.1121/1.3658476

Mcdonald, J. H. (2014). *Handbook of biological statistics* (3rd ed.). Sparky House Publishing.

McEwen, F., Happé, F., Bolton, P., Rijsdijk, F., Ronald, A., Dworzynski, K., & Plomin, R. (2007). Origins of individual differences in imitation: Links with language, pretend play, and socially insightful behavior in two-year-old twins. *Child Development*, *78*(2), 474–492. https://doi.org/10.1111/j.1467-8624.2007.01010.x

Meltzoff, A. N. (2017). Elements of a comprehensive theory of infant imitation. *Behavioral and Brain Sciences*, *40*, e396. https://doi.org/10.1017/S0140525X1600193X

Meltzoff, A. N., & Keith Moore, M. (1994). Imitation, memory, and the representation of persons. *Infant Behavior and Development*, *17*(1), 83–99. https://doi.org/10.1016/0163-6383(94)90024-8

Meltzoff, A. N., & Williamson, R. A. (2013). Imitation: Social, cognitive, and theoretical perspectives. In *The Oxford handbook of developmental psychology (Vol. 1): Body and mind* (pp. 651–682). Oxford University Press.

Mertens, P. (2004). The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model. *Proceedings of Speech Prosody*, *2004*, 23–26.

Molnar-Szakacs, I., & Heaton, P. (2012). Music: A unique window into the world of autism. *Annals of the New York Academy of Sciences*, *1252*, 318–324. https://doi.org/10.1111/j.1749-6632.2012.06465.x

Mottron, L., Peretz, I., Belleville, S., & Rouleau, N. (1999). Absolute pitch in autism: A case study. *Neurocase*, *5*(6), 485–501. https://doi.org/10.1080/13554799908402744

Munson, B. (2004). Variability in /s/ production in children and adults. *Journal of Speech, Language, and Hearing Research*, *47*(1), 58–69. https://doi.org/10.1044/1092-4388(2004/006)

Nguyen, N., & Delvaux, V. (2015). Role of imitation in the emergence of phonological systems. *Journal of Phonetics*, *53*, 46–54. https://doi.org/10.1016/j.wocn.2015.08.004

Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron*, *88*(6), 1281–1296. https://doi.org/10.1016/j.neuron.2015.11.035

Ouimet, T., Foster, N. E. V., Tryfon, A., & Hyde, K. L. (2012). Auditory-musical processing in autism spectrum disorders: A review of behavioral and brain imaging studies. *Annals of the New York Academy of Sciences*, *1252*, 325–331. https://doi.org/10.1111/j.1749-6632.2012.06453.x

Over, H., & Carpenter, M. (2013). The social side of imitation. *Child Development Perspectives*, *7*(1), 6–11. https://doi.org/10.1111/cdep.12006

Over, H., & Gattis, M. (2010). Verbal imitation is based on intention understanding. *Cognitive Development*, *25*(1), 46–55. https://doi.org/10.1016/j.cogdev.2009.06.004

Pagliarini, S., Leblois, A., & Hinaut, X. (2020). Vocal imitation in sensorimotor learning models: A comparative review. *IEEE Transactions on Cognitive and Developmental Systems*, *13*, 1. https://doi.org/10.1109/TCDS.2020.3041179

Patel, A. D. (2008). *Music, language, and the brain*. Oxford University Press.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*, 142. https://doi.org/10.3389/fpsyg.2011.00142

Patel, A. D. (2012). The OPERA hypothesis: Assumptions and clarifications. *Annals of the New York Academy of Sciences*, *1252*, 124–128. https://doi.org/10.1111/j.1749-6632.2011.06426.x

Patel, A. D. (2013). Sharing and nonsharing of brain resources for language and music. In M. Arbib (Ed.), *Language, music, and the brain: A mysterious relationship* (pp. 329–355). MIT Press.

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, *87*(1), B35–B45. https://doi.org/10.1016/S0010-0277(02)00187-7

Patel, A. D., Iversen, J. R., & Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *The Journal of the Acoustical Society of America*, *119*(5), 3034–3047. https://doi.org/10.1121/1.2179657

Patel, A. D., Xu, Y., & Wang, B. (2010). The role of F0 variation in the intelligibility of Mandarin sentences. *Proceedings of Speech Prosody 2010*, 1–4.

Paul, R., Bianchi, N., Augustyn, A., Klin, A., & Volkmar, F. (2008). Production of syllable stress in speakers with autism Spectrum disorders. *Research in Autism Spectrum Disorders*, *2*(1), 110–124. https://doi.org/10.1016/j.rasd.2007.04.001

Peppé, S., Cleland, J., Gibbon, F., O'Hare, A., & Castilla, P. M. (2011). Expressive prosody in children with autism spectrum conditions. *Journal of Neurolinguistics*, *24*(1), 41–53. https://doi.org/10.1016/j.jneuroling.2010.07.005

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and expressive prosodic ability in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research*, *50*(4), 1015–1028. https://doi.org/10.1044/1092-4388(2007/071)

Peretz, I. (2009). Music, language and modularity framed in action. *Psychologica Belgica*, *49*(2-3), 157. https://doi.org/10.5334/pb-49-2-3-157

Peretz, I., Champod, A. S., & Hyde, K. L. (2003). Varieties of musical disorders. The Montreal battery of evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*, 58–75. https://doi.org/10.1196/annals.1284.006.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7), 688–691. https://doi.org/10.1038/nn1083

Peretz, I., Gosselin, N., Nan, Y., Caron-Caplette, E., Trehub, S. E., & Béland, R. (2013). A novel tool for evaluating children's musical abilities across age and culture. *Frontiers in Systems Neuroscience*, *7*, 30. https://doi.org/10.3389/fnsys.2013.00030

Peretz, I., Vuvan, D., Lagrois, M.-É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140090. https://doi.org/10.1098/rstb.2014.0090

Peretz, I., & Zatorre, R. (2005). Brain organization for music processing. *Annual Review of Psychology*, *56*, 89–114.

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness.". *Music Perception*, *25*(2), 95–115. https://doi.org/10.1525/mp.2007.25.2.95

Prince, J. B., & Pfordresher, P. Q. (2012). The role of pitch and temporal diversity in the perception and production of musical sequences. *Acta Psychologica*, *141*(2), 184–198. https://doi.org/10.1016/j.actpsy.2012.07.013

Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven manual: Section 3. Standard progressive matrices*. Oxford Psychologists Press.

Reifinger, J. L. (2006). Skill development in rhythm perception and performance: A review of literature. *Update: Applications of Research in Music Education*, *25*(1), 15–27. https://doi.org/10.1177/87551233060250010103

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2009). A cognitive approach to the development of early language. *Child Development*, *80*(1), 134–150. https://doi.org/10.1111/j.1467-8624.2008.01250.x

RStudio Team. (2020). RStudio: Integrated development for R. RStudio, PBC, Boston, MA. http://www.rstudio.com/

Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology*, *37*(1), 74.

Sammler, D., Koelsch, S., Ball, T., Brandt, A., Elger, C. E., Friederici, A. D., Grigutsch, M., Huppertz, H.-J., Knösche, T. R., Wellmer, J., Widman, G., & Schulze-Bonhage, A. (2009). Overlap of musical and linguistic syntax processing: Intracranial ERP evidence. *Annals of the New York Academy of Sciences*, *1169*(1), 494–498. https://doi.org/10.1111/j.1749-6632.2009.04792.x

Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science*, *14*(3), 262–266. https://doi.org/10.1111/1467-9280.03432

Schellenberg, E. G., & Trehub, S. E. (2008). Is there an Asian advantage for pitch memory? *Music Perception*, *25*(3), 241–252. https://doi.org/10.1525/mp.2008.25.3.241

Schleuter, S. L., & Schleuter, L. J. (1985). The relationship of grade level and sex differences to certain rhythmic responses of primary grade children. *Journal of Research in Music Education*, *33*(1), 23–29. https://doi.org/10.2307/3344755

Sergeant, D., & Roche, S. (1973). Perceptual shifts in the auditory information processing of Young children. *Psychology of Music*, *1*(2), 39–48.

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, 17(8), 675–681. https://doi.org/10.1111/j.1467-9280.2006.01765.x

Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review*, 16(2), 374–381. https://doi.org/10.3758/16.2.374

Smith, B. L. (1978). Temporal aspects of English speech production: A developmental perspective. *Journal of Phonetics*, 6(1), 37–67. https://doi.org/10.1016/S0095-4470(19)31084-8

Smith, B. L., Kenney, M. K., & Hussain, S. (1996). A longitudinal investigation of duration and temporal variability in children's speech production. *The Journal of the Acoustical Society of America*, 99(4), 2344–2349. https://doi.org/10.1121/1.415421

Snow, D. (1998). Children's imitations of intonation contours. *Journal of Speech, Language, and Hearing Research*, 41(3), 576–587. https://doi.org/10.1044/jslhr.4103.576

Stanutz, S., Wapnick, J., & Burack, J. A. (2014). Pitch discrimination and melodic memory in children with autism spectrum disorders. *Autism: The International Journal of Research and Practice*, 18(2), 137–147. https://doi.org/10.1177/1362361312462905

Stelmachowicz, P. G., Hoover, B. M., Lewis, D. E., Kortekaas Reinier, W. L., & Pittman, A. L. (2000). The relation between stimulus context, speech audibility, and perception for Normal-hearing and hearing-impaired children. *Journal of Speech, Language, and Hearing Research*, 43(4), 902–914. https://doi.org/10.1044/jslhr.4304.902

Stone, W. L., & Yoder, P. J. (2001). Predicting spoken language level in children with autism Spectrum disorders. *Autism*, 5(4), 341–361. https://doi.org/10.1177/1362361301005004002

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin*, 113(2), 345.

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, 34(3), 434–464. https://doi.org/10.1111/j.1551-6709.2009.01077.x

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science*, 21(10), 1532–1540. https://doi.org/10.1177/0956797610384142

Tremblay-Champoux, A., Dalla Bella, S., Phillips-Silver, J., Lebrun, M.-A., & Peretz, I. (2010). Singing proficiency in congenital amusia: Imitation helps. *Cognitive Neuropsychology*, 27(6), 463–476. https://doi.org/10.1080/02643294.2011.567258

Turan, F., & Okcun Akcamus, M. C. (2012). An investigation of the development of imitation skill in children with autism and its relations with receptive- expressive language development and the development of symbolic play. *Turkish Journal of Psychiatry*, 24, 111. https://doi.org/10.5080/u7007

Van Santen, J. P. H., Prud'hommeaux, E. T., Black, L. M., & Mitchell, M. (2010). Computational prosodic markers for autism. *Autism: The International Journal of Research and Practice*, 14(3), 215–236. https://doi.org/10.1177/1362361309363281

Vivanti, G., & Hamilton, A. (2014). Imitation in autism spectrum disorders. In *Handbook of autism and pervasive developmental disorders* (4th ed.). American Cancer Society. https://doi.org/10.1002/9781118911389.hautc12

Vivanti, G., Trembath, D., & Dissanayake, C. (2014). Mechanisms of imitation impairment in autism Spectrum disorder. *Journal of Abnormal Child Psychology*, 42(8), 1395–1405. https://doi.org/10.1007/s10802-014-9874-9

Wechsler, D. (2003). *Wechsler intelligence scale for children* (4th ed.). The Psychological Corporation.

Williams, J. H. G., Whiten, A., Suddendorf, T., & Perrett, D. I. (2001). Imitation, mirror neurons and autism. *Neuroscience & Biobehavioral Reviews*, 25(4), 287–295. https://doi.org/10.1016/S0149-7634(01)00014-8

Wisniewski, M., Mantell, J. T., & Pfordresher, P. (2013). Transfer effects in the vocal imitation of speech and song. *Psychomusicology: Music, Mind, and Brain*, 23, 82–99. https://doi.org/10.1037/A0033299

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, 46(3–4), 220–251. https://doi.org/10.1016/j.specom.2005.02.014

Xu, Y. (2013). ProsodyPro—A tool for large-scale systematic prosody analysis. Paper presented at: Proceedings of Tools and Resources for the Analysis of Speech Prosody, pp. 7–10.

Xu, Y., & Prom-on, S. (2014). Toward invariant functional representations of variable surface fundamental frequency contours: Synthesizing speech melody via model-based stochastic learning. *Speech Communication*, 57, 181–208. https://doi.org/10.1016/j.specom.2013.09.013

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from mandarin Chinese. *Speech Communication*, 33(4), 319–337.

Yang, J., & Hofmann, J. (2016). Action observation and imitation in autism spectrum disorders: An ALE meta-analysis of fMRI studies. *Brain Imaging and Behavior*, 10(4), 960–969. https://doi.org/10.1007/s11682-015-9456-7

Young, G. S., Rogers, S. J., Hutman, T., Rozga, A., Sigman, M., & Ozonoff, S. (2011). Imitation from 12 to 24 months in autism and typical development: A longitudinal Rasch analysis. *Developmental Psychology*, 47(6), 1565–1578. https://doi.org/10.1037/a0025418

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 1087–1104. https://doi.org/10.1098/rstb.2007.2161

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.