Individuals with congenital amusia imitate pitches more accurately in singing than in speaking: Implications for music and language processing

Fang Liu, Cunmei Jiang, Peter Q. Pfordresher, James T. Mantell, Yi Xu, Yufang Yang & Lauren Stewart

Springer

Springer

Individuals with congenital amusia imitate pitches more accurately in singing than in speaking: Implications for music and language processing

Fang Liu · Cunmei Jiang · Peter Q. Pfordresher ·
James T. Mantell · Yi Xu · Yufang Yang · Lauren Stewart

Abstract In this study, we investigated the impact of congenital amusia, a disorder of musical processing, on speech and song imitation in speakers of a tone language, Mandarin. A group of 13 Mandarin-speaking individuals with congenital amusia and 13 matched controls were recorded while imitating a set of speech and two sets of song stimuli with varying pitch and rhythm patterns. The results indicated that individuals with congenital amusia were worse than controls in both speech and song imitation, in terms of both pitch matching (absolute and relative) and rhythm matching (relative time and number of time errors). Like the controls, individuals with congenital amusia achieved better absolute and relative pitch matching and made fewer pitch interval and contour errors in song than in speech imitation. These findings point toward domain-general pitch (and time) production deficits in congenital amusia, suggesting the presence of shared pitch production mechanisms but distinct requirements for pitch-matching accuracy in language and music processing.

Keywords Modularity of perception · Music cognition · Sound recognition · Perception and action · Speech production · Temporal processing

F. Liu (✉)
Department of Linguistics and Modern Languages, Chinese University of Hong Kong, Shatin, New Territories, Hong Kong, China
e-mail: liufang@uchicago.edu

C. Jiang
Music College, Shanghai Normal University, Shanghai 200234, China

P. Q. Pfordresher · J. T. Mantell
Department of Psychology, University at Buffalo, State University of New York, Buffalo, NY, USA

Y. Xu
Department of Speech, Hearing and Phonetic Sciences, University College London, London WC1N 1PF, UK

Y. Yang
Institute of Psychology, Chinese Academy of Sciences, Beijing 100864, China

L. Stewart
Department of Psychology, Goldsmiths, University of London, London SE14 6NW, UK

Congenital amusia is a disorder primarily of pitch perception and production that has a profound impact on musical processing, but only minor effects on speech processing (Ayotte, Peretz, & Hyde, 2002; Liu, Patel, Fourcin, & Stewart, 2010; Patel, 2008; Peretz, Ayotte, Zatorre, Mehler, Ahad, Penhune, & Jutras, 2002; Thompson, Marin, & Stewart, 2012). Recent research has suggested that the apparent domain specificity of congenital amusia can be explained partly by the following observations: First, individuals with congenital amusia only demonstrate reduced performance in speech processing when the pitch contrasts involved are relatively small (Hutchins, Gosselin, & Peretz, 2010; Jiang, Hamm, Lim, Kirk, & Yang, 2010; Liu et al., 2010; Liu, Jiang, Thompson, Xu, Yang, & Stewart, 2012; Nan, Sun, & Peretz, 2010; Patel, Wong, Foxton, Lochy, & Peretz, 2008); second, linguistic contexts and acoustic features other than pitch (e.g., duration, intensity) may provide additional cues for speech communication (Liu, Jiang, et al., 2012; Patel, Foxton, & Griffiths, 2005); and finally, the pitch-processing deficits in individuals with congenital amusia are more pronounced with discrete musical pitches than with gliding pitches in speech (Foxton, Dean, Gee, Peretz, & Griffiths, 2004; Liu, Xu, Patel, Francart, & Jiang, 2012). However, evidence is missing with regard to how the different functions of language and music may impact the domain specificity of congenital amusia. Pitch patterns in speech do not need to match a specified standard, but instead merely need to convey contrastive functional information (Xu

2005). By contrast, musical pitch must conform to specific conventions that apply to individual pitches as well as pitch patterns. In other words, the "form" taken by pitch patterns acts as a means of communication in speech, but is the intended end product for music (Patel, 2008). Understanding how musical versus linguistic pitch processing in congenital amusia is affected by the nature of music and language is useful for formulating a model of pitch processing in music and language that takes into account how impairments compromise auditory-processing skills in either domain. Considering four theoretical perspectives, in the present study we examined the characteristics of pitch and rhythm processing in speech versus song imitation in individuals with congenital amusia who speak a tone language, Mandarin.

The relationship between music and language

Much recent research has pointed to shared mechanisms between music and speech processing for individuals of different language and musical backgrounds (Bidelman, Gandour, & Krishnan, 2011; Hutchins, Gosselin, & Peretz, 2010; Jiang et al. 2010, 2012; Liu, Jiang, et al., 2012; Liu et al., 2010; Liu, Xu, et al., 2012; Mantell & Pfordresher, 2013; Nan et al., 2010; Patel, 2008, 2011, 2012a, 2012b; Pfordresher & Brown, 2009; Tillmann, Burnham, Nguyen, Grimault, Gosselin & Peretz 2011a; Tillmann, Rusconi, Traube, Butterworth, Umiltà & Peretz 2011b). However, the case study of a Polish-speaking poor-pitch singer (without pitch perceptual problems) in Dalla Bella, Berkowska, and Sowinski (2011) demonstrated domain-specific performance on pitch imitation in speech (intact) versus song (impaired). Although studies on a larger sample of English-speaking poor-pitch singers suggest that this dissociation may not generalize to the broader population of poor-pitch singers (Mantell & Pfordresher, 2013), it is unclear whether individuals with congenital amusia (with pitch perceptual problems) would demonstrate music-specific pitch production deficits due to the apparent domain specificity of congenital amusia (severely impaired musical perception; relatively spared speech perception) or domain-general pitch production deficits are associated with this disorder.

In the present study, we examined pitch production in congenital amusia through speech and song imitation among speakers of the tone language Mandarin, given that the Mandarin tonal system provides an ideal platform to match speech and song stimuli closely. Mandarin has four lexical tones and a neutral tone (Chen & Xu, 2006; Duanmu, 2007). In order to make speech stimuli most comparable to song stimuli (Liu, Xu, et al., 2012), only phonologically discrete Mandarin tones were used in the speech materials (i.e., high, Tone 1; low, Tone 3; and mid, the neutral tone). In order to account for the possible effects of interval size and rhythm pattern commonly used in speech versus music (Dowling &

Harwood, 1986; Peretz & Hyde, 2003), we included two sets of song stimuli, one with pitch and rhythm patterns similar to those in speech (language-song hereafter), and the other closely resembling Western music (music-song hereafter). We predicted that individuals with congenital amusia would perform worse than controls on both speech and song imitation and that both groups would perform better on song imitation than speech imitation, due to the greater demand on pitch precision imposed by music than speech (Patel, 2008, 2011, 2012b) and because of the fact that when imitating speech materials, individuals tend to imitate the functional goal (e.g., statement/question in English) rather than the form of the utterances (Liu et al., 2010; Over & Gattis, 2010) unless instructed to focus on pitch (Mantell & Pfordresher, 2013).

Pitch/interval/contour processing in music and speech

Absolute pitch, interval (pitch distance between notes), and contour (pitch direction, up vs. down, between notes) play different roles in long-/short-term memory of melodies (Dowling & Bartlett, 1981; Dowling & Fujitani, 1971; Dowling & Harwood, 1986). Previous findings suggest that individuals with congenital amusia tended to produce pitches lower than the targets when imitating single pitches (Hutchins, Zarate, Zatorre, & Peretz, 2010), and they also showed pitch interval and contour errors in singing and pitch matching tasks (Dalla Bella, Giguère, & Peretz, 2009; Loui, Guenther, Mathys, & Schlaug, 2008; Wise, 2009). Although these individuals showed great difficulty recognizing, memorizing, and producing melodies without lyrics (Ayotte et al., 2002; Dalla Bella et al., 2009), it remains unclear which aspects of melodic processing underlie such difficulty: pitch, interval, and/or contour processing? The present study examined pitch/interval/contour processing in speech versus song imitation among Mandarin-speaking individuals with congenital amusia through detailed acoustic analyses. Given that imitation facilitates amusic singing (Tremblay-Champoux, Dalla Bella, Phillips-Silver, Lebrun, & Peretz, 2010) and automatic pitch processing is involved during imitation (Hutchins & Peretz, 2012a; Liu et al., 2010; Loui et al., 2008), we expected these individuals to perform at the normal level on contour processing, but to show reduced performance on pitch/interval processing in speech/song imitation.

Rhythm processing in music and speech

Previous studies have indicated that only around half of individuals with congenital amusia have impaired rhythm processing in music, in terms of their singing of a familiar song (Dalla Bella et al., 2009) and performance on the rhythm subtest of the Montreal Battery of Evaluation of Amusia (MBEA; Peretz, Champod, & Hyde, 2003), which consists

of six subtests on scale, contour, interval, rhythm, meter, and memory processing of musical melodies. Other findings suggest that rhythm deficits in individuals with congenital amusia could only be revealed when test materials were musical (vs. noise bursts; Dalla Bella & Peretz, 2003) and when pitch variations were also involved (Foxton, Nandy, & Griffiths, 2006). The present study explored whether individuals with congenital amusia would show rhythm processing deficits in song and/or speech imitation. Given that rhythmic patterns in speech often resemble those in music (Patel, 2008; Patel, Iversen, & Rosenberg, 2006), on the basis of the findings in Foxton et al. (2006), we expected individuals with congenital amusia to demonstrate domain-general rhythm processing difficulties in speech and song imitation.

The relationship between perception and production

Congenital amusia presents a mixed picture in the relationship between musical perception and production: An association was found in some cases and dissociation in others. For example, when investigating singing in congenital amusia, Dalla Bella et al. (2009) found that for some individuals with congenital amusia, singing performance can be predicted by sensitivity to pitch changes: The poorer the pitch change detection, the worse the singing. However, in exceptional cases, proficient singing was also associated with severe perceptual deficits, and very poor singing with only mild perceptual deficits. An "action–perception mismatch" in congenital amusia was also observed in Loui et al. (2008), in which individuals with congenital amusia were able to imitate the correct direction of a heard pitch interval (intact production), despite their inability to report its direction (impaired perception). Nevertheless, a larger cohort of individuals with congenital amusia demonstrated mixed results in this regard (Williamson, Liu, Peryer, Grierson, & Stewart, 2012). Evidence of perception–production dissociation has also been seen in the speech domain, in which English- or French-speaking individuals with congenital amusia showed better performance on imitation than on identification/discrimination of speech intonation (Hutchins & Peretz, 2012a; Liu et al., 2010), and Mandarin-speaking individuals with congenital amusia demonstrated spared lexical tone production but impaired identification/discrimination (Nan et al., 2010). No research has yet examined whether the pitch perception deficit in congenital amusia would have similar effects on speech and music production when the linguistic and musical materials are closely matched.

In the present investigation, we compared the speech/song imitation abilities of individuals with congenital amusia with their scores on the Montreal Battery of Evaluation of Amusia (MBEA; Peretz et al., 2003) and their psychophysical perceptual thresholds for pitch change detection and pitch direction discrimination (see Table 1; see Liu, Jiang, et al., 2012, and Liu et al., 2010, for detailed task descriptions). Given that the deficits seen in the singing of individuals with congenital amusia cannot be solely attributable to their low-level pitch perception deficits (Dalla Bella et al., 2009), and given the demonstrated unconscious pitch processing during imitation as compared with the identification/discrimination of the same pitch events (Hutchins & Peretz, 2012a; Liu et al., 2010; Loui et al., 2008), we predicted that speech/song imitation in congenital amusia would be accounted for by both psychophysical pitch thresholds and melodic perception abilities.

Method

Participants

A group of 13 Mandarin-speaking individuals with congenital amusia and 13 matched controls were recruited via advertisements in the bulletin board systems in Beijing, China. The Montreal Battery of Evaluation of Amusia (MBEA) was used for diagnosis of congenital amusia in these individuals (Henry & McAuley, 2010; Peretz et al., 2003), with those having a total score of 65 or under (out of 90 trials) on the three pitch-based subtests (scale, contour, and interval) identified as having this musical disorder (Liu, Jiang, et al., 2012; Liu et al., 2010; Liu, Xu, et al., 2012; Thompson et al., 2012; Williamson & Stewart, 2010; Williamson et al., 2012). At the time of testing, all participants were enrolled as undergraduate or Master's students at universities in Beijing with Mandarin Chinese as their native language. In the questionnaire regarding their music, language, and medical/biological background, none of the participants reported having learning or memory problems with their studies, or any neurological/psychiatric disorders or speech/hearing difficulties. None had received formal extracurricular musical training. Table 1 shows the characteristics of these participants, as well as their scores on the MBEA subtests (Peretz et al., 2003) and their thresholds for pitch change detection and pitch direction discrimination (Liu, Jiang, et al., 2012; Liu et al., 2010). As can be seen, the two groups were comparable on all background measures, but individuals with congenital amusia performed worse on the MBEA and obtained higher pitch thresholds than did controls.

Stimuli

As in the procedure followed by Mantell and Pfordresher (2013), the stimuli here were constructed by creating sung tones with pitch patterns and text settings derived from naturally produced speech. Natural speech and song stimuli were recorded in two separate sessions in a soundproof booth at Goldsmiths, University of London, by a 27-year-old Mandarin-speaking female student (the *target speaker*, hereafter). Born

Table 1 Characteristics of the amusic ($n = 13$) and control ($n = 13$) groups

Characteristics	Amusic	Control	t	p
Sex	8 F, 5 M	9 F, 4 M	N/A	N/A
Handedness	2 L, 11R	0 L, 13R	N/A	N/A
Age	24.08 (2.93)	24.69 (1.84)	0.64	.53
Education (in years)	16.62 (2.53)	17.92 (0.95)	1.74	.09
MBEA scale	16.92 (3.33)	27.00 (1.91)	9.46	<.001
MBEA contour	19.31 (2.90)	26.85 (1.72)	8.06	<.001
MBEA interval	18.69 (2.98)	26.38 (1.61)	8.18	<.001
MBEA rhythm	21.92 (4.54)	27.08 (1.71)	3.83	<.001
MBEA meter	19.54 (4.03)	26.31 (2.32)	5.24	<.001
MBEA memory	21.54 (4.48)	28.23 (2.01)	4.91	<.001
MBEA pitch composite	54.92 (6.97)	80.23 (3.59)	11.64	<.001
Pitch change detection threshold	.19 (.07)	.14 (.04)	−2.15	.04
Pitch direction discrimination threshold	.16 (.06)	.11 (.05)	−2.29	.03

F = female; M = male; L = left; R = right; MBEA = Montreal Battery of Evaluation of Amusia (Peretz et al., 2003). The data are means, with SDs in parentheses; scores on the six MBEA subtests are numbers of correct responses out of 30; the pitch composite score is the sum of the scale, contour, and interval scores; pitch thresholds are in semitones; and t is the statistic of the Welch two-sample t test (two-tailed, $df = 24$). As is described by Liu and colleagues (Liu, Jiang, et al., 2012; Liu et al., 2010), the pitch change detection threshold task required participants to report which tone contained a pitch glide among three flat/gliding tones (two flat, and one containing either a rising or falling glide). The pitch direction discrimination threshold task required participants to report which gliding tone had a different direction than the other two (among three rising/falling gliding tones)

and raised in Beijing, the target speaker was an amateur singer/songwriter with 16 years of musical training. In total, 20 Mandarin sentences were used as the speech stimuli, each containing two to six syllables (Table 2).

Acoustic analyses of the speech stimuli were done using Praat (Boersma & Weenink, 2011), with the F_0 (fundamental frequency) and duration of each syllable being extracted using the ProsodyPro script (Xu, 2005–2012). Discrete complex-tone (F_0 plus seven odd harmonics) analogues of these stimuli were then created with a custom-written Praat script, using the technique described in Patel, Peretz, Tramo, and Labreque (1998), to serve as the targets (to be imitated by the target

speaker) of the language-song stimuli. These complex-tone sequences followed the rhythmic patterns of the speech stimuli but contained discrete pitches in the Western pitch class that were closest to the median F_0s of the speech syllables.

In the follow-up recording session for the song stimuli, the target speaker was presented with the auditory stimuli of the complex-tone analogues and the written scripts of the speech materials (in Chinese) and was instructed to reproduce/sing the pitches of the complex tones on the speech syllables while her voice was recorded. During recording, the target speaker made spontaneous adjustments to the rhythmic patterns of the complex-tone analogues in order to make the production more

Table 2 Stimuli used in the experiment

Number of syllables	Stimuli with an early focus	Stimuli with a late focus	Chinese pinyin	English translation
2	黑车?	黑车?	Hei1 che1?	Black car?
2	青天?	青天?	Qing1 tian1?	Blue sky?
3	她的包?	她的包?	Ta1 de0 bao1?	Her bag?
3	三颗星?	三颗星?	San1 ke1 xing1?	Three stars?
4	冬天的风?	冬天的风?	Dong1 tian1 de0 feng1?	The winter's wind?
4	写他书上?	写他书上?	Xie3 ta1 shu1 shang0?	Write on his book?
5	漆黑的天空?	漆黑的天空?	Qi1 hei1 de0 tian1 kong1?	Pitch-black sky?
5	小丁长高了?	小丁长高了?	Xiao3 Ding1 zhang3 gao1 le0?	Xiao Ding grew taller?
6	老郭的猫丢了?	老郭的猫丢了?	Lao3 Guo1 de0 mao1 diu1 le0?	Lao Guo's cat is lost?
6	小方天天加班?	小方天天加班?	Xiao3 Fang1 tian1 tian1 jia1 ban1?	Xiao Fang works overtime every day?

The focused words are in boldface. 0, 1, and 3 in the Chinese pinyin column stand for the neutral, high, and low tones in Mandarin, respectively

song-like. This set of recordings led to language-song stimuli, which featured musical pitches in the chromatic scale that were nearest to the median F_0s of the speech syllables. Consequently, these songs were atonal and contained larger pitch intervals than are common in Western music. In order to create tonal melodies that adhered more closely to the Western diatonic scale, the target speaker was requested to improvise and record another set of songs (music-song stimuli) that approximated the global melodic contours of the stimuli in speech and language song. Figure 1 shows the F_0 (in semitones, or *st*) contours of a set of speech/song examples (accessible at https://sites.google.com/site/fangliuproject/sound-examples), in which black dots represent the target productions, and green diamonds and red squares represent imitations by a control (C08) and an amusic (A08) participant, respectively. As can be seen, although the rhythms in the language-song stimuli mirrored but were proportionally slower than the speech rhythms, the music-song stimuli were mostly isochronous. It is worth noting that neither the language-song nor the music-song stimuli contained vibrato.

In order for participants of different genders to imitate target stimuli of the same gender, the three sets of recorded speech/song stimuli were synthesized into natural-sounding female (preserving the absolute pitches and formant frequencies of the original recordings) and male (changing the original pitches to one octave lower and shifting the frequencies of the original formants by .78 so as to achieve male voice characteristics) voices, using the "change gender" command in Praat. None of the participants commented that either the female or the male voice sounded unnatural, and no significant differences were found in imitation performance between the participants of different genders for either the amusic or the control group. Therefore, the syntheses of the female/male target stimuli were unlikely to have caused any adverse effects on imitation performance.

Table 3 displays the acoustic characteristics of the three types of stimuli in the female target (the values in the male target were 12 semitones lower). Paired *t* tests (shown in

Table 3) indicated that the speech and language-song stimuli did not differ significantly in absolute pitch (measured as the median F_0 of each syllable rhyme, in semitones) and pitch interval (the absolute difference in median F_0s between two consecutive syllable rhymes, in semitones), whereas the music-song stimuli generally had higher absolute pitches and smaller pitch intervals than did the speech and language-song stimuli. Whereas the speech stimuli on average had the shortest syllable durations (length of each syllable rhyme, in milliseconds) and interonset intervals (interval between the onsets of two consecutive syllable rhymes, in milliseconds; *IOIs*, hereafter), the music-song stimuli featured the longest syllable durations and IOIs.

Procedure

The experiments were conducted in a quiet room at the Institute of Psychology, Chinese Academy of Sciences in Beijing, China. Ethical approval was granted by both Chinese Academy of Sciences and Goldsmiths, University of London. Written informed consent forms were obtained from all participants. In previous studies, English-speaking individuals with congenital amusia had shown normal laryngeal control (the contact phase regularity of vocal fold vibration) and pitch production (overall pitch range and pitch regularity) when reading a story and producing three sustained vowels (Liu et al., 2010), and lexical tone production of Mandarin-speaking individuals with congenital amusia achieved near-perfect recognition rates by native listeners (Nan et al., 2010). On the other hand, French- and English-speaking individuals with congenital amusia have been shown to have problems with singing (Ayotte et al., 2002; Dalla Bella et al., 2009; Tremblay-Champoux et al., 2010). On the basis of the expected level of difficulty of the tasks, the three imitation tasks were administered to the participants in the order of (1) speech imitation, (2) language-song imitation, and (3) music-song imitation, with the easiest task presented first. These tasks were separated by
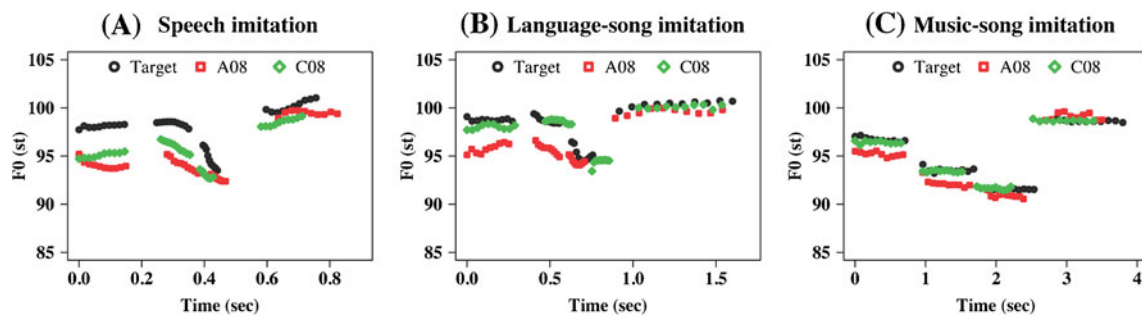

Fig. 1 F_0 contours [in semitones, or *st*; $st = 12 * \log_2(\text{Hz})$, $\text{Hz} = 2^{(st/12)}$, with 1 Hz as the reference frequency] of a set of speech/song examples produced by the target speaker (black dots) and participants C08 (green diamonds) and A08 (red squares): (**a**) speech, (**b**) language-song, and (**c**) music-song. The Mandarin sentence is 冬天的风? ("Dong1tian1 de0 feng1?"; "The wind in the winter?"), where 1 denotes the Mandarin high tone and 0 the neutral tone. Note that although the tones all had level pitches in the phonological/underlying forms, the surface representations may not be flat, because of various articulatory constraints (Xu & Wang, 2001)

Table 3 Acoustic characteristics of the three types of stimuli in the female target production

Characteristics		Speech (Sp)	Language-Song (S1)	Music-Song (S2)	Paired t test (Two-Sided)
Absolute pitch	Mean	95.50	95.53	96.44	Sp vs. S1: $t(79) = .31$, $p = .75$
	SD	3.83	4.02	3.52	Sp vs. S2: $t(79) = 2.58$, $p = .01$
	Range	89.02–101.77	88.74–102.12	89.46–103.17	S1 vs. S2: $t(79) = 2.32$, $p = .02$
Pitch interval	Mean	4.39	4.66	3.40	Sp vs. S1: $t(59) = 1.68$, $p = .10$
	SD	3.10	3.34	2.12	Sp vs. S2: $t(59) = 2.29$, $p = .03$
	Range	0.02–11.48	0.01–13.17	0.09–9.07	S1 vs. S2: $t(59) = 2.63$, $p = .01$
Syllable duration	Mean	153.13	390.02	836.13	Sp vs. S1: $t(79) = 11.87$, $p < .001$
	SD	51.92	215.16	142.12	Sp vs. S2: $t(79) = 52.66$, $p < .001$
	Range	50.84–258.05	89.71–1,016.25	630.85–1,196.52	S1 vs. S2: $t(79) = 28.22$, $p < .001$
IOI	Mean	241.75	414.69	943.69	Sp vs. S1: $t(59) = 14.84$, $p < .001$
	SD	63.50	133.53	47.63	Sp vs. S2: $t(59) = 77.69$, $p < .001$
	Range	111.18–389.85	186.56–818.51	844.65–1,037.78	S1 vs. S2: $t(59) = 31.52$, $p < .001$

Absolute pitch [in semitones, or st; $st = 12 * \log_2(\text{Hz})$, $\text{Hz} = 2^{(st/12)}$, with 1 Hz as the reference frequency] is the median F_0 of each syllable rhyme. Pitch interval (in semitones) is the pitch distance between two consecutive syllables (i.e., the difference in the median F_0s of the syllable rhymes; in absolute value). Syllable duration (in milliseconds) is the duration of each syllable rhyme. IOI (in milliseconds) is the interonset interval between two consecutive syllable rhymes

approximately 15-min gaps, during which the participants carried out the tone/intonation perception tasks as described in Liu, Jiang, et al. (2012). The fixed order of presentation of the speech/song imitation tasks was unlikely to have an impact on the present results for the following reasons. First, no consistent stimulus type effect was observed on imitation performance across the different measures reported in the Results section. Second, experiments that have been set up to look at change in performance across repetitions of the same trial have shown no effect of simple repetition (i.e., no improvement) on pitch matching (Hutchins & Peretz, 2012b) or speech/song imitation (Wisniewski, Mantell, & Pfordresher, in press).

The presentation of the target stimuli and the recording of the imitations were both done using Praat. Four practice trials, with items different from those in experimental trials, were given before the speech imitation task to familiarize the participants with the tasks and procedure. Speech/song stimuli were presented one at a time in pseudorandom order (the same across the three tasks) to the participants, who were required to imitate the pitch and time patterns of the utterances/melodies as closely as possible while their voice was recorded. The participants were encouraged to imitate each stimulus immediately following its presentation, although they could request that the experimenter (author F.L.) replay the stimulus if it was unclear, or repeat the imitation if there was disfluency (both of which rarely happened).

Data analysis

All acoustic analyses were conducted on the imitation data using the Praat script ProsodyPro (Xu, 2005–2012). Given

that musical beats in singing are usually synchronized with the vowel onsets, rather than the initial consonants, of the sung notes (Sundberg & Bauer-Huppmann, 2007), syllable/note duration was calculated as the length of the syllable rhyme, and the onset of syllable rhyme was defined as the syllable/note onset time. The median F_0s of the syllable rhymes were extracted in order to indicate pitch heights. Adapting the acoustic measurements in previous singing or pitch-matching studies (Dalla Bella et al., 2011; Dalla Bella, Giguère, & Peretz, 2007, 2009; Pfordresher & Brown, 2007; Pfordresher, Brown, Meier, Belyk, & Liotti, 2010; Ward & Burns, 1978), the following pitch and time variables were calculated so as to examine imitation accuracy of the participants.

The *absolute pitch deviation* (in semitones) of the imitated syllable/note from the target was the absolute difference in median F_0 between the two. Each participant had 20 values for each stimulus type, averaged across two to six syllables/notes in the 20 utterances/melodies. The bigger the value, the less accurate the imitation in terms of absolute pitch matching.

The *pitch interval deviation* (in semitones) was the absolute difference between the imitated pitch interval (difference in median F_0 between two consecutive syllables/notes; in absolute value) and the target pitch interval. Each participant had 20 values for each stimulus type, averaged across one to five intervals in the 20 utterances/melodies. The bigger the value, the less accurate the imitation in terms of relative pitch matching.

The *signed interval deviation* (in semitones) was the signed difference between each imitated interval (in absolute value) and the corresponding target interval (in absolute

value). Each participant had 60 values for each stimulus type, as the 20 utterances/melodies contained 60 intervals in total. Negative deviations indicate interval compressions, and positive deviations suggest interval expansions. This measure was used for examining patterns of interval compressions/expansions across the three stimulus types for the amusic and control groups, and not for measuring imitation accuracy per se.

The *number of contour errors* (out of the total 60 contours for each stimulus type) was the number of imitated pitch intervals that constituted different pitch directions (up, down, or level) than the corresponding target pitch intervals. Pitch direction was considered to be up or down if the difference in median F_0 between two consecutive syllables/notes was higher or lower by one semitone or more. If the difference in median F_0 between two consecutive syllables/notes was within one semitone, the two syllables/notes were considered to form a level/flat pitch direction.

The *number of pitch interval errors* (out of the total 60 intervals for each stimulus type) was the number of imitated pitch intervals that were larger or smaller than the corresponding target pitch intervals by one semitone. As in Dalla Bella et al. (2007, 2009, 2011), pitch interval errors were counted without considering whether pitch direction errors also occurred. Namely, imitated and target pitch intervals were compared using absolute values, which ignored co-occurring contour errors (if any).

The *duration difference* (in milliseconds) between the imitated syllable/note and the target was the absolute difference in rhyme length between the two. Each participant had 20 values for each stimulus type, averaged across two to six syllables/notes in the 20 utterances/melodies. The bigger the value, the less accurate the imitation in terms of absolute time matching.

The *IOI difference* (in milliseconds) was the absolute difference in interonset interval between two consecutive syllables/notes of the imitated and target productions. Each participant had 20 values for each stimulus type, averaged across one to five IOIs in the 20 utterances/melodies. The bigger the value, the less accurate the imitation in terms of relative time matching.

The *number of time errors* (out of the total 80 syllables/notes for each stimulus type) was the number of imitated syllables/notes that were at least 25 % longer or shorter than the corresponding target syllables/notes (Dalla Bella et al., 2007, 2009).

It is worth noting that—as compared with the absolute pitch deviation, pitch interval deviation, signed interval deviation, and duration/IOI differences—the numbers of contour/pitch interval/time errors are relative measures that are not necessarily affected by the characteristics of the target stimuli.

All statistical analyses were conducted using R (R Development Core Team, 2012). Two-way repeated measures analyses of variance (ANOVAs) were conducted to assess the main effects of group (amusic, control) and stimulus type (speech, language-song, music-song) on imitation accuracy, as well as their interaction. Generalized eta-squared was the measure of effect size calculated (Bakeman, 2005). The "glht" function ("Simultaneous Tests for General Linear Hypotheses") in the R package "multcomp" was used for post-hoc analyses, using Tukey contrasts for multiple comparisons of the means (Hothorn, Bretz, & Westfall, 2008). Kendall's rank correlation tau (τ; two-sided) was used for the correlation analyses. In order to examine whether, and to what extent, target pitches/intervals (durations/IOIs for time variables) affected each group's pitch/time matching accuracy, linear mixed-effects models were fit on the individual syllables/notes (which comprised 80 pitches/durations and 60 intervals/IOIs in total for each stimulus type) using the lme4 package for R (Baayen, Davidson, & Bates, 2008), with group (amusic, control) and target pitches/intervals/durations/IOIs as fixed effects, and individual participants and items as random effects. It is worth mentioning that it is inappropriate to use analysis of covariance to examine the effect of stimulus type with target pitches/intervals/durations/IOIs as covariates, since the two are not independent, and the assumption of the homogeneity of regression slopes for amusic and control groups was not met (Miller & Chapman, 2001).

Results

Absolute pitch deviation

Figure 2 shows box plots of the absolute pitch deviations (in semitones) of the two groups in the three imitation tasks, in which each participant had 20 values for each stimulus type, averaged across two to six syllables/notes in the utterances/melodies. The two-way repeated measures ANOVA revealed a significant main effect of group [$F(1, 24) = 8.03$, $p = .009$, $\eta_p^2 = .24$]. Individuals with congenital amusia produced significantly larger absolute pitch deviations than did controls across all three tasks (post-hoc comparisons between groups: speech, $z = 2.47$, $p = .01$; language-song, $z = 2.93$, $p = .003$; music-song, $z = 2.87$, $p = .004$). The main effect of stimulus type was also significant [$F(2, 48) = 26.28$, $p < .001$, $\eta_p^2 = .07$], with both groups showing better absolute pitch matching for language-song and music-song stimuli than for the speech stimuli (post-hoc analyses for individuals with congenital amusia: language-song vs. speech, $z = -3.78$, $p < .001$; music-song vs. speech, $z = -3.81$, $p < .001$; controls: language-song vs. speech, $z = -6.91$, $p < .001$; music-song vs. speech, $z = -10.44$, $p < .001$). Controls also showed better absolute pitch matching in music-song than in language-song imitation ($z = -3.54$, $p = .001$). No significant Group × Stimulus Type interaction was found for absolute pitch deviations.
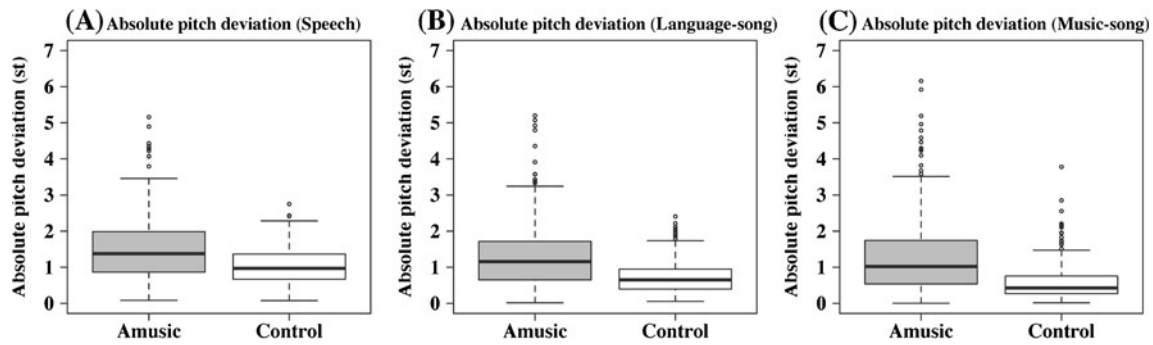
Fig. 2 Absolute pitch deviations (in semitones) of individuals with congenital amusia and controls in speech (**a**), language-song (**b**), and music-song (**c**) imitation. Each participant had 20 values for each stimulus type, averaged across two to six syllables/notes in the 20 utterances/melodies. These box plots show the extreme of the lower whisker, the lower hinge of the box, the median, the upper hinge, and the extreme of the upper whisker. The two hinges are the first and third quartiles, and the whiskers extend to the most extreme data points, which are no more than 1.5 times the interquartile range from the box. The data points that lie beyond the extremes of the whiskers are outliers, denoted by small open circles

Linear mixed-effects models of the absolute pitch deviations (with group and target pitches as fixed effects and individual participants and items as random effects) indicated that target pitch heights were negatively associated with the absolute pitch deviations in all three tasks (speech, $t = -6.00$, $p < .001$; language-song, $t = -2.72$, $p = .007$; music-song, $t = -2.45$, $p = .01$; the higher the target pitch, the smaller the absolute pitch deviation). A significant Group × Target Pitch interaction was found for speech imitation ($t = 2.13$, $p = .03$; the negative effect of target pitch on absolute pitch deviation was stronger for individuals with congenital amusia than controls), but not for language-song or music-song imitation.

Pitch interval deviation

Figure 3 shows box plots of the pitch interval deviations (in semitones) of the two groups in the three tasks. The two-way repeated measures ANOVA revealed significant main effects of group [$F(1, 24) = 12.76$, $p = .002$, $\eta_p^2 = .27$] and stimulus type [$F(2, 48) = 33.00$, $p < .001$, $\eta_p^2 = .30$], but no Group × Stimulus Type interaction on the pitch interval deviations. Post-hoc analyses suggested that individuals with congenital amusia had significantly larger pitch interval deviations than did the controls across all three tasks (speech, $z = 1.99$, $p = .047$; language-song, $z = 3.30$, $p < .001$; music-song, $z = 4.14$, $p < .001$). Both groups showed better relative pitch matching for music-song than speech or language-song stimuli (individuals with congenital amusia: music-song vs. speech, $z = -6.86$, $p < .001$; music-song vs. language-song, $z = -4.78$, $p < .001$; controls: music-song vs. speech, $z = -10.54$, $p < .001$; music-song vs. language-song, $z = -4.64$, $p < .001$). Although controls also showed better relative pitch matching in language-song than speech imitation ($z = -5.90$, $p < .001$), this difference was only marginally significant for individuals with congenital amusia ($z = -2.09$, $p = .09$).

Linear mixed-effects models of pitch interval deviations on group and target interval revealed a significant main effect of target interval (speech, $t = 15.10$, $p < .001$; language-song, $t = 10.88$, $p < .001$; music-song, $t = 8.87$, $p < .001$) and a significant Group × Target Interval interaction (speech, $t = -3.96$, $p < .001$; language-song, $t = -5.61$, $p < .001$; music-song, $t = -3.60$, $p < .001$) in all three tasks, as we observed a positive association between target interval and interval deviation (the larger the target interval, the greater the interval deviation), and this association was stronger in individuals with congenital amusia than in controls.

Signed interval deviation

Figure 4 shows the mean signed interval deviations (and standard errors) of the two groups against the target intervals (rounded to integers) for speech, language-song, and music-song imitation. Repeated measures ANOVAs were conducted on the signed interval deviations for each stimulus type, with Target Interval as the within-subjects factor and Group as the between-subjects factor. Given that theoretically only interval expansions are possible for the Target Interval 0 semitone (Dalla Bella et al., 2009), this interval size was not included in the ANOVA models. A significant group difference was found for language-song imitation [individuals with congenital amusia, mean $(SD) = -.97$ (1.76); controls $= -.46$ (1.25); $F(1, 24) = 5.29$, $p = .03$, $\eta_p^2 = .09$], but not for speech or music-song imitation. We found a significant main effect of target interval for all three types of stimuli [speech, $F(10, 240) = 26.07$, $p < .001$, $\eta_p^2 = .32$; language-song, $F(12, 288) = 48.15$, $p < .001$, $\eta_p^2 = .58$; music-song, $F(6, 144) = 70.67$, $p < .001$, $\eta_p^2 = .71$], with larger target intervals generally leading to greater interval compressions in imitation. The Target Interval × Group interaction was also significant for all three types of stimuli [speech, $F(10, 240) = 1.92$, $p = .04$, $\eta_p^2 = .03$; language-song, $F(12, 288) = 4.25$, $p < .001$, $\eta_p^2 =$
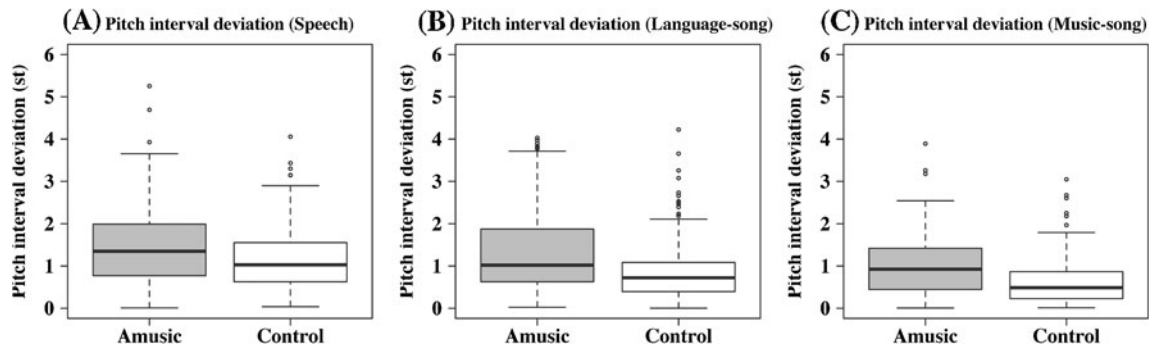
Fig. 3 Pitch interval deviations (in semitones) of individuals with congenital amusia and controls in speech (**a**), language-song (**b**), and music-song (**c**) imitation. Each participant had 20 values for each stimulus type, averaged across one to five intervals in the 20 utterances/melodies

.06; music-song, $F(6, 144) = 4.45$, $p < .001$, $\eta_p^2 = .11$], as the two groups demonstrated different degrees of interval expansions/compressions across the spectrum of the target interval sizes.

Number of pitch interval errors

Figure 5 shows the numbers of pitch interval errors made by the two groups out of the total of 60 pitch intervals in each stimulus type. The two-way repeated measures ANOVA revealed significant main effects of group [$F(1, 24) = 13.75$, $p = .001$, $\eta_p^2 = .28$] and stimulus type [$F(2, 48) = 26.63$, $p < .001$, $\eta_p^2 = .27$], and a significant Group × Stimulus Type interaction [$F(2, 48) = 4.11$, $p = .02$, $\eta_p^2 = .05$], on numbers of pitch interval errors. Post-hoc analyses suggested that the group effect was only significant for language-song ($z = -3.14$, $p = .002$) and music-song ($z = -4.52$, $p < .001$) imitation, but not for speech imitation. In addition, individuals with congenital amusia achieved fewer interval errors in music-song than in speech imitation ($z = -3.04$, $p = .007$), and controls' pitch interval errors were significantly different across the three tasks (speech > language-song, $z = 4.15$, $p < .001$; speech > music-song, $z = 7.51$, $p < .001$; language-song > music-song, $z = 3.36$, $p = .002$).

Generalized linear mixed-effects models revealed a positive association between number of pitch interval errors and target interval in all three tasks (speech, $z = 9.27$, $p < .001$; language-song, $z = 7.23$, $p < .001$; music-song, $z = 5.00$, $p < .001$), as both groups made more pitch interval errors when the target intervals were relatively large.

Number of contour errors

Figure 6 shows numbers of contour errors made by the two groups out of the total 60 contours in each stimulus type. No significant group effect or Group × Stimulus Type interaction was observed. The main effect of stimulus type was significant [$F(2, 48) = 54.98$, $p < .001$, $\eta_p^2 = .48$], as both groups made significantly fewer contour errors with music-song than with speech/language-song stimuli (individuals with congenital amusia: speech > music-song, $z = 6.17$, $p < .001$; language-song > music-song, $z = 4.92$, $p < .001$; controls: speech > music-song, $z = 8.03$, $p < .001$; language-song > music-song, $z = 6.98$, $p < .001$).

Generalized linear mixed-effects models revealed a negative association between numbers of contour errors and target intervals for both speech ($z = -6.47$, $p < .001$) and language-
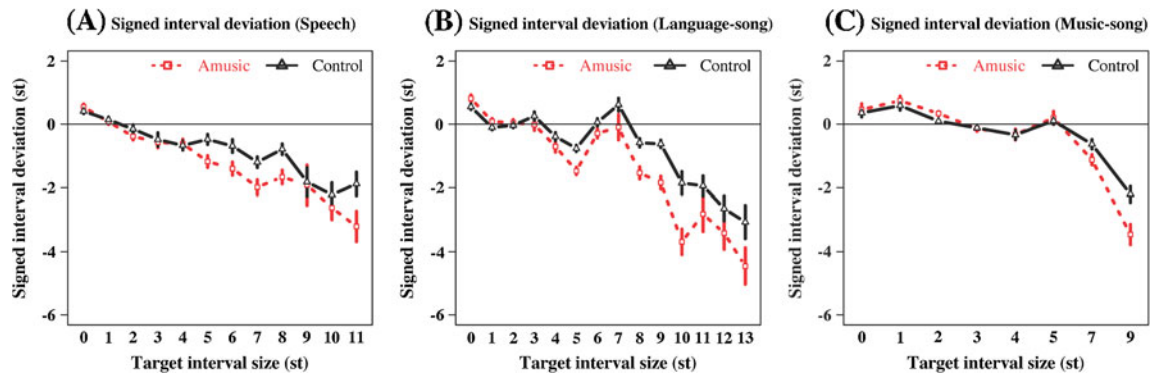

Fig. 4 Mean signed interval deviations (in semitones; with standard errors) of individuals with congenital amusia (red dashed lines) and controls (black straight lines) against target intervals (rounded to integers, in semitones) in speech (**a**), language-song (**b**), and music-song (**c**) imitation. Each participant had 60 values for each stimulus type, as the 20 utterances/melodies contained 60 intervals in total. Negative deviations indicate interval compressions, and positive deviations suggest interval expansions
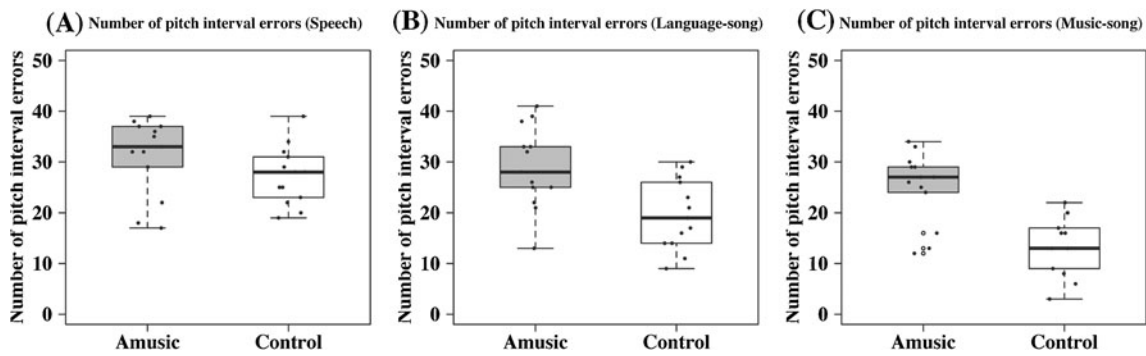
Fig. 5 Numbers of pitch interval errors (out of the total number of 60 pitch intervals in the speech/song stimuli) of individuals with congenital amusia and controls in speech (**a**), language-song (**b**), and music-song (**c**) imitation. Individual participants (13 in each group) are represented by black dots, with those at the same horizontal level having identical values, and those lying beyond the whiskers being outliers (which are further indicated by open circles along the midline)

song ($z = -6.03$, $p < .001$) imitation, but not for music-song imitation. That is, for both groups, smaller pitch intervals in the target production were more likely to lead to contour errors in the imitation than were larger pitch intervals in both speech and language-song imitation, but not in music-song imitation.

Duration difference

Figure 7 illustrates the duration differences (in milliseconds) between target and imitation by the two groups in the three tasks. The two-way repeated measures ANOVA revealed a significant main effect of stimulus type [$F(2, 48) = 137.07$, $p < .001$, $\eta_p^2 = .73$] and a significant Group × Stimulus Type interaction [$F(2, 48) = 3.79$, $p = .03$, $\eta_p^2 = .07$] on duration differences. The main effect of group was only marginally significant [$F(1, 24) = 4.22$, $p = .051$, $\eta_p^2 = .08$]. Post-hoc analyses suggested that individuals with congenital amusia showed significantly larger duration differences than did controls in speech ($z = 2.14$, $p = .03$) and music-song ($z = 2.15$, $p = .03$) imitation, but not in language-song imitation. Both groups showed the smallest duration differences in speech imitation and the biggest duration differences in music-song imitation

(individuals with congenital amusia: speech < language-song, $z = -11.61$, $p < .001$; speech < music-song, $z = -32.77$, $p < .001$; language-song < music-song, $z = -21.16$, $p < .001$; controls: speech < language-song, $z = -13.83$, $p < .001$; speech < music-song, $z = -30.22$, $p < .001$; language-song < music-song, $z = -16.39$, $p < .001$).

Linear mixed-effects models revealed a positive association between target duration and duration difference in all three tasks (speech, $t = 7.97$, $p < .001$; language-song, $t = 15.31$, $p < .001$; music-song, $t = 15.53$, $p < .001$), but this association was weaker for controls than for individuals with congenital amusia in music-song imitation (Group × Target Duration, $t = -3.06$, $p = .002$).

IOI difference

Figure 8 illustrates the IOI differences (in milliseconds) between target and imitation by the two groups in the three tasks. The two-way repeated measures ANOVA revealed significant main effects of group [$F(1, 24) = 13.34$, $p = .001$, $\eta_p^2 = .15$] and stimulus type [$F(2, 48) = 52.96$, $p < .001$, $\eta_p^2 = .60$] on the IOI differences, but no Group × Stimulus Type interaction. Post-hoc analyses suggested that individuals with congenital
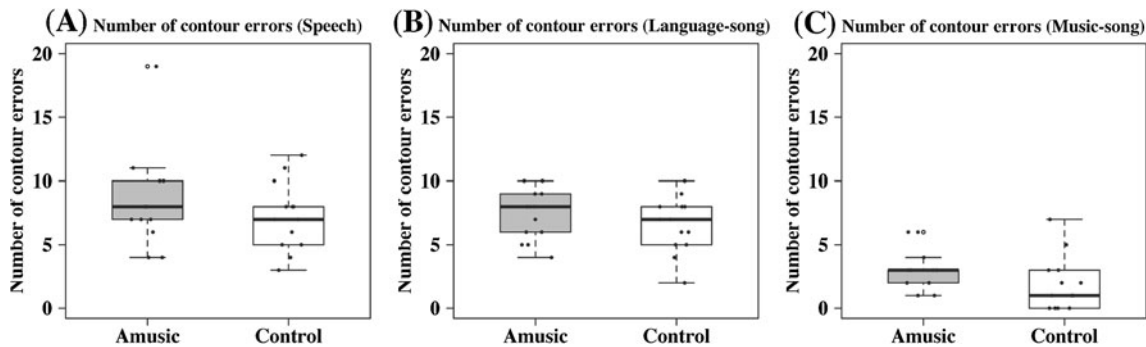

Fig. 6 Numbers of contour errors (out of the total number of 60 contours in the speech/song stimuli) of individuals with congenital amusia and controls in speech (**a**), language-song (**b**), and music-song (**c**) imitation. Individual participants (13 in each group) are represented by black dots, with those at the same horizontal level having identical values, and those lying beyond the whiskers being outliers (which are further indicated by open circles along the midline)
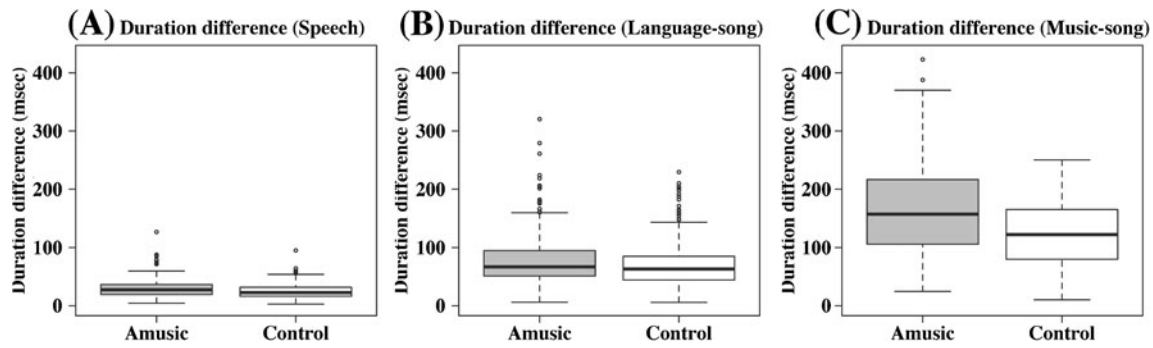
Fig. 7 Duration differences (in milliseconds) of individuals with congenital amusia and controls in speech (**a**), language-song (**b**), and music-song (**c**) imitation. Each participant had 20 values for each stimulus type, averaged across two to six syllables/notes in the 20 utterances/melodies

amusia showed significantly larger IOI differences than did controls in language-song ($z = 2.27$, $p = .02$) and music-song ($z = 2.43$, $p = .02$) imitation, but the difference was only marginally significant in speech imitation ($z = 1.78$, $p = .08$). Both groups showed the smallest IOI differences in speech imitation and the biggest IOI differences in music-song imitation (individuals with congenital amusia: speech < language-song, $z = -3.96$, $p < .001$; speech < music-song, $z = -17.32$, $p < .001$; language-song < music-song, $z = -13.36$, $p < .001$; controls: speech < language-song, $z = -4.21$, $p < .001$; speech < music-song, $z = -15.63$, $p < .001$; language-song < music-song, $z = -11.42$, $p < .001$).

Linear mixed-effects models revealed a positive association between target IOI and IOI difference in all three tasks (speech, $t = 2.76$, $p = .006$; language-song, $t = 6.49$, $p < .001$; music-song, $t = 6.52$, $p < .001$), and this association was weaker for controls than for individuals with congenital amusia in language-song imitation (Group × Target Duration, $t = -2.34$, $p = .02$).

Number of time errors

Figure 9 shows the numbers of time errors (out of 80) made by the two groups during speech/song imitation. The main effect of group was significant [$F(1, 24) = 8.65$, $p = .007$, $\eta_p^2 = .17$],

as controls made fewer time errors than did individuals with congenital amusia across the three tasks (speech, $z = -2.36$, $p = .02$; language-song, $z = -1.75$, $p = .08$; music-song, $z = -2.50$, $p = .01$). No significant effect of stimulus type or Group × Stimulus Type interaction was observed.

Generalized linear mixed-effects models revealed a negative association between number of time errors and target duration for speech imitation ($z = -2.94$, $p = .003$; the shorter the target duration, the greater the number of time errors), but a positive association between number of time errors and target duration for music-song imitation ($z = 6.70$, $p < .001$; the longer the target duration, the greater the number of time errors). No significant effect of target duration on number of time errors was observed for language-song imitation.

Correlations between imitation performance, MBEA scores, and pitch thresholds in individuals with congenital amusia

In order to investigate the relationship between production and perception in congenital amusia, correlation analyses were conducted between imitation performance and MBEA scores in individuals with congenital amusia (controls' data are omitted in the interest of space). Negative correlations suggest that better scores on the MBEA (i.e., number of correct responses
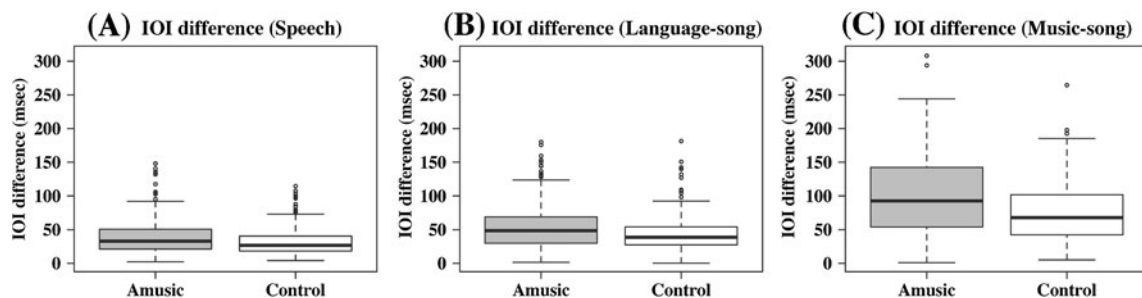

Fig. 8 Interonset interval (IOI) differences (in milliseconds) of individuals with congenital amusia and controls in speech (**a**), language-song (**b**), and music-song (**c**) imitation. Each participant had 20 values for each stimulus type, averaged across one to five IOIs in the 20 utterances/melodies
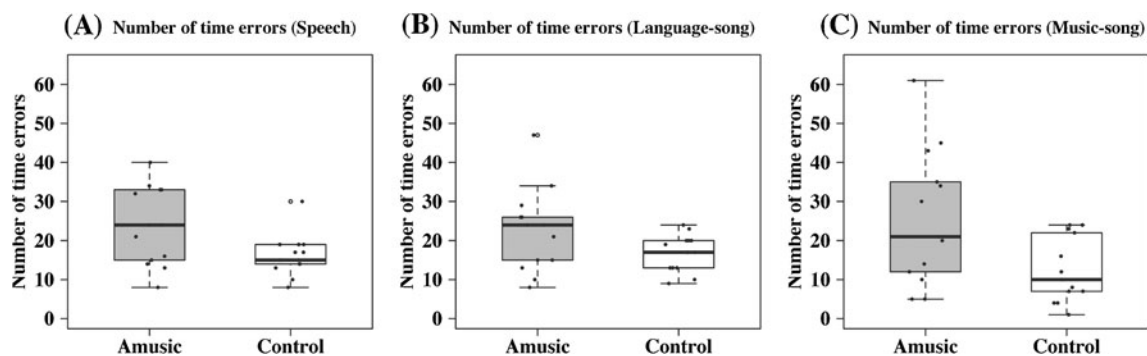
Fig. 9 Numbers of time errors (out of 80) of individuals with congenital amusia and controls in speech (a), language-song (b), and music-song (c) imitation. Individual participants (13 in each group) are represented by black dots, with those at the same horizontal level having identical values, and those lying beyond the whiskers being outliers (which are further indicated by open circles along the midline)

out of 30) were associated with better speech/song imitation performance (i.e., smaller values of pitch/time variables). First, MBEA scale scores were negatively correlated with pitch interval deviations ($\tau = -.45$, $p = .04$) and numbers of pitch interval ($\tau = -.59$, $p = .006$) and contour errors ($\tau = -.54$, $p = .02$) in speech imitation. Second, negative correlations were observed between MBEA interval scores and duration difference ($\tau = -.50$, $p = .02$) and number of time errors ($\tau = -.57$, $p = .01$) in speech imitation. Third, MBEA rhythm scores were negatively associated with duration difference ($\tau = -.48$, $p = .03$) and number of time errors ($\tau = -.49$, $p = .03$) in speech imitation. Fourth, negative associations were observed between MBEA meter scores and pitch interval deviations ($\tau = -.48$, $p = .02$) and number of pitch interval errors ($\tau = -.62$, $p = .004$) in speech imitation, and between MBEA meter scores and duration difference ($\tau = -.56$, $p = .008$), IOI difference ($\tau = -.48$, $p = .02$), and number of time errors ($\tau = -.50$, $p = .02$) in music-song imitation. Finally, MBEA memory scores were negatively correlated with absolute pitch deviations in speech ($\tau = -.43$, $p = .04$) and music-song ($\tau = -.43$, $p = .04$) imitation, with numbers of contour errors in music-song imitation ($\tau = -.58$, $p = .01$), and with IOI difference in language-song imitation ($\tau = -.51$, $p = .02$).

Correlation analyses between pitch thresholds and imitation performance in individuals with congenital amusia revealed a positive correlation between pitch direction discrimination thresholds and IOI difference in language-song imitation ($\tau = .61$, $p = .005$): The higher (worse) the thresholds, the worse the relative time matching.

Discussion

The relationship between music and language

In the present study, we investigated pitch and rhythm processing in speech versus song imitation among Mandarin-speaking individuals with and without congenital amusia in order to examine whether the functional differences between music and language would have any impact on pitch/rhythm processing in either domain. The finding of reduced speech and song imitation abilities in individuals with congenital amusia provides further evidence for shared mechanisms between music and language processing (Liu et al., 2010; Patel, 2008, 2012a). Given the important role that imitation plays in phonological development (Plaut & Kello, 1999), the observed speech imitation impairment in congenital amusia seems rather surprising, since it might potentially hinder the language (in this case, Mandarin) acquisition of these individuals. Nevertheless, individuals with congenital amusia rarely report language problems in everyday life (in English, Liu et al., 2010; in Mandarin, Jiang et al., 2010). The apparent paradox may be explained by the different natures of speech and music: Speech is function-driven, and music is form-driven (Patel, 2008). In particular, pitch patterns in speech are used for representing functional contrasts (e.g., lexical tone/stress, focus, sentence modality, etc.), and as such their execution only needs to satisfy contrastive adequacy (Xu, 2005). For music, musical understanding or communication relies on pitch accuracy and aesthetics, which are obvious aspects to be perfected in performances (Patel, 2008, 2011, 2012b). Indeed, research has demonstrated that the exact control of F_0 is "unnecessary" in speaking but "preferable" in singing (Natke, Donath, & Kalveram, 2003; Patel, 2012b). The present results are consistent with such claims, in that imitation of song was generally more accurate than imitation of speech with respect to pitch and time for individuals with or without congenital amusia (see also the similar results of Mantell & Pfordresher, 2013, for English-speaking individuals who do not have congenital amusia). That is, although both groups had mean absolute pitch deviations and pitch interval deviations above one semitone in speech imitation, controls' absolute/relative pitch deviations in song imitation were on average below one semitone, whereas those of individuals with congenital amusia were close to or above one semitone. Therefore, it seems that although neither group was very accurate in speech imitation, controls achieved increased accuracy for

song imitation, whereas individuals with congenital amusia were unable to do so, as evidenced by the significant Group × Stimulus Type interaction on number of pitch interval errors in speech/song imitation (the group effect was only significant for language-song and music-song imitation, but not for speech imitation).

Pitch/interval/contour processing in music and speech

The individuals with congenital amusia in the present study showed reduced performance relative to controls on both absolute (absolute pitch deviation for all three tasks) and relative (pitch interval deviation for all three tasks, number of pitch interval errors for language-song and music-song) pitch matching in speech/song imitation. Acoustic analyses revealed a positive association between target interval and interval deviation (the larger the target interval, the greater the interval deviation), and this association was stronger in individuals with congenital amusia than in controls. This indicates that these individuals were more likely than controls to compress large pitch intervals in both speech and song imitation (as evidenced by the results on signed interval deviations in the two groups; Fig. 4).

The present findings also suggest that the reduced performance on speech/song imitation in individuals with congenital amusia was due mostly to inaccurate pitch and interval processing, but not to inaccurate contour processing. This is consistent with the results of Loui et al. (2008), but not with Dalla Bella et al. (2009) and Wise (2009). Note that Pfordresher and Brown (2007) also found no differences between good- and poor-pitch singers with respect to contour errors. This discrepancy may be due to task (imitation vs. singing from memory) or stimulus (lyrics vs. tones) differences among these studies.

Rhythm processing in music and speech

When singing a familiar song from memory, individuals with congenital amusia have been shown to perform similarly to controls in terms of tempo, number of time errors, and rubato consistency, although they showed greater temporal variability than did controls (Dalla Bella et al., 2009). In the present study of speech/song imitation, the group effect was found to be significant for both IOI difference (significance for language-song and music-song, and marginal significance for speech) and number of time errors (significance for all three tasks). Furthermore, the significant Group × Stimulus Type interaction on duration difference (the main effect of group was marginally significant) suggests a significant group effect on duration differences in speech and music-song imitation, but not in language-song imitation. It is worth noting that IOI differences in the present study measured localized relative time matching between individual imitated IOIs and target IOIs,

which is equivalent to the measures of neither tempo (mean IOI of the quarter note) nor temporal variability (coefficient of variation of quarter-note IOIs) in Dalla Bella et al. (2009). The discrepancy between the present findings and those in Dalla Bella et al. (2009) concerning the number of time errors made by individuals with congenital amusia may result from the familiarity of the song materials used in the two studies. Namely, it may be that reduced time-matching abilities in individuals with congenital amusia are more likely to be revealed when singing or imitating unfamiliar speech/song materials (the present study) than when singing or imitating familiar ones (Dalla Bella et al., 2009).

It is worth noting that although both groups in the present study showed greater duration and IOI differences in song than in speech imitation (music-song > language-song > speech), this finding may not necessarily imply that absolute and relative time matching during speech imitation were superior to song imitation for both groups. Two of our results motivate this interpretation. First, no significant main effect of stimulus type was observed on number of time errors (a relative measure of time matching) for either group. Second, we found a positive association between the target duration/IOI and duration/IOI differences in all three tasks. Given that music-song stimuli contained the longest target durations/IOIs among the three stimulus types (Table 3), it is likely that the largest duration/IOI differences in music-song imitation were caused by the positive associations between target durations/IOIs and duration/IOI differences. This effect simply replicates the well-known association between target duration and timing variability in production (e.g., Wing & Kristofferson, 1973). Interestingly, the association between target duration and duration difference was weaker for controls than it was for individuals with congenital amusia within the music-song imitation condition. Similarly, the association between target IOI and IOI difference was weaker for controls than for individuals with congenital amusia within the language-song imitation condition. The fact that control participants were less strongly affected by target duration may partly explain their superior performance on time matching in speech/song imitation.

The relationship between perception and production

The extent to which singing and pitch-matching abilities can be predicted by pitch perception thresholds is a debated issue (Amir, Amir, & Kishon-Rabin, 2003; Bradshaw & McHenry, 2005; Dalla Bella et al., 2007, 2009; Hutchins & Peretz, 2011, 2012b; Nikjeh, Lister, & Frisch, 2009; Pfordresher & Brown, 2007). Upon observing the complex pitch production and perception associations and dissociations in congenital amusia, Dalla Bella et al. (2009) concluded that amusic singing could not be accounted for by a fine-grained pitch discrimination deficit alone. The results from the present

study further support this conclusion, as the speech/song imitation performance of individuals with congenital amusia was largely associated with their scores on the MBEA melodic perception tests, but rarely with their pitch change detection or pitch direction discrimination thresholds (see the details in the Results section). Acoustic analyses of the speech/song imitation data suggest that, although like controls, individuals with congenital amusia were more likely to make contour errors on smaller target intervals than on larger ones (especially in speech and language-song imitation), both groups made more pitch interval errors when the target intervals were relatively large (across the three tasks). Furthermore, individuals with congenital amusia showed a stronger positive association between target interval and pitch interval deviation than did controls: The larger the target interval, the greater the pitch interval deviation (mostly due to interval compression, as in Dalla Bella et al., 2009; Pfordresher & Brown, 2007). These findings indicate that the pitch imitation deficits of individuals with congenital amusia cannot be explained solely by their impaired abilities to discriminate fine-grained pitch changes, and thus the core deficit of amusia may go beyond low-level pitch-processing impairments (Patel, 2008; Patel et al., 2005).

Finally, although the present study did not measure cognitive abilities such as working memory capacity, the findings are unlikely to have resulted from the possible differences in cognitive ability between the two groups for the following reasons. First, individuals with congenital amusia demonstrated working memory capacities comparable to those of controls (Williamson & Stewart, 2010). Second, ranging from two to six syllables/notes, our speech/song stimuli were relatively short verbal sound sequences, for which individuals with congenital amusia show normal short-term memory (Tillmann, Schulze, & Foxton, 2009). However, future studies will be required in order to explore whether individual differences in cognitive abilities such as working memory are associated with musical and speech-processing abilities in congenital amusia.

Conclusion

The present study is the first to report the reduced speech and song imitation abilities of individuals with congenital amusia, despite the fact that these individuals are proficient speakers of a tone language, Mandarin. The domain-general pitch/time production deficits in congenital amusia provide a new line of evidence for the shared mechanisms in pitch/time processing between language and music. However, similar to controls, individuals with congenital amusia demonstrated better pitch matching in song than in speech imitation, suggesting that the apparent domain specificity of congenital amusia may partly be due to the different functions that music and language serve

in everyday life. That is, pitch patterns in speech are used for representing functional contrasts. For music, pitch accuracy is a crucial requirement for musical communication. Therefore, although individuals with congenital amusia are able to imitate pitches more accurately in singing than in speaking, the degree of precision is still not enough for music processing, but is already sufficient for speech processing.

References

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

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain, 125,* 238–251.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. doi:10.1016/j.jml.2007.12.005

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods, 37,* 379–384. doi:10.3758/BF03192707

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience, 23,* 425–434.

Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer (Version 5.2.19) [Computer program]. Retrieved from www.praat.org

Bradshaw, E., & McHenry, M. A. (2005). Pitch discrimination and pitch matching abilities of adults who sing inaccurately. *Journal of Voice, 19,* 431–439.

Chen, Y., & Xu, Y. (2006). Production of weak elements in speech: Evidence from $F(0)$ patterns of neutral tone in Standard Chinese. *Phonetica, 63,* 47–75.

Dalla Bella, S., Berkowska, M., & Sowiński, J. (2011). Disorders of pitch production in tone deafness. *Frontiers in Psychology, 2,* 164. doi:10.3389/fpsyg.2011.00164

Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2007). Singing proficiency in the general population. *Journal of the Acoustical Society of America, 121,* 1182–1189.

Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2009). Singing in congenital amusia. *Journal of the Acoustical Society of America, 126,* 414–424.

Dalla Bella, S., & Peretz, I. (2003). Congenital amusia interferes with the ability to synchronize with music. *Annals of the New York Academy of Sciences, 999,* 166–169.

Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval formation in long-term memory for melodies. *Psychomusicology, 1,* 30–49.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America, 49,* 524–531.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition*. San Diego, CA: Academic Press.

Duanmu, S. (2007). *The phonology of Standard Chinese* (2nd ed.). New York, NY: Oxford University Press.

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying "tone deafness. *Brain, 127,* 801–810.

Foxton, J. M., Nandy, R. K., & Griffiths, T. D. (2006). Rhythm deficits in "tone deafness. *Brain and Cognition, 62,* 24–29.

Henry, M. J., & McAuley, J. D. (2010). On the prevalence of congenital amusia. *Music Perception, 27,* 413–418.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50,* 346–363.

Hutchins, S., Gosselin, N., & Peretz, I. (2010a). Identification of changes along a continuum of speech intonation is impaired in congenital amusia. *Frontiers in Psychology, 1,* 236. doi:10.3389/fpsyg.2010.00236

Hutchins, S., & Peretz, I. (2011). Perception and action in singing. *Progress in Brain Research, 191,* 103–118.

Hutchins, S., & Peretz, I. (2012a). Amusics can imitate what they cannot discriminate. *Brain and Language, 123,* 234–239.

Hutchins, S., & Peretz, I. (2012b). A frog in your throat or in your ear? Searching for the causes of poor singing. *Journal of Experimental Psychology. General, 141,* 76–97.

Hutchins, S., Zarate, J. M., Zatorre, R. J., & Peretz, I. (2010b). An acoustical study of vocal pitch matching in congenital amusia. *Journal of the Acoustical Society of America, 127,* 504–512.

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2010). Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. *Neuropsychologia, 48,* 2630–2639. doi:10.1016/j.neuropsychologia.2010.05.009

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2012). Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics. *Memory & Cognition, 40,* 1109–1121. doi:10.3758/s13421-012-0208-2

Liu, F., Jiang, C., Thompson, W. F., Xu, Y., Yang, Y., & Stewart, L. (2012). The mechanism of speech processing in congenital amusia: Evidence from Mandarin speakers. *PLoS One, 7,* e30374. doi:10.1371/journal.pone.0030374

Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: Discrimination, identification, and imitation. *Brain, 133,* 1682–1693.

Liu, F., Xu, Y., Patel, A. D., Francart, T., & Jiang, C. (2012). Differential recognition of pitch patterns in discrete and gliding stimuli in congenital amusia: Evidence from Mandarin speakers. *Brain and Cognition, 79,* 209–215.

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action–perception mismatch in tone-deafness. *Current Biology, 18,* R331–R332.

Mantell, J. T., & Pfordresher, P. Q. (2013). Vocal imitation of song and speech. *Cognition, 127,* 177–202.

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110,* 40–48.

Nan, Y., Sun, Y., & Peretz, I. (2010). Congenital amusia in speakers of a tone language: Association with lexical tone agnosia. *Brain, 133,* 2635–2642.

Natke, U., Donath, T. M., & Kalveram, K. T. (2003). Control of voice fundamental frequency in speaking versus singing. *Journal of the Acoustical Society of America, 113,* 1587–1593.

Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2009). The relationship between pitch discrimination and vocal production: Comparison of vocal and instrumental musicians. *Journal of the Acoustical Society of America, 125,* 328–338.

Over, H., & Gattis, M. (2010). Verbal imitation is based on intention understanding. *Cognitive Development, 25,* 46–55.

Patel, A. D. (2008). *Music, language, and the brain*. New York, NY: 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. doi:10.3389/fpsyg.2011.00142

Patel, A. D. (2012a). Language, music, and the brain, a resource-sharing framework. In P. Rebuschat, M. Rohrmeier, J. Hawkins, & I. Cross (Eds.), *Language and music as cognitive systems* (pp. 204–223). Oxford, UK: Oxford University Press.

Patel, A. D. (2012b). The OPERA hypothesis, assumptions and clarifications. *Annals of the New York Academy of Sciences, 1252,* 124–128.

Patel, A. D., Foxton, J. M., & Griffiths, T. D. (2005). Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain and Cognition, 59,* 310–313.

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. *Journal of the Acoustical Society of America, 119,* 3034–3047.

Patel, A. D., Peretz, I., Tramo, M., & Labreque, R. (1998). Processing prosodic and musical patterns: A neuropsychological investigation. *Brain and Language, 61,* 123–144.

Patel, A. D., Wong, M., Foxton, J., Lochy, A., & Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Perception, 25,* 357–368.

Peretz, I., Ayotte, J., Zatorre, R., Mehler, J., Ahad, P., Penhune, V., & Jutras, B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron, 33,* 185–191.

Peretz, I., Champod, S., & Hyde, K. (2003). Varieties of musical disorders: The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences, 999,* 58–75.

Peretz, I., & Hyde, K. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences, 7,* 362–367.

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness. *Music Perception, 25,* 95–115.

Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception, & Psychophysics, 71,* 1385–1398. doi:10.3758/APP.71.6.1385

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., & Liotti, M. (2010). Imprecise singing is widespread. *Journal of the Acoustical Society of America, 128,* 2182–2190.

Plaut, D. C., & Kello, C. T. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In B. MacWhinney (Ed.), *The emergence of language* (pp. 381–415). Mahwah, NJ: Erlbaum.

R Development Core Team. (2012). R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing. Retrieved from www.R-project.org

Sundberg, J., & Bauer-Huppmann, J. (2007). When does a sung tone start? *Journal of Voice, 21,* 285–293.

Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences, 109,* 19027–19032.

Tillmann, B., Burnham, D., Nguyen, S., Grimault, N., Gosselin, N., & Peretz, I. (2011a). Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages. *Frontiers in Auditory Cognitive Neuroscience, 2,* 120. doi:10.3389/fpsyg.2011.00120

Tillmann, B., Rusconi, E., Traube, C., Butterworth, B., Umiltà, C., & Peretz, I. (2011b). Fine-grained pitch processing of music and speech in congenital amusia. *Journal of the Acoustical Society of America, 130,* 4089–4096.

Tillmann, B., Schulze, K., & Foxton, J. M. (2009). Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds. *Brain and Cognition, 71,* 259–264.

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,* 463–476.

Ward, W. D., & Burns, E. M. (1978). Singing without auditory feedback. *Journal of Research in Singing, 1,* 24–44.

Williamson, V. J., Liu, F., Peryer, G., Grierson, M., & Stewart, L. (2012). Perception and action de-coupling in congenital amusia: Sensitivity to task demands. *Neuropsychologia, 50,* 172–180.

Williamson, V. J., & Stewart, L. (2010). Memory for pitch in congenital amusia: Beyond a fine-grained pitch perception problem. *Memory, 18,* 657–669.

Wing, A. M., & Kristofferson, A. B. (1973). The timing of interresponse intervals. *Perception & Psychophysics, 13*(455), 460.

Wise, K. J. (2009). *Understanding "tone-deafness": A multi-compositional analysis of perception, cognition, singing and self-perception in adults reporting musical difficulties*. Ph.D. dissertation, Keele University.

Wisniewski, M. G., Mantell, J. T., & Pfordresher, P. Q. Transfer effects in the vocal imitation of speech and song. *Psychomusicology: Music, Mind and Brain.* (in press)

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication, 46,* 220–251.

Xu, Y. (2005–2012). ProsodyPro.praat. Retrieved from www.phon.ucl.ac.uk/home/yi/ProsodyPro/

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