Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages

Barbara Tillmann[1,2,3,4]*, Denis Burnham[4], Sebastien Nguyen[5,6], Nicolas Grimault[1,2,3], Nathalie Gosselin[5,6] and Isabelle Peretz[5,6]

[1] *CNRS, UMR5292; INSERM, U1028; Lyon Neuroscience Research Center, Auditory Cognition and Psychoacoustics Team, Lyon, France*
[2] *University Lyon, France*
[3] *University Lyon 1, Villeurbanne, France*
[4] *MARCS Auditory Laboratories, University of Western Sydney, Sydney, NSW, Australia*
[5] *International Laboratory for Brain, Music and Sound Research (BRAMS), Université de Montréal, Montréal, QC, Canada*
[6] *Department of Psychology, Université de Montréal, Montréal, QC, Canada*

Congenital amusia is a neurogenetic disorder that affects music processing and that is ascribed to a deficit in pitch processing. We investigated whether this deficit extended to pitch processing in speech, notably the pitch changes used to contrast lexical tones in tonal languages. Congenital amusics and matched controls, all non-tonal language speakers, were tested for lexical tone discrimination in Mandarin Chinese (Experiment 1) and in Thai (Experiment 2). Tones were presented in pairs and participants were required to make same/different judgments. Experiment 2 additionally included musical analogs of Thai tones for comparison. Performance of congenital amusics was inferior to that of controls for all materials, suggesting a domain-general pitch-processing deficit. The pitch deficit of amusia is thus not limited to music, but may compromise the ability to process and learn tonal languages. Combined with acoustic analyses of the tone material, the present findings provide new insights into the nature of the pitch-processing deficit exhibited by amusics.

Keywords: congenital amusia, pitch perception, music processing, tone-language processing

INTRODUCTION

A highly debated question is to what extent music and language share processing components (e.g., Patel, 2008). Our study contributes to this debate by investigating pitch-processing across domains in congenital amusia (or tone-deafness). Congenital amusia refers to a lifelong disorder of music processing that occurs despite normal hearing and other cognitive functions as well as normal exposure to music. We investigated here whether the impaired musical pitch perception typically found in congenital amusia might reflect a domain-general deficit that also affects pitch processing in speech.

Pitch processing is crucial in music, but also in speech processing, notably for discriminating questions and statements, as well as for emotional expressions in non-tone intonation languages (e.g., English, French); while in tone languages (e.g., Mandarin, Thai, Vietnamese), it is used for all these as well as for understanding word meaning. Tone-languages comprise 70% of the world's languages (Yip, 2002) and are spoken by more than 50% of the world's population (Fromkin, 1978). In these languages, tone variation changes (comprising predominantly FØ height and contour parameters) at the syllabic level have the same effect on word meaning as do vowel and consonant variations in non-tone languages. For examples, see **Figure 1**. In the present study, we used natural speech samples of tone languages to investigate whether pitch variations in speech might be affected by the previously described musical pitch deficit in congenital amusia.

Expertise or training in a tonal language can facilitate pitch perception and production with musical material: Mandarin, Vietnamese, and Cantonese speakers have been found to be more accurate at imitating musical pitch and discriminating intervals than English speakers (Pfordresher and Brown, 2009; see also Hove et al., 2010), as can be also reflected in subcortical pitch tracking (e.g., Krishnan et al., 2005). The influence of tone-language background has been mostly observed for relative pitch processing (e.g., intervals), and it might even lead to difficulties in pitch contour processing when non-speech target sounds resemble features of linguistic tones (Bent et al., 2006). However, it has been found that listeners with tone-language background did not differ from listeners with non-tone-language background for absolute pitch discrimination of non-speech sounds (e.g., Bent et al., 2006; Pfordresher and Brown, 2009). Interestingly, in musically trained participants, there is a link between tone-language background and single pitch processing: absolute pitch (i.e., the ability to label a tone without a reference pitch) is more prevalent among tone-language speakers than among non-tone-language speakers (Deutsch et al., 2006, 2009).

Conversely, it has been shown that musical training or expertise can improve pitch perception not only in musical contexts, but also in speech contexts. For example, musicians show improved pitch processing for the prosody of non-tonal language material (Schön et al., 2004; Magne et al., 2006) and for tone-language material, such as Thai tones (Burnham and Brooker, 2002; Schwanhäußer and Burnham, 2005) and Mandarin tones (Alexander et al., 2005; Wong et al., 2007; Lee and Hung, 2008; Delogu et al., 2010; Bidelman et al., 2011).

Previous research has thus shown some positive influences between music and speech due to expertise in music or in tone languages, and these effects suggest common pitch-processing mechanisms in
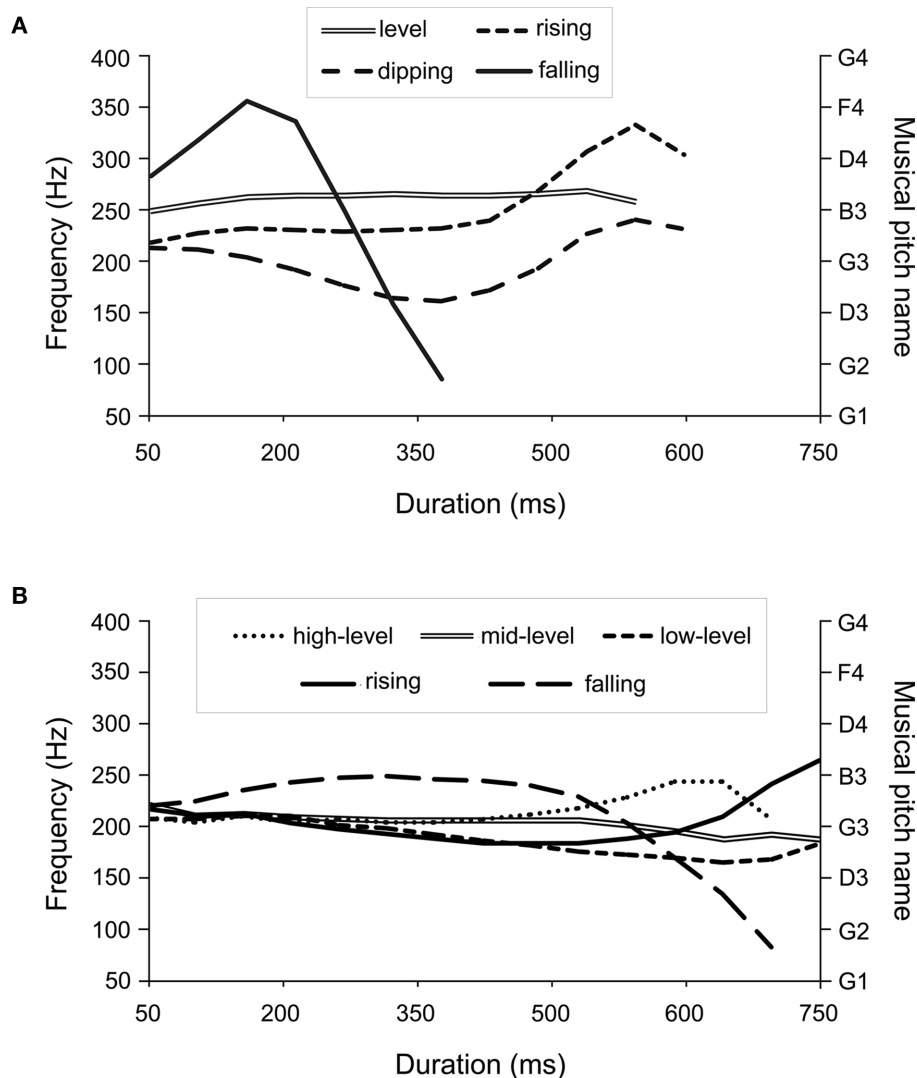
FIGURE 1 | (A) Fundamental frequency contours of the four Mandarin tones (spoken by a female speaker of Mandarin). Each tone on the syllable/ma/ represents a different lexical item. "ma1" is the level tone, "ma2" the rising tone, "ma3" the dipping tone, and "ma4" the falling tone. **(B)** Fundamental frequency contours of the five Thai tones (spoken by a female speaker of Thai). Each tone on the syllable/ma/represents a different lexical item. "ma0" is the mid level tone, "ma1" the low level tone, "ma2" the falling tone, "ma3" the high level tone, and "ma4" the rising tone.

music and speech. For example, musical training might shape basic sensory circuitry as well as corticofugal tuning of the afferent system, which is context-general and thus also has positive side-effects on linguistic pitch processing (e.g., Wong et al., 2007). Similar findings suggesting experience-dependent corticofugal tuning have been recently reported for the effects of tone-language expertise on musical pitch processing (Bidelman et al., 2011). In parallel to this previously observed training-related improvement of pitch processing from one domain to the other, the experiments reported here investigate the influence of a pitch perception deficit for music, as observed in congenital amusia, on pitch perception in speech.

Up to recently, congenital amusia has been thought to result from a musical pitch-processing disorder. Individuals with congenital amusia have difficulties recognizing familiar tunes without lyrics and detecting an out-of-key or out-of-tune note. They have impaired perception of pitch directions for pure tones (Foxton et al., 2004) and for detecting pitch deviations that are smaller than two semitones in sequences of piano notes (Hyde and Peretz, 2004) as well as in note pairs (Peretz et al., 2002). Initial reports have suggested that the deficit was restricted to pitch processing in music, and did not extend to pitch processing in speech material. Individuals with congenital amusia have been reported to be unimpaired in language and prosody tasks, such as learning and recognizing lyrics, classifying a spoken sentence as statement or question based on final falling or rising pitch information (e.g., Ayotte et al., 2002; Peretz et al., 2002).

Peretz and Hyde (2003) suggested that the difference between pitch perception in speech and music is related to the relative size of relevant pitch variations. In speech (of non-tonal languages), pitch variations are typically coarse (e.g., more than 12 semitones
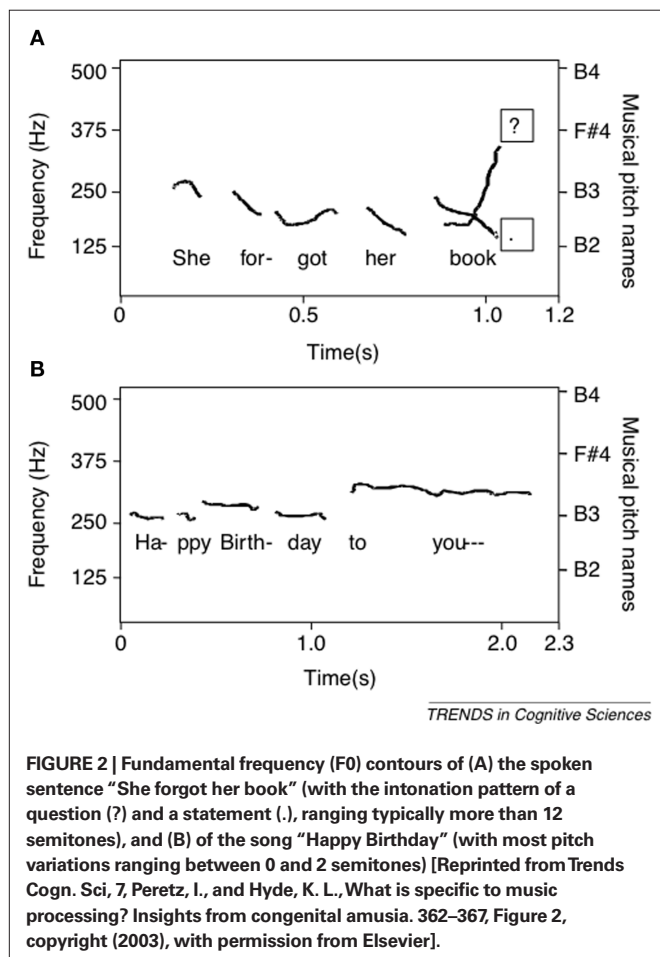
in the final pitch rise indicative of a question; see Fitzsimons et al., 2001), whereas in music, these are more fine-grained (1 or 2 semitones; Vos and Troost, 1989; see **Figure 2**). Accordingly, amusics' pitch deficit would affect music more than speech not because their deficit is music-specific, but because music is more demanding in pitch resolution. Thus, congenital amusia would represent a music-relevant deficit, not necessarily a music-specific deficit. However, when pitch changes of spoken sentences were embedded in a non-speech context (i.e., musical analogs preserving gliding-pitch changes or transforming these into discrete steps), amusics failed to discriminate these pitch changes – in contrast with their high performance level for the same pitch changes in the sentences (Patel et al., 2005). Conversely, recent data have shown that, for some amusic cases, the pitch-processing deficit can also affect the processing of speech intonation in the amusics' mother tongue (Patel et al., 2008; Jiang et al., 2010; Liu et al., 2010). In particular, a slow rate of gliding-pitch change might have deleterious effects on pitch perception in English, but not in French speech (Patel et al., 2008), although the influence of glide rate has not been replicated in a subsequent study for English in British amusics (Liu et al., 2010).

In addition to differences in the size of pitch changes, musical, and linguistic materials differ in their use of discrete, segmented events versus continuous pitch changes (i.e., glides), respectively. Foxton et al. (2004) have shown that congenital amusics have higher thresholds for segmented tones (exceeding one semitone) than for continuous

tone glides (below one semitone). These higher thresholds would affect music perception more markedly, because pitch in music is based upon discrete notes, while the better thresholds for glides might lead to less impairment for pitch processing in speech signals, with its gliding, continuous pitch changes. However, this is unlikely since amusics perform equally poorly on tone analogs of sentences made of pitch glides and discrete events (Patel et al., 2005). In a recent study using discrete segmented events (a tone for the musical material, the syllable /ka/ for the verbal material), we have observed that fine-grained pitch discrimination (i.e., 25 cents) can be impaired in amusics not only for musical sounds, but also for verbal sounds. Interestingly, pitch discrimination is better when the pitch is carried by verbal material than by musical material (Tillmann et al., submitted).

Pitch perception in congenital amusia might thus be affected by the size of pitch changes and the nature of the material (verbal, musical). The present study investigates amusics' pitch processing in tonal language material in order to address the question: do congenital amusics show deficits for lexical tone perception, thus for speech material with continuous (rather than discrete) pitch changes, and with pitch changes larger than those that are relevant in music, but smaller than those used in statements/questions in their mother tongue (see **Figures 1 and 2**; Fitzsimons et al., 2001)? We tested French-speaking amusics for their perception of Mandarin tones (Experiment 1) and Thai tones (Experiment 2). We here used monosyllabic words (in contrast to sentences or phrases in previous studies) to keep memory load relatively low, in particular as amusic individuals show impaired short-term memory for pitch information (Gosselin et al., 2009; Tillmann et al., 2009; Williamson et al., 2010). Experiment 2 additionally tested the perception of the same pitch changes in non-verbal, musical analogs. Furthermore, for both Experiments 1 and 2, we present acoustic analyses of the tone-language stimulus materials, and compare the acoustic features of the stimulus materials with participants' behavioral performance, in order to locate the critical acoustic information used by amusic and control participants.

The overall objective of our study is to further understand the nature of the pitch-processing deficit experienced by individuals with congenital amusia, particularly because congenital amusia is now known to have neurogenetic correlates (Drayna et al., 2001; Peretz et al., 2007), and these may not be music-specific, but apply also to speech, as suggested by recent work on amusics' perception of pitch in speech of their mother tongue (e.g., Liu et al., 2010; Nan et al., 2010). Testing the perception of tone-language materials allowed us to use natural speech that contained smaller pitch differences than those occurring in native non-tonal speech (English or French), as in the sentences used in Patel et al. (2005) and Liu et al. (2010). Another advantage of using non-native pitch variations in speech is that, as the words have no meaning for the participants, they are free to respond to the acoustic parameters of the speech without any added complication of semantic significance. This also has the advantage that the non-native tones can be presented as speech that has full speech-shaped spectral information, albeit devoid of semantic significance; and as non-speech – in which the same tones are converted to musical stimuli. In this way, the same acoustic aspects of speech, notably here the pitch information, can be presented in speech and non-speech contexts in which the main difference between speech and more musical non-speech stimuli is the differences in spectral

FIGURE 2 | Fundamental frequency (F0) contours of (A) the spoken sentence "She forgot her book" (with the intonation pattern of a question (?) and a statement (.), ranging typically more than 12 semitones), and (B) of the song "Happy Birthday" (with most pitch variations ranging between 0 and 2 semitones) [Reprinted from Trends Cogn. Sci, 7, Peretz, I., and Hyde, K. L., What is specific to music processing? Insights from congenital amusia. 362–367, Figure 2, copyright (2003), with permission from Elsevier].

make-up. In addition, testing amusics who were non-native speakers with speech signals that were tone-language materials also allowed us to aim for converging evidence with findings on tone-language processing recently reported for amusics who were native speakers (Jiang et al., 2010; Nan et al., 2010). Note that previous research has shown that non-native (non-amusic) listeners still engage a speech listening mode (as reflected by the linguistic constraints of their mother tongue) when processing non-native tone-language materials (e.g., Burnham et al., 1996; Burnham et al., submitted).

EXPERIMENT 1

Mandarin Chinese uses four tones characterized by their pitch trajectories, traditionally numbered as tones 1–4: tone 1 is high level, tone 2 is mid-rising, tone 3 mid-dipping (or mid-falling–rising), and tone 4 is high-falling (see **Figure 1A**). Tone 1 has little fundamental frequency (FØ) movement (and so is often referred to as a level tone), whereas tones 2–4 have more FØ movement (and so often referred to as contour or dynamic tones, Abramson, 1978): tone 2 has a rising FØ pattern, tone 3 a falling–rising FØ pattern, and tone 4 a falling FØ pattern.

Experiment 1 tested native French-speaking congenital amusics' for their pitch discrimination with this unfamiliar language material. Even though normal French listeners do not perceive tone contrasts categorically, they are sensitive to tone contour variations (Hallé et al., 2004). A same-different paradigm using monosyllabic Mandarin Chinese words was employed; it was taken from Klein et al. (2001) who showed that normal English or Mandarin speaking participants reached high levels of performance (although English-speakers performed at a level slightly below that of Mandarin speakers, 93 versus 98% accuracy). If the hypothesis of domain-general pitch-processing mechanisms is true, then it can be predicted that amusics' musical pitch deficit should lead to impaired performance in this discrimination task.

METHOD

Participants

The amusic group and the control group each comprised 20 adults who were native French speakers (from Canada and France). The groups were matched for gender, age, education, and musical training (see **Table 1**). All participants completed the Montreal battery of evaluation of amusia (MBEA; Peretz et al., 2003), which is currently widely used in research investigating congenital amusia. The MBEA involves six tests that aim to assess the various components that are known to contribute to melody processing. The stimuli are novel melodic sequences, played one note at a time on a piano; they are written in accordance to the rules of the tonal structure of the Western idiom. These melodies are arranged in various tests so as to assess abilities to discriminate pitch and rhythmic variations, and to recognize musical sequences heard in prior tests of the battery. Peretz et al. (2003) tested a large population and defined a cut-off score (78%) under which participants can be defined as amusics or above which participants were normal. Participants' individual scores for the full battery were below cut-off for the amusic group, but not for the control group (**Table 1**). One amusic participant reported living in China from age 6 to 10; he indicated that he took lessons in Mandarin Chinese but with difficulty and that he spoke either French (he attended a French school) or English, not Mandarin, during that period. Note that 16 out of the 20 amusics have also participated in the experiment testing pitch change detection in verbal and non-verbal materials (Tillmann et al., submitted).

Material

The 98 recordings of monosyllabic Mandarin words (produced by a native female Mandarin speaker) from Klein et al. (2001) were used (see **Table 3** for acoustic descriptors). In all there were 51 different words (13 words with level tone, 10 words with rising

---

Table 1 | Number of participants per group (amusic/controls) for Experiments 1 and 2, followed by mean age (years), gender distribution, mean education (years), level of musical instruction as well as mean scores obtained on the Montreal battery of evaluation of amusia (MBEA), for the entire test (global score), and the subtests.

	Experiment 1 (Mandarin)			Experiment 2 (Thai)		
	Amusics ($n = 20$)	Controls ($n = 20$)	t-Test	Amusics ($n = 19$)	Controls ($n = 19$)	t-Test
Age	48.5 (17.2); 19–71	47.9 (17.1); 19–73	$p = 0.91$	50.5 (15.9); 21–71	49.5 (15.9); 22–72	$p = 0.85$
Gender	12 F; 8 M	13 F; 7 M	–	11 F; 8 M	12 F; 7 M	–
Education (years)	16.1 (2.5)	15.2 (2.6)	$p = 0.28$	16.3 (2.3)	15.3 (2.5)	$p = 0.19$
Musical instruction[1]	1.6 (0.9)	1.4 (0.7)	$p = 0.36$	1.6 (0.9)	1.5 (0.8)	$p = 0.68$
MBEA[2]	–	–	–	–	–	–
Global	66.4 (6.9)	89.9 (4.2)	$p < 0.001$	65.8 (6.9)	90.1 (4.1)	$p < 0.001$
Scale	62.1 (8.3)	90.7 (7.0)	$p < 0.001$	62.5 (8.0)	91.2 (7.1)	$p < 0.001$
Contour	61.1 (8.0)	90.5 (7.1)	$p < 0.001$	60.8 (8.3)	91.1 (6.9)	$p < 0.001$
Interval	61.4 (9.5)	88.0 (7.0)	$p < 0.001$	60.8 (9.8)	88.1 (7.4)	$p < 0.001$
Rhythm	76.2 (14.4)	92.2 (5.8)	$p < 0.001$	74.6 (14.4)	91.2 (6.9)	$p < 0.001$
Meter	66.5 (13.3)	88.3 (6.7)	$p < 0.001$	64.6 (14.0)	88.6 (6.0)	$p < 0.001$
Memory	71.8 (10.3)	91.2 (6.2)	$p < 0.001$	71.9 (9.9)	92.1 (6.1)	$p < 0.001$

Standard deviations are indicated in brackets. For age, the range is also indicated with minimum and maximum age (years).
F, female; M, male; [1]based on an interval scale of musical training (years): 1 = less than a year, 2 = 1–3 years, 3 = 4–6 years, 4 = 7–10 years, 5=more than 10 years; [2]percentage of correct responses.

---

tone, 15 words with dipping tone, and 13 words with falling tone), with multiple recordings of 25 words for use in same-word pairs (thus leading to acoustic variability between words used in the same-word condition). These words consisted of various consonant–vowel (CV) combinations (e.g., /nju//kuaI//tʃuən/). Words were presented in 49 pairs: 24 composed of word pairs with the same CV combination but differing in the tone, and 25 composed of different renditions of the same words, and so having the same tone[1]. For all participants, word combinations presented in a pair (and word order within each pair) were the same. Each of the 98 recordings was used once in the task. The experiment was run with E-Prime software (Schneider et al., 2002).

Procedure
Within each pair, the first word was followed by a silent period of 350 ms, followed by the second word. Following each pair, listeners were asked to judge whether the two words of the pair were the same or different, by pressing one of the indicated keys on a computer keyboard. Participants were not explicitly told that the relevant dimension for discrimination was pitch. Listeners were first familiarized with the task by means of three practice pairs followed by error feedback, and then moved to the 49 experimental pairs without feedback. After participants' responses, the next pair was presented after a delay of 2 s. The order of presentation of word pairs was randomized for each participant. The experimental session lasted for about 10 min.

RESULTS
Performance was analyzed by calculating proportions of Hits (number of correct responses for different trials/number of different trials) minus False Alarms (FAs; number of incorrect responses for same trials/number of same trials). The amusic group performance was significantly below that of the control group, $F(1,38) = 11.63$, $p = 0.002$. Nevertheless, as can be seen in **Figure 3A**, there was substantial overlap between the groups. Only three amusic individuals performed 2 SD below average control performance. The amusic participant who had spent some time in China reached a performance level of 0.46, i.e., in the lower performance range of the amusic group and within the 2 SD of average control performance.

Correlations[2] between performance and the six subtests of the MBEA reached significance only for the interval subtest in amusics, $r(18) = 0.47$, $p = 0.038$.

To specify whether the observed group difference was associated with a reduced sensitivity to lexical tone pitch changes, or rather to a propensity to judge a "same" pair to entail a change, we ran two two-sided, independent t-tests for Hits and FAs, respectively. Amusics made fewer hits (0.66) and more FAs (0.09) than controls (0.79 versus 0.05), $t(38) = 2.70$, $p = 0.010$, and $t(38) = 2.23$, $p = 0.031$, respectively. Amusics performed more poorly than controls by both failing to discriminate pairs that were different, and erroneously judging same-word pairs to be different. In general then, it appears that amusics had a less clear grasp of this speech-based pitch discrimination task than did controls.

---

[1]One "different" pair had to be discarded due to distortion in the acoustic signal.

[2]All reported correlations were Pearson correlation coefficients that were tested with two-tailed tests.

An additional analysis separated tone pairs as a function of the different tone comparisons (Level–Rising, Level–Dipping, Level–Falling, Rising–Dipping, Rising–Falling, Dipping–Falling). An 6×2 ANOVA on Hit–FA rates with tone comparisons as within-participants factor and Group (amusics/controls) as between-participants factor revealed a main effect of tone comparison [$F(5,190) = 10.81$, $p < 0.0001$, MSE = 0.05] and a main effect of group [$F(1,38) = 11.31$, $p = 0.002$, MSE = 0.17], but no interaction ($p = 0.39$). Control participants performed generally better than the amusic participants, but both groups showed higher performance level for the pairs comparing level and dipping tones (0.68 for amusics and 0.87 for controls) and for pairs comparing dipping and falling tones (0.71 for amusics and 0.86 for controls) than for the other tone pair comparisons (0.48 for amusics and 0.66 for controls).

DISCUSSION
Experiment 1 revealed that amusics, who were speakers of a non-tonal language (French), encountered difficulties in Mandarin lexical tone discrimination (in comparison to their matched controls). In addition, amusics' performance correlated with their performance in the interval test of the MBEA, which requires the discrimination of tone sequences differing by interval sizes: the lower their performance on the interval test with melodies, the lower their performance in the lexical tone discrimination for Mandarin. These findings support the conclusion that amusics' pitch deficit in melodies extends to the perception of pitch in speech material.

While amusics' average performance was below the group performance of controls, there was considerable overlap in performance ranges between the two groups. The relatively comparable performance of the amusics might be due to some pitch variations in the Mandarin tones (or the comparisons in some of the pairs) that might exceed amusics' thresholds (e.g., larger than two semitones). To increase the difficulty of the task and to assess the generality of the findings, Experiment 2 tested amusics and controls with a same-different paradigm using Thai tones.

EXPERIMENT 2
Standard Thai uses five tones: three level tones (low, mid, high) and two contour tones (rising and falling), referred to as static and dynamic tones respectively (see **Figure 1B**; Abramson, 1962). The tone systems of Thai and Mandarin are different in relation to the number of tones as well as their pitch height, durations and start/end points. They also show some similarities. For example, both Thai and Mandarin have one rising and one falling contour tone and at least one level tone. Thai, however, contains five (not four) different tones, and these tones are based on smaller pitch changes together with weaker contribution of durational differences than in Mandarin. Previous studies have shown that English-speaking children and adults can discriminate Thai tones (Burnham et al., 1996; Burnham and Francis, 1997; Burnham and Brooker, 2002), and Experiment 2 tested for the first time congenital amusics on this material.

In addition, to further our understanding of domain-generality versus -specificity of pitch processing, Experiment 2 compared amusics' performance for the lexical tones to their performance on musical analogs thereof (pitch variations of the Thai tones applied
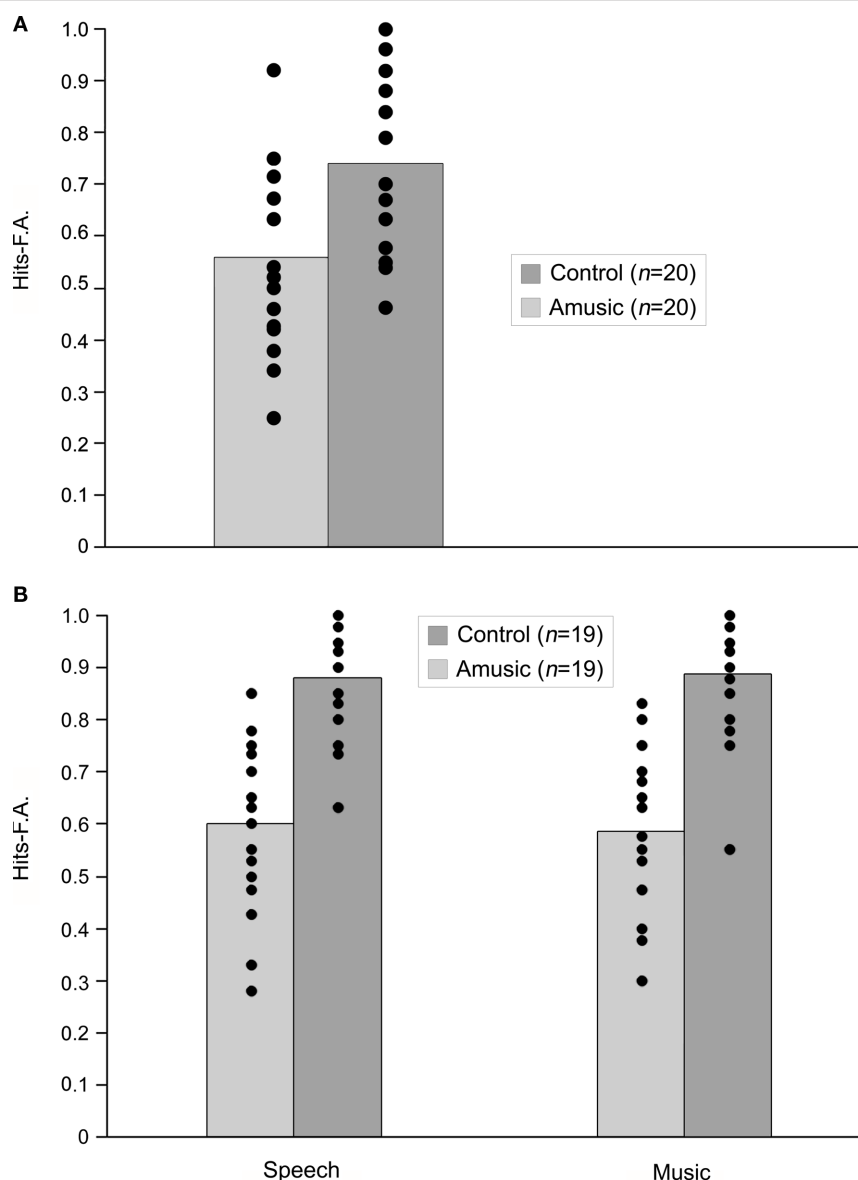
FIGURE 3 | Performance [expressed as Hits – False Alarms (FA)] for amusic and control participants in Experiment 1 with the Mandarin material (A) and in Experiment 2, with the Thai material (speech) and its musical analogs [music; (B)]. Dots represent individual participants.

to a violin sound). Previous studies have shown that (1) normal, English-speaking listeners performed worse for the speech signals than for musical analogs or low-pass filtered versions of the speech signals (Burnham et al., 1996; Burnham and Brooker, 2002), and (2) musical training boosted overall performance levels, with musicians without absolute pitch performing better than non-musicians, and musicians with absolute pitch performing the best (Burnham and Brooker, 2002). Based on these positive transfer effects of musical training, Burnham and Brooker (2002) concluded that speech and music perception are not independent, and that musical training and absolute pitch ability can affect speech processing. Thus, amusics' musical deficit would predict impaired processing for pitch in the speech material here, in line with data of Experiment 1 as well

as some previous work on amusics' perception of their mother tongue, whether non-tonal (Patel et al., 2008; Liu et al., 2010) or tonal (Nan et al., 2010).

Impaired pitch processing for speech material has been also reported for a task requiring fine-grained pitch change detection in sequences of a repeated syllable (/ka/) (Tillmann et al., submitted). Even though impaired, amusics' performance was less impaired for these syllable sequences than for sequences with repeated tones (carefully matched to the syllables for their acoustic features). This performance difference between speech and musical sounds might be linked to differences in the energy distribution of the sounds' spectrum, notably by the presence versus absence of formants, and/ or to higher-level processing related to strategic influences (see

below for further discussion). In the Tillmann et al. (submitted) study, the to-be-detected pitch changes were instantiated between syllables (or tones), thus between segmented events, and not within a given event with continuous pitch changes as in the materials of Experiments 1 and 2 here. In Experiment 2, we thus hypothesized that for the processing of pitch in Thai tones, amusics' performance should benefit from the speech signal, leading to some boost in pitch processing (compared to musical analogs), at least for the most severely impaired amusics (Tillmann et al., submitted). Controls, however, should perform better for the musical material (Burnham and Brooker, 2002; Tillmann et al., submitted).

METHOD

Participants

The amusic group and the control group each comprised 19 French-speaking adults (from Canada and France); 18 of the amusics and 17 of the controls also participated in Experiment 1 with the Mandarin tones. As in Experiment 1, the groups were matched for gender, age, education, and musical training, with MBEA scores below cut-off for the amusic group only (**Table 1**). Note that 14 out of the 19 amusics tested have also participated in the experiment testing pitch change detection in verbal and non-verbal material (Tillmann et al., submitted).

Materials

Five tokens of /ba/ for each of the five Thai tones, recorded by a female speaker, were taken from Burnham et al. (1996) (see **Table 3** for acoustic descriptors). Instead of using musical stimuli that were created by a professional musician imitating the lexical tones on the violin as in Burnham et al. (1996), the musical analogs were created by applying the pitch contour, temporal envelope, duration, and intensity of the Thai tones to a violin sound (a steady-state violin sound of an original duration of 1 s, which was then shortened to that of the tones). First, overall duration, pitch contour, and temporal envelope (computed as the half-wave rectified signal low-pass filtered at 80 Hz) were extracted from each of the verbal tokens using STRAIGHT (Kawahara et al., 1999). Then, duration and pitch contour of the verbal sound were applied to the violin sound with STRAIGHT. Finally, the temporal envelope and the RMS value of the verbal sound were applied to the transformed violin sound. Overall, 25 musical sounds, corresponding to the 25 verbal sounds (five tokens for each of the five Thai tones) were generated. For each type of material (verbal, musical), 40 same pairs and 40 different pairs were created. For the 40 same pairs, 8 pairs were created for each of the five tones. For the 40 different pairs, each tone was presented with one of the other four tones four times. Same pairs consisted of different tokens of the same tone and over all pairs, different tokens were used across participants. The experiment was run with PsyScope software (Cohen et al., 1993).

Procedure

Within each pair, the first item was followed by a silent period of 500 ms, followed by the second item (as in Burnham et al., 1996; Burnham and Brooker, 2002). For each pair, listeners were asked to judge whether the speaker (or the musician) pronounced (played) the two syllables (notes) in the same way or in a different way. As in Experiment 1, no explicit references were made to the pitch

dimension, the changes were referred to as "different ways of pronouncing the syllable" or "different ways of producing the violin sound." Participants indicated their answers by pressing one of two keys on a computer keyboard. Listeners were first familiarized with the materials by two example pairs (one different pair, one same pair), which could be repeated for clarification if required. The 80 verbal and 80 musical pairs were each separated into two blocks. The four resulting blocks were presented in either the order verbal–musical–musical–verbal or the order musical–verbal–verbal–musical, counterbalanced across participants. Within each block, trials were presented in randomized orders for each participant. No feedback was given for the experimental trials, and the next trial started when participants pressed a third key. The experimental session lasted for about 20 min.

Pretest

A pretest was run to confirm that the musical material newly constructed for Experiment 2 replicated the result pattern of Burnham et al. (1996). Twenty English-speaking students from the University of Western Sydney participated in the pretest (mean age: 23.2 ± 8.73), with average instruction on a musical instrument of 0.33 years ± 0.73 and a median of 0 years. Performance was analyzed by calculating proportions of Hits (number of correct responses for different trials/number of different trials) minus FAs (number of incorrect responses for same trials/number of same trials). Results replicated better performance for musical material (0.85 ± 0.03) than for verbal material (0.78 ± 0.04), $F(1,19) = 4.37$, $p = 0.05$, as previously observed by Burnham et al. (1996), Burnham and Brooker (2002), Burnham and Francis (1997) in non-musicians. In addition, as in previous data, performance reflected the degree of acoustic changes between the sounds of a pair: performance was better for pairs combining two contour tones (rising, falling) than for pairs with two level tones (low, mid, high) or mixed pairs, $F(2,38) = 6.68$, $p = 0.003$.

RESULTS

As in Experiment 1, proportions of Hits (number of correct responses for different trials/number of different trials) and FAs (number of incorrect responses for same trials/number of same trials) were calculated. These proportions were analyzed by a 2×2 ANOVA with Material (verbal, musical) as a within-participant factor and Group (amusics, controls) as a between-participants factor. Only the main effect of group was significant, $F(1,36) = 46.08$, $p < 0.0001$: the amusic group performed below the level of the control group, although there was substantial overlap between the groups (see **Figure 3B**). Twelve amusics for the verbal material and 11 amusics for the musical material performed 2 SD below average control performance. The amusic who had lived in China (see Experiment 1) had performance levels of 0.48 and 0.40 for verbal and musical materials, respectively. No other effects were significant, $ps > 0.36$. Thus, performance did not differ significantly for verbal and musical materials, either for amusics or controls.

For the verbal material, the only correlation between performance and scores in the subtests of the MBEA was for the meter subtest in controls, $r(17) = 0.47$, $p < 0.04$. In amusics, the correlation with the interval test was marginally significant, $r(17) = 0.41$, $p < 0.08$, in agreement with the findings for the Mandarin material

in Experiment 1. For the musical material, no correlations were significant, although the correlation with the meter subtest for controls was marginally significant, $r(17) = 0.45, p < 0.053$. Correlation for performance between verbal and musical tasks were significant both for amusics, $r(17) = 0.64, p < 0.003$, and for controls, $r(17) = 0.91, p < 0.0001$.

As for Experiment 1, we also analyzed group differences with two-tailed, independent t-tests for Hits and FAs, respectively (**Table 2**). As in Experiment 1, amusics made fewer Hits and more FAs than controls for both the verbal material, $t(36) = 6.13, p < 0.0001$, and $t(36) = 3.39, p = 0.001$, respectively, and the non-verbal material, $t(36) = 5.86, p < 0.0001$, and $t(36) = 2.75, p = 0.009$, respectively.

As in Burnham et al. (1996), we separated performance (Hits–FAs) for pairs combining the two contour tones (i.e., rising, falling), or two of the level tones (i.e., low, mid, or high) or mixed pairs with one contour tone and one level tone (**Table 2**). Performance was analyzed in a $3 \times 2 \times 2$ ANOVA with Type (contour–contour, level–level, contour–level) and Material (verbal, musical) as within-participant factors and Group (amusics, controls) as a between-participants factor. The main effects of group and of type were significant, $F(1,36) = 40.85, p < 0.0001$, and $F(2,72) = 62.05, p < 0.0001$, respectively, as was their interaction, $F(2,72) = 23.38, p < 0.0001$. Overall amusics performed worse than controls; this difference was smaller for pairs with two contour tones, while still being significantly below controls, $F(1, 36) = 5.11, p = 0.03$. Even though the three-way interaction between Group, Type, and Material was not significant ($p = 0.38$), we ran two additional 3×2 ANOVAs with Type and Material as within-participants factors in each group separately, with the goal to further investigate amusics' and controls' sensitivity to the type of comparison pairs. Only the main effect of type was significant in both amusics, $F(2,36) = 57.54, p < 0.0001$, MSE $= 0.02$, and controls, $F(2,36) = 7.76, p = 0.002$, MSE $= 0.01$. For controls, performance was better for contour–contour pairs than for contour–level ($p = 0.003$) and level–level pairs ($p = 0.01$), while these latter two did not differ ($p = 0.23$). For amusics, performance was also better for contour–contour pairs than for contour–level pairs ($p < 0.0001$) and level–level pairs ($p < 0.0001$), but in addition, amusics were sensitive to the difference between these two latter pairs, with lower performance

for the level–level pairs ($p = 0.004$). Note that for the Level–Level pairs, the advantage of the verbal material over the musical material observed in the overall mean performance was not significant in amusics $p = 0.24$. No other effects were significant, $ps > 0.13$. The relative difficulty of the pair types was also reflected in the number of amusics performing 2 SD below average control performance: only 2 amusics (out of 19) performed below average control performance for contour–contour pairs, but 12 for contour–level and level–level, respectively (i.e., for verbal materials).

The analyses of the entire amusic group, reported above, did not show a significant advantage of the verbal material over the musical material, either for the overall material set or for pairs comparing two level tones. Following Tillmann et al. (submitted), in which the advantage of verbal over musical material was mostly observed for amusics who had severe pitch deficits, we divided the amusics based on their pitch thresholds. Fifteen of the 19 amusics had previously participated in a pitch perception threshold test[3]; their overall average threshold was 1.49 semitones (± 1.13), but with thresholds ranging from 0.13 to 4. We separated amusics into two groups: eight amusics with thresholds below one semitone (mean of 0.68, ranging from 0.13 to 0.97) and seven amusics with thresholds above one semitone (mean of 2.41, ranging from 1.3 to 4).

For amusics with thresholds above one semitone, performance was significantly better for the verbal material than the musical material for the entire material set (0.55 versus 0.45, $p = 0.03$) and for pairs comparing two level tones (0.52 versus 0.29, $p = 0.04$). For both comparisons, six of the seven amusic showed higher mean performance for the verbal material. This advantage was not observed for amusics with thresholds below one semitone, either for the entire material set (0.60 versus 0.65, $p = 0.38$) or for the level–tone pairs (0.51 versus 0.52, $p = 0.90$). To further investigate this group comparison, we run a 2×2 ANOVA with the two subgroups of amusics as between-participants factor and material (verbal, non-verbal) as within-participants factor. Amusics with thresholds below 1 ST tended to perform better than amusics with thresholds above one semitone, $F(1,13) = 3.49, p = 0.08$, MSE $= 0.03$. The main effect of material was not significant ($p = 0.38$), but material interacted with group, $F(1,13) = 5.77, p = 0.03$, MSE $= 0.007$: for amusics with thresholds below one semitone, performance did not differ between the two sets of materials ($p = 0.30$), while for amusics with thresholds above one semitone, performance was better for the verbal material than for the musical material ($p = 0.04$). When the amusics with threshold below one semitone were directly compared to the control group in an additional ANOVA, the interaction between group and material was not significant ($p = 0.41$). Control participants performed better than amusic participants, as shown by the main effect of Group, $F(1,25) = 33.28, p < 0.0001$, MSE $= 0.02$. In addition, overall performance tended to be better for the musical material than the verbal material, even though the effect of material failed to reach significance, $F(1,25) = 3.21, p = 0.09$, MSE $= 0.004$.

Table 2 | Discrimination performance for amusic and control participants for the Thai tones (speech material) and their musical analogs (music) in Experiment 2. The two top lines present Hits and False Alarms. The three bottom lines present Hits minus False Alarms separated for pairs comparing two types of contour tones (rising, falling), two types of level tones (low, mid, high) or one contour tone with a level tone.

	Amusic group		Control group	
	Speech	**Music**	**Speech**	**Music**
Hits	0.71 (0.03)	0.70 (0.04)	0.92 (0.02)	0.94 (0.02)
False alarms	0.11 (0.02)	0.12 (0.02)	0.04 (0.01)	0.05 (0.02)
Contour–contour	0.86 (0.04)	0.86 (0.04)	0.95 (0.02)	0.94 (0.02)
Level–level	0.54 (0.04)	0.47 (0.05)	0.86 (0.03)	0.86 (0.04)
Contour–level	0.58 (0.04)	0.60 (0.04)	0.87 (0.02)	0.90 (0.02)

SE are indicated in brackets.

---

[3]To determine pitch perception thresholds, a two-alternative forced-choice task was employed with adaptive tracking using a two-down, one-up staircase procedure. Participants were presented with two pairs of pure tones: in one pair the tones had the same pitch, and in the other pair the tones differed in pitch. They were asked to decide whether the first or the second pair contained a pitch difference (see Tillmann et al., 2009, for details).

Note that in the 15 amusics considered, thresholds correlated with performance level for musical material, $r(13) = -0.62$, $p < 0.02$, but not for the verbal material, $r(13) = -0.29$.

DISCUSSION

Experiment 2 tested amusics' perception of Thai tones and their musical analogs in a same-different paradigm. Findings supported those of Experiment 1 on Mandarin tones and suggest that amusics' pitch deficit extends to the perception of pitch in speech material. As in Experiment 1, the French-speaking amusics encountered difficulties in lexical tone discrimination, and their performance tended to correlate with their score on the interval test of the MBEA. While we observed again an overlap in performance ranges in amusic and control groups, considerably more amusics performed below the controls' distribution (i.e., 12 out of the 19 amusics for the verbal material). This observation suggests that for amusics the task with the Thai material was more difficult than the task with the Mandarin material. This might be due to the larger set of tones used (five instead of four), the smaller pitch range covered by Thai tones (see **Figure 1B**), the more standardized material solely using the syllable /ba/ (instead of using a range of CV syllables), or the fact that Thai tones do not vary in duration as much as do Mandarin tones. In addition, the delay between the to-be-compared syllables was slightly longer in Experiment 2 than in Experiment 1 (500 versus 350 ms). In this regard, the recently reported memory deficit of amusics for pitch material might thus have contributed to make the task with the Thai materials more difficult (e.g., Gosselin et al., 2009; Tillmann et al., 2009; Williamson et al., 2010). However, even if we cannot exclude the contribution of any pitch memory deficit in amusics, its contribution should be rather minor as the delays in our tasks were considerably shorter than the delays tested in the previous studies (e.g., between 1 and 15 s in Williamson et al., 2010).

The analyses separating pairs as a function of tone categories (i.e., pairs comparing contour tones only, level tones only or mixed pairs) revealed amusics' sensitivity to the acoustic features in the presented material. As for controls in the present experiment as well as participants in previous studies (see Burnham and Brooker, 2002), amusics performed better with pairs that required the comparison of two contour tones, which involve larger acoustic differences than the comparison of two level tones, for example.

Experiment 2 tested amusics' perception not only with lexical tones, but also with musical analogs of these tones. For amusics, the findings support previous conclusions for discrete pitches (Tillmann et al., submitted): even though amusics appeared impaired overall for speech and musical materials, the amusics with the largest pitch deficits benefited from the speech material, leading to improved performance. In controls, the reverse pattern (worse performance for the speech materials) was only observed in the Australian-English language student group in the pretest, but not for the French-language matched control group (even though the mean performance difference pointed in the expected direction, but $p = 0.14$). This difference between the control group and participants in our pretest as well as participants in Burnham and Brooker (2002) might be attributed to the fact that our controls were French-speaking, while participants in the pretest and the study by Burnham and Brooker were English-speaking (see Patel

et al., 2008, for pitch-processing differences in French- and English-speaking participants). The comparison of performance patterns (0.88 and 0.89 Hits–FAs for matched controls for verbal and musical material, versus 0.60 and 0.80 in Burnham and Brooker, 2002) suggest that both language groups perform similarly for the music material, while English-speaking participants show lower performance level for the speech materials than did the French-speaking participants. This might be linked to the observation that native English-speakers disregard supra-segmental cues to stress in word recognition (e.g., Cutler 2009), which might thus attenuate their performance for the speech material here[4]. Despite these differences between English and French controls, the important point here is that the amusics with greater pitch deficits *benefited* from speech material, whereas for controls, be they French or English speakers, the reverse is the case – there is a benefit for the non-speech, musical materials.

ACOUSTIC CORRELATES OF PERFORMANCE

In order to investigate the acoustic information used by amusic and control participants in Experiments 1 and 2, we analyzed Mandarin and Thai tones for the information contained in the pitch dimension (FØ mean, slope and movement, and duration of the voiced pitch component of the syllables) as well as the overall sound duration of each word and mean intensity. We added these latter non-pitch features into the analyses as we reasoned that amusic participants might use these alternative cues to aid their discrimination. The purpose of these analyses was to estimate the difficulty of the same-different task for the experimental pairs on the basis of the acoustic differences for the stimuli. Accordingly, we calculated for each acoustic feature the distance between the two items presented in each pair. If listeners use a given acoustic feature in their judgments, then larger differences on this acoustic dimension between the items should lead to higher accuracy, and smaller differences to lower accuracy.

ACOUSTIC ANALYSES

All acoustic analyses were conducted in a Matlab computing environment. STRAIGHT (Kawahara et al., 1999) was used to compute pitch contours for all stimuli. All pitch contours were subsequently fitted in a least squares sense with a 4th-degree polynomial to avoid fast pitch variation and to capture the overall shape of the pitch contour. Prior visual inspection of the pitch contours suggested the fitting of this polynomial degree: it allowed for four inflexion points or three reversals in the pitch contour. These pitch contours are displayed both individually and as average pitch contours in **Figure 4**. Note that the degree of similarity between tokens in the Thai material (all based on the syllable /ba/) was understandably higher than that between the items of the Mandarin material (which were based on various CV syllables).

For each sound file, the following steps were performed. First, the duration of the whole spoken sound was extracted. Second, several parameters were computed from the fitted pitch contour

---

[4]Note that for the more artificial speech material (synthesized syllable /ka/ and its matched musical analog in Tillmann et al., submitted), the French speakers showed a disadvantage for pitch change detection solely for the smallest pitch changes used (i.e., 25 cents), but not for the larger changes for which performance was overall high.

for each wave file: the *duration* of the pitch contour, the mean *pitch* (average FØ over the entire syllable), the mean *slope* (in semitone/s), which provided an estimate of the direction of variation of the pitch contour, and the mean of the *absolute value of the slope* of the pitch contour (in semitone/s), which provided an estimate of the overall pitch movement. Third, the mean *RMS* value was calculated using an arbitrary dB scale. These values were then averaged for each of the linguistic categories for Mandarin tones (level tone, rising, dipping, falling) and for Thai tones (low, mid, high, rising, falling) and are presented in **Table 3**.

"ACOUSTIC DISTANCES" BETWEEN THE PAIR MEMBERS AND LINK TO BEHAVIORAL DATA

Calculation of acoustic distances

For each acoustic feature, we calculated the "acoustic distance" between the two items of a pair using Euclidean distance computation. For the Mandarin tones, for which each participant was presented with the exact same set of tone pairs with only order of trials differing between participants, we calculated the distances between the items of each different pair as well as each same pair (i.e., consisting of different recordings of the same word, see
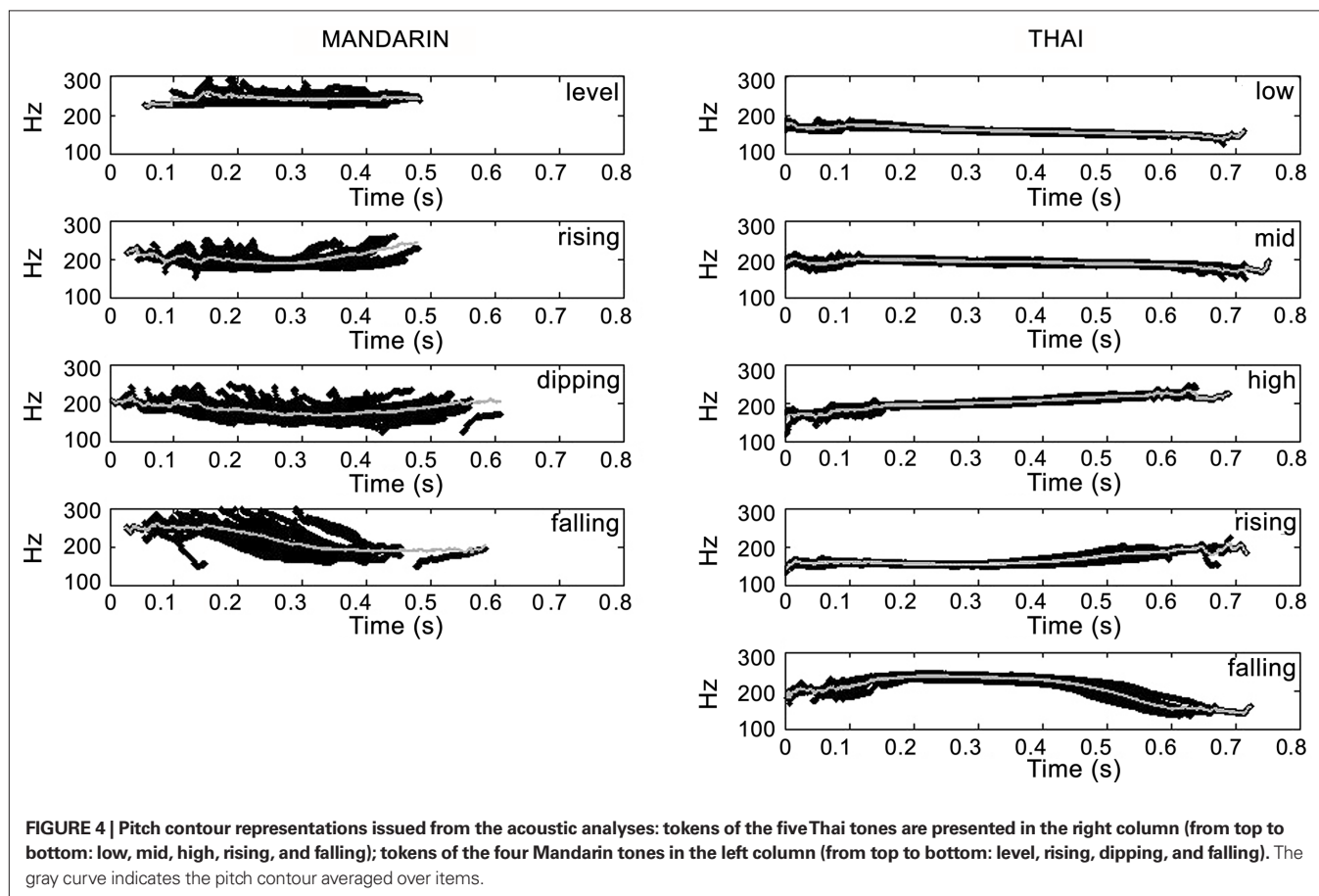

FIGURE 4 | Pitch contour representations issued from the acoustic analyses: tokens of the five Thai tones are presented in the right column (from top to bottom: low, mid, high, rising, and falling); tokens of the four Mandarin tones in the left column (from top to bottom: level, rising, dipping, and falling). The gray curve indicates the pitch contour averaged over items.

Table 3 | Mean values of the acoustic descriptors for the four Mandarin tones and the five Thai tones.

Language	Tone	Pitch, FØ (Hz)	Slope (ST/s)	Absolute slope (ST/s)	Pitch duration (s)	Sound duration (s)	RMS (arb dB)
Mandarin	Level	243.41 (8.03)	−4.90 (3.95)	13.18 (5.26)	0.28 (0.03)	0.41 (0.04)	47.05 (2.80)
	Rising	204.64 (5.31)	8.35 (3.74)	28.80 (7.20)	0.32 (0.03)	0.38 (0.05)	47.78 (3.34)
	Dipping	184.03 (5.94)	−0.85 (4.44)	28.75 (6.15)	0.40 (0.07)	0.51 (0.05)	45.61 (2.29)
	Falling	212.22 (5.56)	−14.42 (10.53)	32.21 (7.70)	0.30 (0.07)	0.37 (0.05)	47.33 (3.24)
Thai	Low	158.99 (1.42)	−4.38 (1.01)	5.20 (0.81)	0.66 (0.05)	0.67 (0.06)	66.24 (2.84)
	Mid	188.75 (3.94)	−3.07 (1.35)	5.99 (1.88)	0.71 (0.03)	0.72 (0.02)	69.01 (1.53)
	High	199.64 (2.31)	9.91 (2.58)	10.39 (2.48)	0.61 (0.02)	0.66 (0.06)	65.53 (2.73)
	Rising	164.53 (1.71)	6.58 (0.74)	10.64 (1.56)	0.64 (0.06)	0.75 (0.05)	62.56 (1.13)
	Falling	209.04 (3.76)	−9.30 (1.78)	23.75 (1.79)	0.63 (0.02)	0.69 (0.05)	68.47 (1.64)

Mean values were calculated over all items in one linguistic category. For Mandarin, there were 23 items for level tone and rising tone, 25 for the dipping tone, and 27 for the falling tone. For Thai, there were five items in each category. SD are indicated in brackets. ST, semitones; arb dB, arbitrary dB units.

Method of Experiment 1). For the Thai tones, we calculated acoustic distances using the average values over the five tokens because the specific associations of tokens for tones varied across participants. Consequently, distances were calculated for different pairs only, as the average distance in the same pairs would be 0.

Link with behavioral performance

To investigate whether acoustic distances correlated with participants' performance, we calculated mean performance (percentage correct) for the 24 same pairs and the 25 different pairs with Mandarin tones and for the 10 different pairs with Thai tones in each set of material (verbal or musical). Mean performance over item pairs was then correlated with mean distances for each acoustic descriptor.

For Mandarin "same pairs," negative correlations were observed in amusics for mean pitch [$r(23) = -0.44$, $p < 0.05$] and pitch duration [$r(23) = -0.44$, $p < 0.05$]: the larger the changes in pitch and in pitch duration, the lower their performance. This suggests that amusics detected some changes in mean pitch and pitch duration, and that these changes led them to respond "different" and thus to err. In contrast to amusics, controls showed such a "distraction" effect only for the absolute slope [$r(23) = -0.49$, $p < 0.05$]. For Mandarin "different pairs," no correlations were significant.

For Thai "different pairs," positive correlations (ps < 0.05) were observed in both amusics and controls for distances in mean pitch [amusics: $r(8) = 0.79$ (verbal), $r(8) = 0.74$ (musical); controls: $r(8) = 0.80$ (verbal only)] and in absolute slope [amusics: $r(8) = 0.64$; controls: $r(8) = 0.75$, both for musical materials]: the larger the changes in pitch and in absolute slope, the higher participants' performance for the different pairs. Finally, it is interesting to point out that the slope seems to have been used more strongly by controls than amusics in the musical material [$r(8) = 0.54$ for

controls; $r(8) = 0.09$ for amusics], although this correlation in controls did not reach significance, which might be related to controls' overall high performance level.

In sum, these correlational analyses suggest that amusics' performance was influenced by changes in pitch parameters, notably mean pitch or absolute slope (i.e., pitch movement). While this information correlated positively and thus facilitated performance for the Thai material, which was based on the same CV syllable, it was misleading for the Mandarin material. In addition, the amusic participants also erroneously used the duration of the pitch information in Mandarin, which resulted in attenuated performance. As Mandarin tones do differ considerably in duration, it appears that the amusics focused on this cue (given their impaired sensitivity to pitch), but that this cue was ultimately misleading.

To further investigate the link between acoustic distances and behavioral performance, we performed an additional analysis for the Thai material (both verbal and musical). For this, we related for each participant the acoustic distances to performance differences as follows: we selected among the 10 different tone pairs, the three pairs with the largest acoustic distance and the three pairs with the smallest acoustic distance, for each of the descriptors. For these subgroups of test items, we computed mean performance (in Hits–FAs) for each participant. **Figure 5** represents the performance difference between stimulus pairs with large and with small acoustic distances. Positive differences represent better performance when the items of a pair differing markedly on the acoustic descriptor. Negative differences represent better performance when the items of a pair differed only weakly on the acoustic descriptor. Two-tailed t-tests on amusics' data showed that the observed differences were significantly different from 0 for all predictors and both the verbal and the musical materials (ps < 0.05), except for pitch duration in both speech and musical materials (ps > 0.42) and for RMS in the musical material ($p = 0.11$).

FIGURE 5 | Performance differences of amusic participants (left) and the control participants (right) for the Thai material (speech and musical analogs) between trials selected for large acoustic differences and trials selected for small acoustic differences. Positive values represent better performance when the items of a pair differed strongly on a given acoustic descriptor. Negative values represent better performance when the items of a pair differed only weakly on a given descriptor. Note that controls' data were presented here only for the sake of completeness as control participants reached relatively high performance levels, thus leaving little room for measuring changes as a function of acoustic differences.

The difference graph (**Figure 5**) shows that amusics benefited from large differences in mean pitch and in absolute slope, but less so from differences in slope. This might be due to amusics' previously described difficulty in using pitch direction information (see Foxton et al., 2004). For total sound duration and for intensity, however, smaller differences led to better performance (or larger differences led to worse performance). When the two items in a pair differed markedly on these features (i.e., these were misleading cues), amusics seemed to use this rather irrelevant information and thereby perform less accurately. When, however, the two items differed only minimally on these features, the strategy to base judgments on these cues (duration, intensity) was less disruptive. In such cases, amusics might then also consider other features that might be more difficult for them to perceive (but that are more relevant, such as pitch), leading to better performance.

GENERAL DISCUSSION

The present study investigated pitch perception in congenital amusia for tone-language materials (Experiments 1 and 2) and musical analogs (Experiment 2). Our goal was to investigate whether the previously described musical pitch-processing deficit of congenital amusics might also impair pitch processing in tone-language speech signals. Overall, our findings suggest that this is indeed the case: amusics showed impaired performance for lexical tones. This finding is consistent with recent studies showing mild deficits of amusics in speech intonation discrimination, identification, and imitation in their native language, notably British English (Liu et al., 2010) and Mandarin Chinese (Jiang et al., 2010). In Mandarin Chinese, Nan et al. (2010) further showed that a subgroup of native speakers who are amusic shows impaired identification of lexical tones.

Our results show that amusics' pitch deficit can extend to non-native, meaningless speech material, that is when no semantic content might distract listeners' attentional focus away from the task-relevant pitch information. The use of non-native speech material (Mandarin, Thai) also allowed us to investigate French native speakers (who are amusic) with natural speech material that required the processing of pitch changes smaller than those used in previous intonation tasks (Patel et al., 2008; Liu et al., 2010). The comparison of the observed deficits for Mandarin and Thai materials, respectively, further shows the impact of the size of the to-be-processed pitch changes on amusics' performance (see below). In addition, it is worth noting that the deficit was here observed with an experimental task that required the comparison of two syllables rather than two sentences (as in Patel et al., 2005; Liu et al., 2010), thereby decreasing memory load. Our results further reveal that amusics' deficit is less important for speech than for music, at least in the most severe cases of amusia. Combined with the acoustic analyses, the present study provides new insights into the nature of the pitch-processing deficit exhibited by amusics as set out below.

Amusics' pitch deficit was first documented with musical material (thus leading to the term "amusia"). Subsequent tests of the consequences of this pitch deficit on speech processing have implications for the understanding of the overall phenomenon of what has been labeled "amusia": the pitch deficit does not seem to be domain-specific and restricted to musical material, but is rather a domain-general deficit that was first discovered in a musical setting. Indeed, the findings show that a pitch deficit can be observed in speech material, even though it is not systematic (not all amusics show the deficit), relatively mild (both quantitatively and qualitatively small), and is not as pronounced as in musical analogs.

Experiments 1 and 2 revealed that congenital amusics, as a group, showed deficits in tasks requiring discrimination of Mandarin and Thai tones, respectively. This deficit also correlated with amusics' deficit in a musical task that required interval processing between tones (i.e., a subtest of the MBEA). The additional analyses of the Thai tone pairs further showed that amusics performed only 9% below controls for the contour–contour pair (the rising versus falling tones), but dropped to 33% below controls for other pairs requiring finer pitch contour discriminations (averaged over verbal and musical materials; **Table 2**).

In Experiment 2 with the Thai tones, the comparison between speech material and musical analogs suggests that in the presence of a severe musical pitch deficit, pitch information might be slightly better perceived in speech materials. Speech might thus enhance pitch processing in amusics, even if it does not restore normal processing. As discussed in Tillmann et al. (submitted) for discrete pitch changes, it remains to be investigated whether this boost for verbal material might be due to acoustic features of the speech sound that facilitate pitch extraction, or whether top-down influences come into play to modulate pitch-processing depending on material type (speech, music). According to the latter view, which suggests influences related to strategies, attention or memory, pitch extraction of tones in congenital amusia might not be the sole impairment, but rather the deficit may also incorporate later processing stages and be related to material-specific top-down processes.

Comparisons between Experiments 1 and 2 suggested that the amusics' deficit affected more strongly the processing of Thai tones than of Mandarin tones: for Thai, 63% of the amusics performed 2 SD below the controls' mean, whereas this was only the case of 15% of the amusics for Mandarin. The greater difficulty level of the Thai material might be linked to the set of five tones and the acoustic features of the to-be-detected changes, which were less varied in the Thai than Mandarin stimuli. We further acknowledge that the difference in performance for the two experiments might be exacerbated by the longer delay between the two items for the Thai tone pairs, which may have selectively influenced amusics' performance due to their pitch memory deficit (Gosselin et al., 2009; Tillmann et al., 2009; Williamson et al., 2010). Furthermore, the discrimination of Mandarin tones (but less so of Thai tones) might also be partly based on cues other than pitch, such as length and intensity contours (Whalen and Xu, 1992; Hallé et al., 2004), even if those cues are less reliable, and as was found here, misleading.

Even though the group of amusics performed below the level of the group of controls, our findings showed some overlap in performance between the groups (see **Figure 3**). This overlap might reflect the influence of large individual differences reported in the normal population, notably for the perception of pitch change and pitch direction (Semal and Demany, 2006; Foxton et al., 2009) as well as for the learning of pitch contours in syllables (Golestani and Zatorre, 2004; Wong and Perrachione, 2007). Some of these previous findings also suggest that pitch processing/learning is not independent between music and speech: for example, Wong and Perrachione (2007) reported an association between-participants' ability to learn pitch patterns in syllables and their ability to perceive pitch patterns in

non-lexical contexts as well as with their previous musical experience. Finally, in light of the observed variability in learning, Golestani and Zatorre (2004) discussed a "connectivity hypothesis," notably that fast and slow learners might differ in white matter connectivity, with greater myelination leading to more rapid neural transmission. This suggestion, along with our present findings, even if based on behavior only, might be related to recent data in congenital amusics having neural anomalies in white matter concentration, cortical thickness and fiber tracts in the right hemisphere (Hyde et al., 2006, 2007; Loui et al., 2009), and in functional connectivity between the auditory and inferior frontal gyrus (Hyde et al., 2011). Previous brain imaging research has suggested that pitch processing in tone-languages involves a right-hemisphere network rather than a left-hemisphere network for non-native listeners (e.g., Hsieh et al., 2001). This might explain the overall deficit for the amusic group observed here for the tone-language material (all non-native listeners). However, other brain imaging data have suggested the implication of left-hemisphere networks in speech processing independently of language background, but rather depending on the acoustic features of verbal sounds (see Zatorre and Gandour, 2007 for a review). In our study, the observed advantage of Thai tones over non-verbal, musical analogs in the amusics who exhibited higher pitch thresholds might suggest some implication of the left-hemispheric network, which seems to be unimpaired in congenital amusia.

Aiming to further understand amusics' performance and deficits, we calculated a series of acoustic measures and conducted analyses to investigate their relation to the behavioral data. These analyses provide some interesting insights into features related to amusics' discrimination performance. It was found that that amusics' tone discrimination is related not only to some pitch characteristics of the sounds, but also to features that are unrelated to pitch changes or pitch movement. Regarding first the pitch-related features, the analyses of the Thai material (**Figure 5**) suggested that small differences in FØ mean and absolute slope are associated with more difficult item pairs: amusics benefited from large differences in FØ mean and absolute slope (i.e., mean of the absolute value of the slope of the pitch contour), but this was less the case for slope (i.e., providing information about the direction of the pitch movement). This finding is in agreement with previous observations suggesting that amusics can detect pitch movement, but that they have difficulties in perceiving the direction of these movements (e.g., Foxton et al., 2004; Patel et al., 2008). Another pitch-related feature that seemed to be used by the amusics for the Mandarin materials is the duration over which pitch information was present in the tones but here, larger differences in pitch duration was a misleading cue; it led to more rather than less errors.

The acoustic analyses further showed that, in contrast to control participants, amusics seemed to use non-pitch-related information, notably sound duration and intensity, which are less relevant for lexical tone discrimination. The result pattern suggests that amusics use these acoustic features as some kind of replacement strategy: as amusics are less efficient in using pitch cues, they focus on these cues, whose variations they *can* discriminate well and this erroneously results in lower performance levels. When, however, items of a pair differ less strongly on these non-pitch features, amusics seem able to use other cues (probably also pitch-related cues), leading to higher performance levels.

In sum, the various acoustic analyses suggest that same-different task performance might not be based on single features, but that participants (in particular controls who have better performance on the task) used a weighted combination of acoustic features. Relative weighting of acoustic cues in lexical tone perception has been previously shown for native and non-native speakers (for example, for Yoruba and Thai, see Gandour and Harshman, 1978). Amusic individuals might (a) use the various pitch features less efficiently (because of their pitch deficit), (b) not weight/combine them in the adequate way (or also under-use some components), and (c) get waylaid by other, irrelevant but, for them, more easily discriminable cues (duration, intensity). Based on our findings, future experiments can now directly (i.e., parametrically) manipulate the various pitch and non-pitch cues in extent as well as in combination to further investigate the importance of these parameters, their relative weighting, as well as the deficits of amusics in processing these parameters.

This study provides evidence that the pitch deficit of congenital amusics previously observed for musical material extends to speech material, notably lexical tones of Mandarin and Thai. Our findings provide further motivation to investigate processing of pitch in amusics' native language (e.g., non-tonal languages). Specifically, future studies should focus not only on statements/questions, as has been done up to now, but investigate also prosody-based perception of emotion and humor, as well as the use of subtle pitch cues. Regarding this latter aspect, pitch variations can make a syllable more salient, create focus stress in syllables or allow for differentiation between, for example, "a hot dog" and "a hotdog." For example, Spinelli et al. (2010) tested the influence of FØ changes in the first vowel for word segmentation in French (e.g., "la fiche" versus "l'affiche"). Future research should investigate the use of these subtle pitch cues in amusics' native language perception (e.g., English, French), and even extend this to language production. Finally, our results further suggest that congenital amusics may experience difficulties in acquiring a tonal language. Actually, this possible relation between amusia and tone-language processing has been recently confirmed by Nan et al. (2010) who reported cases of congenital amusia among native speakers of Mandarin. Importantly, the musical deficit was found to be associated to impairments in lexical tone discrimination and identification, but not production – a finding that mirrors the previously observed mismatch between perception and production of musical intervals in congenital amusia (Loui et al., 2008).

REFERENCES

Abramson, A. (1978). Static and dynamic acoustic cues in distinctive tones. *Lang. Speech* 21, 319–325.

Abramson, A. S. (1962). *The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments.* Bloomington, IN: Indiana University Research Center in Anthropology, Folklore and Linguistics, 20.

Alexander, J. A., Wong, P. C. M., and Bradlow, A. R. (2005). "Lexical tone perception in musicians and non-musicians," in *Proceedings of Interspeech 2005*, Lisbon.

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

Bent, T., Bradlow, A. R., and Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 97–103.

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

Burnham, D., and Brooker, R. (2002). "Absolute pitch and lexical tones: tone perception by non-musician, musician, and absolute pitch non-tonal language speakers," in *7th International Conference on Spoken Language Processing*, Denver, 257–260.

Burnham, D., and Francis, E. (1997). "The role of linguistic experience in the perception of Thai tones," in *Southeast Asian Linguistic Studies in Honour of Vichin Panupong* (Science of Language, Vol. 8.), ed. A. S. Abramson (Bangkok: Chulalongkorn University Press), 29–47.

Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F., and Keller, P. (1996). "Perception of lexical tone across languages: evidence for a linguistic mode of processing," in *Fourth International Conference on Spoken Language Processing (ICSLP'96)*, Philadelphia.

Cohen, J., MacWhinney, B., Flatt, M., and Provost, J. (1993). PsyScope: an interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behav. Res. Methods Instrum. Comput.* 25, 257–271.

Cutler, A. (2009). Greater sensitivity to prosodic goodness in non-native than in native listeners. *J. Acoust. Soc. Am.* 125, 3522–3525.

Delogu, F., Lampis, G., and Belardinelli, M. O. (2010). From melody to lexical tone: musical ability enhances specific aspects of foreign language perception. *Eur. J. Cogn. Psychol.* 22, 46–61.

Deutsch, D., Dooley, K., Henthorn, T., and Head, B. (2009). Absolute pitch among students in an American music conservatory: association with tone language fluency. *J. Acoust. Soc. Am.* 125, 2398–2403.

Deutsch, D., Henthorn, T., Marvin, E., and Xu, H.-S. (2006). Absolute pitch among American and Chinese conservatory students: prevalence differences, and evidence for a speech-related critical period. *J. Acoust. Soc. Am.* 119, 719–722.

Drayna, D., Manichaikul, A., de Lange, M., Snieder, H., and Spector, T. (2001). Genetic correlates of musical pitch recognition in humans. *Science* 291, 1969–1972.

Fitzsimons, M., Sheahan, N., and Staunton, H. (2001). Gender and the integration of acoustic dimensions of prosody: implications for clinical studies. *Brain Lang.* 78, 94–108.

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

Foxton, J. M., Weisz, N., Bauchet-Lecaignard, F., Delpuech, C., and Bertrand, O. (2009). The neural bases underlying pitch processing difficulties. *Neuroimage* 45, 1305–1313.

Fromkin, V. (1978). *Tone: A Linguistic Survey.* New York: Academic.

Gandour, J., and Harshman, R. (1978). Cross-language differences in tone perception: a multidimensional scaling investigation. *Lang. Speech* 21, 1–33.

Golestani, N., and Zatorre, R. J. (2004). Learning new sounds of speech: reallocation of neural substrates. *Neuroimage* 21, 494–506.

Gosselin, N., Jolicoeur, P., and Peretz, I. (2009). Impaired memory for pitch in congenital amusia. *Ann. N. Y. Acad. Sci.* 1169, 270–272.

Hallé, P., Chang, Y.-C., and Best, C. (2004). Categorical perception of Taiwan Mandarin Chinese tones by Chinese versus French native speakers. *J. Phon.* 32, 395–421.

Hove, M. J., Sutherland, M. E., and Krumhansl, C. L. (2010). Ethnicity effects in relative pitch. *Psychon. Bull. Rev.* 17, 310–316.

Hsieh, L., Gandour, J., Wong, D., and Hutchins, G. D. (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain Lang.* 76, 227–252.

Hyde, K., Zatorre, R., and Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cereb. Cortex* 21, 292–299.

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., and Peretz, I. (2007). Cortical thickness in congenital amusia: when less is better than more. *J. Neurosci.* 27, 13028–13032.

Hyde, K. L., and Peretz, I. (2004). Brains that are out of tune but in time. *Psychol. Sci.* 15, 356–360.

Hyde, K. L., Zatorre, R. J., Griffiths, T. D., and Peretz, I. (2006). Morphometry of the amusic brain: a two-site study. *Brain* 129, 2562–2570.

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

Kawahara, H., Masuda-Katsuse, I., and Cheveigne, A. D. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous frequency-based f0 extraction. *Speech Commun.* 27, 187–207.

Klein, D., Zatorre, R. J., Milner, B., and Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *Neuroimage* 13, 646–653.

Krishnan, A., Xu, Y., Gandour, J., and Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res. Cogn. Brain Res.* 25, 161–168.

Lee, C.-Y., and Hung, T.-H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *J. Acoust. Soc. Am.* 124, 325–3248.

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

Loui, P., Alsop, D., and Schlaug, G. (2009). Tone deafness: a new disconnection syndrome? *J. Neurosci.* 29, 10215–10220.

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

Magne, C., Schön, D., and Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *J. Cogn. Neurosci.* 18, 199–211.

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

Nguyen, S., Tillmann, B., Gosselin, N., and Peretz, I. (2009). Tonal language processing in congenital amusia. *Ann. N. Y. Acad. Sci.* 1169, 490–493.

Patel, A. D. (2008). *Music, Language, and the Brain.* New York, NY: Oxford University Press.

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

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

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

Peretz, I., Champod, S., and Hyde, K. (2003). Varieties of musical disorders: the Montreal battery of evaluation of amusia. *Ann. N. Y. Acad. Sci.* 999, 58–75.

Peretz, I., Cummings, S., and Dubé, M. P. (2007). The genetics of congenital amusia (tone deafness): a family-aggregation study. *Am. J. Hum. Genet.* 81, 582–588.

Peretz, I., and Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends Cogn. Sci. (Regul. Ed.)* 7, 362–367.

Pfordresher, P. Q., and Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Atten. Percept. Psychophys.* 71, 1385–1398.

Schneider, W., Eschmann, A., and Zuccolotto, A. (2002). *E-Prime User's Guide.* Pittsburgh, PA: Psychology Software Tools, Inc.

Schön, D., Magne, C., and Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41, 341–349.

Schwanhäußer, B., and Burnham, D. (2005). "Lexical tone and pitch perception in tone and non-tone languages speakers," in *Proceedings of the 9th European Conference on Speech Communication and Technology* (Bonn: ISCA), 1701–1704.

Semal, C., and Demany, L. (2006). Individual differences in the sensitivity to pitch direction. *J. Acoust. Soc. Am.* 120, 3907–3915.

Spinelli, E., Grimault, N., Meunier, F., and Welby, P. (2010). An intonational cue to word segmentation in phonemically identical sequences. *Atten. Percept. Psychophys.* 2, 775–787.

Tillmann, B., Schulze, K., and Foxton, J. (2009). Congenital amusia: a short-term memory deficit for nonverbal, but not verbal sounds. *Brain Cogn.* 71, 259–264.

Vos, P. G., and Troost, J. M. (1989). Ascending and descending melodic intervals: statistical findings and their perceptual relevance. *Music Percept.* 6, 383–396.

Whalen, D. H., and Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments, *Phonetica* 49, 25–47.

Williamson, V., McDonald, C., Deutsch, D., Griffiths, T., and Stewart, L. (2010). Faster decline of pitch memory over time in congenital amusia. *Adv. Cogn. Psychol.* 6, 15–22.

Wong, P. C. M., and Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Appl. Psycholinguist.* 28, 565–585.

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422.

Yip, M. (2002). *Tone.* Cambridge: Cambridge University Press.

Zatorre, R. J., and Gandour, J. T. (2007). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1087–1104.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.