https://doi.org/10.1016/j.neuropsychologia.2017.01.033

Neural bases of congenital amusia in tonal language speakers

Caicai Zhang[a,b,*], Gang Peng[a,b,*], Jing Shao[a], and William S-Y. Wang[a,b]

[a] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China

[b] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, 518055

Corresponding authors *:

Gang Peng: Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China. Tel: (+852) 3400 8462. E-mail address: gpeng@polyu.edu.hk.

Caicai Zhang: Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China. Tel: (+852) 3400 8465. Email address: caicai.zhang@polyu.edu.hk.

Abstract

Congenital amusia is a lifelong neurodevelopmental disorder of fine-grained pitch processing. In this fMRI study, we examined the neural bases of congenial amusia in speakers of a tonal language – Cantonese. Previous studies on non-tonal language speakers suggest that the neural deficits of congenital amusia lie in the music-selective neural circuitry in the right inferior frontal gyrus (IFG). However, it is unclear whether this finding can generalize to congenital amusics in tonal languages. Tonal language experience has been reported to shape the neural processing of pitch, which raises the question of how tonal language experience affects the neural bases of congenital amusia. To investigate this question, we examined the neural circuitries sub-serving the processing of relative pitch interval in pitch-matched Cantonese level tone and musical stimuli in 11 Cantonese-speaking amusics and 11 musically intact controls. Cantonese-speaking amusics exhibited abnormal brain activities in a widely distributed neural network during the processing of lexical tone and musical stimuli. Whereas the controls exhibited significant activation in the right superior temporal gyrus (STG) in the lexical tone condition and in the cerebellum regardless of the lexical tone and music conditions, no activation was found in the amusics in those regions, which likely reflects a dysfunctional neural mechanism of relative pitch processing in the amusics. Furthermore, the amusics showed abnormally strong activation of the right middle frontal gyrus and precuneus when the pitch stimuli were repeated, which presumably reflect deficits of attending to repeated pitch stimuli or encoding them into working memory. No significant group difference was found in the right IFG in either the whole-brain analysis or region-of-interest analysis. These findings imply that the neural deficits in tonal language speakers might differ from those in non-tonal language speakers, and overlap partly with the neural circuitries of lexical tone processing (e.g. right STG).

Keywords: neural bases, congenital amusia, lexical tone, music, Cantonese, fMRI

1. Introduction

Congenital amusia is a lifelong neurodevelopmental disorder that influences musical pitch processing (Ayotte et al. 2002; Peretz et al. 2002; Hyde and Peretz, 2003, 2004; Foxton et al. 2004). It is estimated to affect about 3-4% of the population (Peretz et al. 2008). Earlier studies suggest that this disorder is music-specific, leaving pitch processing in language intact (Ayotte et al. 2002; Peretz and Hyde, 2003). However, recent studies with more refined design found that amusia does affect pitch processing in language (Patel et al. 2005; Nguyen et al. 2009; Liu et al. 2010; Tillmann et al. 2011a, 2011b). Among non-tonal language speakers, the amusics were found to be impaired in intonation processing when pitch differences in speech stimuli were controlled and reduced (Liu et al., 2010). The amusics also had difficulty with accurate discrimination of non-native lexical tones (Nguyen et al. 2009; Tillmann et al. 2011a). Among tonal language speakers, those with congenital amusia exhibited inferior performance in the perception of native tones as well as nonnative tones (Nan et al. 2010; Jiang et al. 2012b; Wang and Peng, 2014; Liu et al. 2013, 2015; Huang et al. 2015a,b). For instance, Nan et al. (2010) found that Mandarin-speaking amusics performed worse than musically intact controls in the identification and discrimination of Mandarin tones, though there was individual variation in the severity of lexical tone impairment. Furthermore, there is some evidence that the deficiency in tonal language speakers is not confined to auditory pitch processing, but prevails to higher-level phonological processing, impeding the categorical perception of lexical tone (Jiang et al. 2012b; Huang et al. 2015a). All these findings suggest that the musical pitch deficit prevails to pitch processing in language.

In spite of the ample behavioral evidence of pitch impairment in music and speech, however, the neural bases of congenital amusia remain unclear, especially in different language populations. Understanding the neural bases of congenital amusia is important, not only for shedding light on the nature of congenital amusia, but also for guiding the intervention. In non-tonal languages, although one study has found deficient pitch processing in the auditory cortices in congenital amusia (Albouy et al., 2013), the majority of neuroimaging studies have found that the auditory cortices of the amusics respond normally to pitch, especially in implicit pitch processing tasks (Omigie et al. 2013; Peretz et al. 2005, 2009; Hyde et al. 2011; Moreau et al. 2013; Norman-Haignere et al. 2016). For instance, in a functional MRI (fMRI) study, Hyde et al. (2011) found that the auditory cortices of the amusics responded normally to unattended pitch changes in a sequence of pitch stimuli. Instead, the functional deficits are localized in the music-selective neural circuitry in the right inferior frontal gyrus (IFG), which is believed to support musical pitch encoding and pitch memory (Zatorre et al. 1994; Holcomb et al. 1998; Griffiths et al. 1999). Furthermore, the right IFG showed anatomical abnormality in terms of the white and grey matter concentration in the amusic brain (Hyde et al. 2006; 2007; Albouy et al. 2013). Functional and structural connectivity between the right IFG and right auditory cortex is also disrupted (Hyde et al. 2011; Loui et al. 2009; Albouy et al. 2013).

The aforementioned studies have provided important data regarding the neural bases of congenital amusia in non-tonal language speakers. However, it remains unclear whether these findings can generalize to congenital amusia in different language populations, or might be under the influence of (non-tonal) language experience to some extent. In particular, it is worth examining the neural bases of congenital amusia in tonal language speakers. Long-term experience of a tonal language has been widely demonstrated to shape the neural processing of pitch (Bidelman et al. 2011; Bidelman et al. 2013; Tong et al. 2005; Luo et al. 2006; Gandour et al. 2002; 2004; Gu et al. 2013). In tonal languages, pitch is systematically used to convey information, like in music. Probably because of the overlap in pitch usage, cross-domain transfer between music and tonal language experience in pitch processing has been widely report at both behavioral and neural levels (Bidelman et al. 2011; Bidelman et al. 2013; Pfordresher and Brown, 2009; Deutsch et al. 2006; Peng et al. 2013; Wong et al. 2012; Alexander et al. 2005; Delogu et al. 2006; 2010; Lee and Hung, 2008; Lee et al. 2014; Smayda et al. 2015). Furthermore, as mentioned before, tonal language speakers with congenital amusia have been found to exhibit degraded performance in lexical tone perception (Nan et al. 2010; Jiang et al. 2012b; Wang and Peng, 2014; Liu et al. 2013, 2015; Huang et al. 2015a,b), which also suggests that musical processing is intimately linked with lexical tone processing. Such cross-domain transfer hints at possibly shared neural pathways of musical and lexical tone processing in tonal language speakers with congenital amusia. However, this hypothesis has not been systematically examined at the neurobiological level before. As the majority of previous neuroimaging studies focused on non-tonal language speakers, little is known about the neural bases of amusia in tonal language speakers.

To fill in the gap, in the current fMRI study, we examined the effect of tonal language experience on the neural circuitries of pitch processing in lexical tone and musical stimuli in Cantonese speakers with congenital amusia. We hypothesize that the neural circuitries sub-serving pitch processing in lexical tone and music are commonly impaired in Cantonese-speaking congenital amusics. In particular, we aim to examine whether the neural deficits in tonal language speakers lie in music-selective neural circuitry in the right IFG as in non-tonal language speakers, or overlap with the neural circuitries of lexical tone processing (cf. Tong et al. 2005; Luo et al. 2006; Gandour et al. 2002; 2004; Wong et al. 2004; Li et al. 2010; Gu et al. 2013; Zhang et al. 2016). We chose Cantonese in this study because it has three level tones that contrast a relatively flat pitch contour at various pitch height, which can be matched with musical notes in pitch (see Fig. 1).

We adopted a *group* (amusic and normal) × *domain* (speech and music) × *pitch interval* (repetition, fixed pitch interval, and varied pitch interval) design to examine the neural bases of relative pitch perception, a fundamental human perceptual ability in both lexical tone and musical perception (Saffran and Griepentrog, 2001; Itoh et al. 2005). It has been found that the absolute pitch height of a tone varies dramatically in the productions of speakers with different pitch ranges, and thus cannot be a reliable index of the tone category perceptually (Peng et al. 2012; Zhang et al. 2012, 2013). On the other hand, the pitch location of a tone within a speaker's pitch range is relatively constant across speakers (e.g. Wong and Diehl, 2003). For example, a high tone tends to be located close to the upper bound of a speaker's pitch range, while a low tone tends to be located close to the lower bound of a speaker's pitch range, no matter whether this speaker speaks with a high or low pitch range (Peng et al. 2012).

It has been found that tonal language speakers rely on the relative pitch height of a tone with reference to a speaker's pitch range in lexical tone perception (Zhang et al. 2012, 2013; Zhang and Chen, 2016). Similarly, relative pitch relationship is critical for musical perception. A musical melody can be perceived as the same melody when presented at different keys, presumably because the relative pitch intervals between notes are constant across keys. Recognizing the same melody presented at different keys is to some extent analogous to recognizing the same tone produced by speakers with different pitch ranges, both of which rely on the constancy of the relative pitch relationship between notes/tones.

We used an adaptation paradigm (e.g. Celsis et al. 1999; Chandrasekaran et al. 2011; Joanisse et al. 2007; Zhang et al. 2016) to examine the neural activities underlying relative pitch processing in lexical tone and musical stimuli, via a comparison of three conditions – repetition, fixed interval, and varied interval. The repetition condition, where a pair of tone/music sounds was simply repeated eight times in a block, served as the control condition. The fixed interval condition included eight pairs of tone/music sounds with repeated pitch interval at various pitch height in a block. The pitch interval was identical between the fixed interval and repetition condition, but the pitch height was different. The varied interval condition included eight pairs of tone/music sounds with varied pitch intervals at various pitch height in a block. Repetition of the pitch interval in the repetition and fixed interval conditions is expected to habituate the Blood Oxygenation Level Dependent (BOLD) signal in brain regions sensitive to relative pitch processing. The varied interval condition would result in a release from habituation, increasing the BOLD signal in those same regions. It has been found that implicit lexical tonal changes presented in an adaptation paradigm activated the superior temporal gyrus (STG) bilaterally in Cantonese speakers, among other brain regions (Zhang et al. 2016). While the experimental design of the previous study was different from the current study to a certain extent, it provides some information of the neural bases of lexical tone processing in Cantonese speakers. Based on the previous study, it is possible that the processing of relative pitch intervals also activates the bilateral STG and/or other brain regions. Importantly, the amusics, who are less sensitive to refined pitch differences, might fail to detect the variations in pitch interval size in the varied interval condition (which were 1 semitone larger or smaller than the pitch interval size in the fixed interval and repetition block; see Stimuli below). If so, the amusics would show a strong habituation even in the varied interval condition, thus exhibiting less or no activation (i.e. less/no release from habituation) in the processing of relative pitch intervals in the bilateral STG and/or other brain regions compared to the controls. Apart from the bilateral STG, we also focus on examining whether Cantonese-speaking amusics exhibit deficient pitch processing in the right IFG as in non-tonal language speakers (cf. Hyde et al. 2011).

2. Material and methods

2.1 Participants

Eleven amusics and 11 musically intact controls matched one by one in age, gender, and years of education participated in the experiment. All participants were native speakers of Hong Kong Cantonese and university students in Hong Kong. They were all right-handed, and reported no hearing impairment, no history of neurological illness or musical training.

The amusic participants were identified using the Online Identification Test of Congenital Amusia (Peretz et al. 2008; http://www.brams.umontreal.ca/onlinetest), which has been used as a screening test in recent studies on congenital amusia (e.g. Liu et al, 2010; Zendel et al, 2015). All amusic participants scored 73 or lower, whereas the control participants scored 80 or higher in the global score of the test, which is the average of all three subtests – out-of-key, offbeat and mistuned subtests. Among the 11 amusic participants, nine of them even scored below 70 in the global score. Note that the score for selecting amusics (73) used in the current study was below the cut-off score (78.4) reported in a previous prevalence study of Cantonese-speaking amusics using the same test (Wong et al. 2012). A lower score was used in the current study partly because the Cantonese participants in the previous study (Wong et al. 2012) had a high level of musical training, which was likely to inflate the cut-off score for determining the amusics to some extent. Independent-samples t-test confirmed that the global score of the amusics was significantly lower than that of the controls ($t(20) = -14.337$, $p < 0.001$) in the current study. This result was further corroborated by the results of a behavioral study that examined the performance of Cantonese-speaking amusics selected using the same criterion (eight out of 11 amusics in the current study participated in the behavioral study) in lexical tone identification and discrimination, which showed that these amusics demonstrated an impairment in lexical tone perception, apart from the deficient musical perception performance (Shao et al. 2016). Demographic characteristics of the participants are summarized in Table 1. For the reason that the participants were only tested on the online amusia test, the possibility that some participants might have other cognitive deficits such as attentional deficits cannot be completely ruled out. The procedures of the fMRI experiment were approved by the Shenzhen Institutes of Advanced Technology Institutional Review Board. Informed written consent was obtained from participants in compliance with the experiment protocols.

2.2 Stimuli

The stimuli were comprised of speech stimuli generated from natural productions of Cantonese words contrasting the three level tones, and musical stimuli in the piano timbre that were matched in pitch with the speech stimuli. The three lexically contrastive level tones were high level tone (醫 /ji55/ ‘a doctor’), mid level tone (意 /ji33/ ‘meaning’), and low level tone (二 /ji22/ ‘second’). A female Cantonese speaker was recorded reading aloud the three target words minimally contrasting the three level tones (醫 /ji55/ ‘a doctor’, 意 /ji33/ ‘meaning’, 二 /ji22/ ‘second’) in a carrier sentence (呢個字係__ /li55 ko33 tsi22 hɐi22 __/ ‘This word is __’) for ten times. The F0 was measured at 21 sampling points of 5% intervals across the whole syllable using the Praat (Boersma and Weenick, 2014). The F0 trajectory of the three tones averaged from all repetitions is shown in Fig. 1. The three level tones were primarily distinguished by pitch height, despite some small and perceptually negligible fluctuation in the pitch contour (Peng et al. 2012). The high level tone /ji55/ was roughly 3 semitones higher than the mid level tone /ji33/, which was roughly 2 semitones higher than the low level tone /ji22/.

One clearly produced token of /ji33/ was selected for manipulation. It was normalized in duration to 500 ms and in average intensity to 80 dB using the Praat. Then the F0 trajectory of /ji33/ was raised by 3 semitones to create the /ji55/ stimulus, and lowered by 2 semitones to create the /ji22/ stimulus. Resynthesized stimuli based on /ji33/ instead of naturally produced /ji55/ and /ji22/ were used, in order to minimize acoustic differences between the three tones other than the F0 (Peng et al. 2012). This manipulation helps to ensure that any observed group differences between amusics and controls in the neural processing are not due to different processing abilities of duration, formant frequency or other acoustic cues, other than the F0.

The musical stimuli were 500-ms long piano tones generated using a Kurzweil K2000 synthesizer tuned to the standard A4 of 440 Hz (Peng et al. 2013). The F0 trajectories of the three musical notes (C♯4, A♯3, and G♯3) that were pitch-matched with the three level tones are shown in Fig. 1.

The three tonal/musical stimuli were grouped into six pairs (/55/-/33/, /33/-/55/, /33/-/22/, /22/-/33/, /55/-/22/, /22/-/55/), and each pair was presented in the three conditions – repetition, fixed interval, and varied interval (see Table 2). In the repetition condition, a tone/music pair was simply *repeated* eight times. The fixed interval condition included eight variants of the same tone/music pair, with *identical pitch interval* as in the repetition condition, but *varied pitch height*. Among the eight variants, four variants were 1-4 semitones higher in pitch height than the pair in the repetition condition, while the other four variants were 1-4 semitones lower. The varied interval condition included eight variants of the same tone/music pair with *varied pitch interval* and *varied pitch height*. Among the eight variants, four variants had pitch interval one semitone larger than the pair in the fixed interval condition, and the other four variants had pitch interval one semitone smaller. This was achieved by switching the second sound in the eight variants in the fixed interval condition to create varied pitch intervals for each tonal/musical pair (e.g., D♯4-C4 and D4-B3 in the fixed interval condition was paired as D♯4-B3 and D4-C4 in the varied interval condition).

In total, all tone pairs presented in the three conditions covered 14 speech stimuli, and the musical stimuli covered 14 pitch-matched musical notes in the piano timbre, ranging from E3 to F4 (see Table 2).

2.3 Procedure

The stimuli were presented in a block design. There were 36 blocks in total, including 18 speech blocks and 18 music blocks, which were further divided into equal numbers of repetition, fixed and varied interval conditions (six blocks per condition).

Each block was 20 seconds in length, followed by a 12 seconds rest block. In the first 16 seconds of a block, eight pairs of stimuli were randomly presented. Each pair of stimuli was one second in length, containing two stimuli of 500-ms each with no pause in between; neighboring pairs were separated by one second. After the stimulus presentation, a question appeared on the screen, instructing participants to judge whether the heard stimuli were speech or music within four seconds, by pressing buttons (1 or 2) with the left and right thumb respectively. The manual responses were counterbalanced, with half of participants making "speech" responses with left thumb and "music" responses with right thumb, and the other half of participants with switched left/right thumb responses. The task was only to ensure that the participants paid sustained attention to the auditory stimuli. Both amusics and controls were highly accurate in the speech/music judgment. Independent-samples t-tests found no significant difference between amusics and controls in either accuracy ($t(88.262) = -1.649$, $p = 0.101$; amusics: mean = 98.0%, SD = 6.6%; controls: mean = 99.5%, SD = 2.9%) or reaction time (RT) ($t(119.619) = 1.158$, $p = 0.249$; amusics: mean = 1045 ms, SD = 254; controls: mean = 984 ms, SD = 344). An example block of the musical stimuli in the repetition, fixed interval and varied interval conditions is displayed in Fig. 2. Examples of audio files of the speech and musical stimuli (using the pair /55/-/33/ as an example) in the repetition, fixed interval and varied interval conditions can be found in the supplementary materials.

For each participant, the 36 blocks were randomly divided into three runs, with 12 test blocks per run together with 12 rest blocks. Each run included six speech blocks and six music blocks (two test blocks for each of the repetition, fixed interval and varied interval conditions). The order of the 12 test blocks was randomized in a run. Each run lasted 384 seconds.

2.4 fMRI data acquisition

fMRI data were acquired using a 3T Magnetom TRIO Scanner (Siemens, Erlangen, Germany) equipped with a 12-channel phased array receive-only head coil at the Shenzhen Institutes of

Advanced Technology, the Chinese Academy of Sciences. 3D MPRAGE was applied to obtain continuous high-resolution T1-weighted anatomical images (scan repetition time (TR) = 2530 ms; echo time (TE) = 2.01 ms; flip angle (FA) = 7°; field of view (FOV) = 256 mm; in-plane resolution 1.0 mm × 1.0 mm × 1.0 mm; 192 slices in total). Functional gradient-echo planar images (EPI) were acquired (TR = 2000 ms; TE = 30 ms; FA = 90°; FOV = 220 mm; 4 mm slice thickness; no gap; 64 × 64 matrix; 32 slices) continuously in ascending interleaved axial slices.

2.5 fMRI data analysis

Three imaging runs, each containing 198 TRs, were obtained from each participant. Data analysis was performed using AFNI (Cox, 1996). The first six TRs of each run were disregarded from analysis. Images were aligned to the first TR, corrected for slice acquisition time, motion corrected using a six-parameter rigid body transform, and spatially smoothed with an 8 mm Gaussian filter. Images exceeding 3 mm displacement or 3° rotation measured in TR-to-TR change were discarded. The high-resolution anatomical scan for each participant was normalized to Talairach and Tournoux stereotaxic space using the Colin27 template. All data were transformed to this same space using a single concatenated transform from EPI to high-resolution anatomical to Colin27 template. Single-subject BOLD signals were scaled and submitted to a regression analysis with the idealized hemodynamic responses as regressors at each voxel, which were created by convolving the timing of a condition with a gamma function for each condition respectively. The six parameters from the motion-correction process were included as nuisance regressors, as were baseline, linear, and quadratic trend.

Regression coefficients from the single-subject analysis were input to the group-level analysis. Whole-brain and region-of-interest (ROI) analysis were conducted using the AFNI.

For the whole-brain analysis, a mixed-effects *group* × *domain* × *pitch interval* ANOVA analysis with subjects as a random factor was conducted at each voxel at a whole-brain level using 3dMVM of AFNI. Contrast maps were obtained for main and interaction effects of *group*, *domain* and *pitch interval*, with a focus on the main effects of *group* and the interaction between *group* and other factors. Additionally, in order to fully explore the differences between the amusics and controls in brain activities, we also examined the main effects of *group* in each level of *domain* (speech, music), *pitch interval* (repetition, fixed interval, varied interval) and *domain* × *pitch interval* condition (speech-repetition, speech-fixed interval, speech-varied interval, music-repetition, music-fixed interval, music-varied interval). More importantly, partly following the previous study (Hyde et al. 2011), *within-group* analyses of relative pitch processing were conducted on the data of the amusics and controls respectively, in order to examine the neural circuitries sub-serving the processing of relative pitch interval in each group, which is the main focus of this study. As mentioned above, repetition of pitch interval in the repetition and fixed interval conditions is expected to habituate the BOLD signal in brain regions sensitive to relative pitch interval processing, whereas the varied interval condition would result in a release from habituation, increasing the BOLD signal in those same regions. Thus those brain regions involved in relative pitch interval processing are expected to show the following pattern: large BOLD signal in the varied interval condition and comparably reduced BOLD signal in the repetition and fixed interval conditions (Varied interval > Fixed interval = Repetition). Using this criterion, contrast maps were obtained for speech and music conditions respectively for the amusic and control group. For all comparisons, statistic images were assessed for cluster-wise significance using a cluster-defining threshold calculated with 3dClustSim of AFNI: the uncorrected threshold was $p = 0.001$, and the FWE-corrected critical cluster size at the threshold of $p = 0.05$ was 36.2 voxels (NN level 1, faces must touch and 2-sided threshold).

As for the ROI analyses, we focused on the right IFG (pars orbitalis), where functional deficit was reported in non-tonal language speakers (Hyde et al. 2011). A 5 mm sphere was generated centering on the coordinates of the right IFG (MNI coordinates: x = 34, y = 32, z = 0) reported in Hyde et al. (2011), and the mean beta values of all voxels within the sphere were obtained. Note that spheres of other sizes (3-8 mm) have also been used, and the results were similar. ROI analysis was also conducted on the clusters significantly activated in the whole-brain analysis mentioned above. The mean beta values of all voxels within each mask/cluster were input to the *group* × *domain* × *interval* ANOVA analysis using SPSS (IBM Corp., 2011).

3. Results

All significant clusters are reported in Table 3. In the text below, we focused on reporting the results of *group* effects and within-group analyses of relative pitch processing. Significantly activated clusters and the mean beta values for each condition are shown in Fig. 3. Fig. 4 shows the ROI analysis of the right IFG.

For the mixed-effects *group* × *domain* × *pitch interval* analysis, the only condition that showed significant activations was the interaction between *group* and *interval*, where a cluster primarily located in the right precuneus was significantly activated. ROI analysis revealed that the group difference was significant only in the repetition condition ($t(42) = 2.133$, $p = 0.039$), where the mean beta value was significantly larger in the amusics (mean = 0.056, SD = 0.135) than in the controls (mean = −0.031, SD = 0.136). No other effects were significant.

As for the analyses of main effects of *group* in each level, a significant *group* effect was found in the repetition condition, where a cluster primarily located in the right middle frontal gyrus (MFG) was significantly activated. The mean beta values in the right MFG were significantly larger in the amusics (mean = 0.066, SD = 0.072) than in the controls (mean = −0.016, SD = 0.067). No other brain regions were significantly activated in other group analyses.

As for within-group analyses of relative pitch processing, two clusters were significantly activated in the control group, one cluster in the right STG in the speech condition, and the other cluster in the cerebellum in general. No significant activations were found in response to relative pitch processing in the amusics in these two regions or any other brain region. This indicates that only controls exhibited habituation and a release from habituation to the constancy of pitch interval specifically in the speech condition in the right STG, and habituation and a release from habituation to the constancy of pitch interval regardless of speech and music conditions in the cerebellum.

The ROI analysis was conducted on the mean beta values of the right STG cluster. Given that this cluster was activated specifically in the speech condition, *group* × *interval* ANOVA analysis was conducted on the mean beta values of the speech condition. There were a significant *group* by *interval* interaction ($F(2, 40) = 4.630$, $p = 0.016$), and a significant main effect of *interval* ($F(2, 40) = 7.727$, $p = 0.001$). Post-hoc analyses revealed a near-significant trend of differences in pitch interval processing in the controls ($F(2, 30) = 2.282$, $p = 0.120$), where the mean beta value was larger in the varied interval condition (mean = 0.707, SD = 0.231) than the fixed interval condition (mean = 0.545, SD = 0.234) and the repetition condition (mean = 0.514, SD = 0.217), while the latter two were quite similar to each other. No significant effect or such trend was found in the amusics, where the mean beta value was actually largest in the fixed interval condition (repetition: mean = 0.477, SD = 0.232; fixed interval: mean = 0.590, SD = 0.207; varied interval: mean = 0.549, SD = 0.215).

The ROI analysis on the cerebellum also confirmed that there were significant group differences in the processing of relative pitch interval. *Group* × *domain* × *interval* ANOVA analysis on the mean beta values of voxels in the cerebellum cluster revealed a significant *group* by *interval* interaction ($F(2, 40) = 10.184$, $p < 0.001$), and a significant main effect of *interval* ($F(2, 40) = 8.027$, $p = 0.001$). Post-hoc analyses showed that there was a significant main effect of *interval* in the controls only ($F(2, 63) = 5.146$, $p = 0.009$). For the controls, Bonferroni-corrected pairwise comparisons showed that the mean beta value was significantly stronger in the varied interval condition than the fixed interval condition ($p = 0.007$; varied interval: mean = 0.088, SD = 0.106; fixed interval: mean = −0.027, SD = 0.157), while there was no significant difference between the fixed interval and repetition condition (repetition: mean = 0.014, SD = 0.088). The main effect of *interval* was not significant in the amusics. Moreover, independent-samples t-tests showed that the group difference was significant in the fixed interval condition ($t(42) = 2.076$, $p = 0.044$), where the mean beta value was significantly stronger in the amusics (mean = 0.053, SD = 0.086) than the controls (mean = −0.027, SD = 0.157). No group difference was found in the other conditions. These results indicate that only controls exhibited habituation to repeated pitch intervals regardless of whether they were lexical tone or speech stimuli in the cerebellum.

Lastly, for the right IFG, no significant activations were found in the whole-brain analysis. ROI analysis on the mean beta values of a 5 mm sphere centered on the coordinates reported in Hyde et al. (2011) also failed to find a significant main effect of *group* or interaction effects of *group* with other factors in this region. The only significant effect revealed by the ROI analysis was the main effect of *domain* ($F(1, 20) = 4.656$, $p = 0.043$), where the mean beta value was stronger in the speech condition (mean = 0.046, SD = 0.142) than in the music condition (mean = 0.004, SD = 0.159) in the right IFG.

4. Discussion

In this fMRI study, we examined the neural deficits of pitch processing in lexical tone and musical stimuli in Cantonese speakers with congenital amusia and matched controls. We found that Cantonese-speaking amusics exhibited abnormal activities in a widely distributed neural network in music and lexical tone conditions. Whereas the controls exhibited significant activation in the right STG in the lexical tone condition and in the cerebellum regardless of the lexical tone and music conditions during the processing of relative pitch interval, no activation was found in the amusics in those regions. The amusics also demonstrated abnormally strong activation in response to the repetition of pitch stimuli in the right MFG and right precuneus.

4.1 Relative pitch processing deficits

Relative pitch perception is a fundamental human perceptual ability, which requires the extraction of constancy in relative pitch intervals from auditory stimuli. This ability is crucial for speech and musical perception (e.g. Saffran and Griepentrog, 2001; Itoh et al. 2005). For lexical tone, the category of a tone is more reliably indexed by its pitch location relative to a speaker’s pitch range than its absolute pitch location, the latter of which varies substantially in the productions of different speakers (Peng et al. 2012; Zhang et al. 2012, 2013, 2016; Zhang and Chen, 2016). Similarly, a musical melody presented at different keys can be perceived as the same melody, presumably because relative pitch intervals between notes are constant across keys.

We found that the cerebellum responded to the constancy of relative pitch intervals in general, showing stronger activation in the varied interval condition than in the fixed interval condition (and similar activation between the fixed interval and repetition condition), regardless of speech or music conditions, in the control group only. This indicates that the cerebellum in the control participants' brain exhibited habituation by the constancy of pitch intervals in general, and release from habituation when such constancy was interrupted in the varied interval condition. On the contrary, no activation in the cerebellum was found in the amusics. It has been found that the cerebellum is involved in automatic recognition, motor learning, and language-related skills (Ito, 2000; Nicolson et al. 2001). For instance, dyslexia – a reading disorder with a core deficit in phonological processing (e.g., Ramus et al. 2013; Boets et al. 2013; Frost et al. 2009), has been associated with cerebellar impairment, among other reported neural deficits, with the cerebellar impairment found in about 80% of the cases (Nicolson et al. 2001). The cerebellum was also activated in the automatic detection of talker changes (which are indexed by pitch differences to a large extent) in the speech stimuli while listeners were doing a lexical tone judgment task (Zhang et al. 2016). Given the previously reported functional involvement of cerebellum, it is likely that a general mechanism for automatically detecting sound patterns is subserved by the cerebellum. More importantly, this general mechanism in the cerebellum appears to be dysfunctional during the detection of constancy in relative pitch intervals in the amusics.

While the cerebellum responded to the constancy of pitch intervals in a domain-general manner (regardless of speech and music condition), the right STG responded to pitch interval constancy in speech specifically. We found that the right STG was significantly activated in the speech condition in the control group only, whereas the amusic group exhibited no activation. The following ROI analyses also revealed a trend of habituation by the constancy of pitch intervals in the fixed interval condition and release from habituation in the varied interval condition in the control group only. The right STG has been found to be involved in the processing of spectral information (Zatorre et al. 1992; Zatorre and Belin, 2001). It is also part of the neural circuitries of lexical tone processing (Tong et al. 2005; Zhang et al. 2016). For instance, Zhang et al. (2016) found that the implicit processing of lexical tone changes entangled with talker changes in the speech stimuli activated the bilateral STG in Cantonese speakers. Thus the right STG probably plays a pivotal role in the processing of lexical tone with phonetic variation. Importantly, the neural function in the right STG appears to be impaired in the amusics during the implicit processing of constancy in pitch intervals.

4.2 Potential influence of language experience on the neural bases of congenital amusia

Most previous neuroimaging studies on non-tonal language speakers have found that pitch processing in the auditory cortices is normal in the congenital amusics, especially in implicit pitch processing tasks (Peretz et al. 2005, 2009; Hyde et al. 2011; Moreau et al. 2013; Lu et al. 2015; Norman-Haignere et al. 2016). For instance, Hyde et al. (2011) found that the bilateral auditory cortices of the amusics responded normally to pitch changes, showing increasing activation on a par with increasing pitch distance in a sequence of pitch stimuli. Consistent with this finding, a recent study reported that the amusics exhibited normal responses to pitch in pitch-responsive clusters in the auditory cortices, which are comparable in extent and anatomical location to those in the controls (Norman-Haignere et al. 2016). The functional deficit of the congenital amusia was found in the right IFG, which showed abnormal deactivation towards implicit pitch changes (Hyde et al. 2011). The right IFG also showed anatomical abnormality in terms of white and grey matter concentration in the amusics (Hyde et al. 2006, 2007; Albouy et al. 2013). Hyde et al. (2011) argued that the neural deficit in the right IFG is related to impairment in higher-level processing in musical pitch encoding and pitch memory in the amusics (Zatorre et al. 1994; Griffiths et al. 1999; Holcomb et al. 1998). The functional and structural connectivity between the right auditory cortex and the right IFG is also disrupted (Loui et al. 2009; Hyde et al. 2011; Albouy et al. 2013). Interestingly, although the auditory cortices appear to function normally in the amusical speakers of non-tonal languages, anatomical abnormality in the right auditory cortex has been reported (Hyde et al. 2006, 2007; Albouy et al. 2013).

While the above findings have provided important information on the neural bases of congenital amusia on non-tonal language speakers, the neural bases in tonal language speakers remain largely unknown. As mentioned before, cross-domain transfer between musical and tonal language experience in pitch processing has been widely reported (Bidelman et al. 2011; Bidelman et al. 2013; Pfordresher and Brown, 2009; Deutsch et al. 2006; Peng et al. 2013; Wong et al. 2012; Alexander et al. 2005; Delogu et al. 2006; 2010; Lee and Hung, 2008; Lee et al. 2014; Smayda et al. 2015), which raises the question of how long-term experience of a tonal language affects the neural bases of congenital amusia.

In the present study, we found that tonal language experience appears to affect the neural bases of congenital amusia to some extent, for the reason that the neural deficits in Cantonese-speaking amusics seem to be different from those reported in non-tonal language speakers. The differences between Cantonese-speaking amusics and amusical speakers of non-tonal languages are discussed in detail below.

First, we found that Cantonese-speaking amusics exhibited abnormal lack of activation during relative pitch processing in the right STG in the speech condition and in the cerebellum regardless of speech and music conditions. The finding of functional brain deficits in the right STG in particular seems to deviate from the aforementioned findings on non-tonal language speakers with congenital amusia, who usually exhibit normal pitch processing in the auditory cortices (Peretz et al. 2005, 2009; Hyde et al. 2011; Moreau et al. 2013; Lu et al. 2015; Norman-Haignere et al. 2016). It should be noted that the right STG is part of the neural network of lexical tone processing. It has been found that lexical tone processing recruits the bilateral STG (Tong et al. 2005; Luo et al. 2006; Zhang et al. 2016), amid a widely distributed neural network in frontal, temporal and parietal regions (Gandour et al. 2002; 2004; Wong et al. 2004; Li et al. 2010; Gu et al. 2013). Implicit processing of lexical tone changes also activates the bilateral STG in Cantonese speakers (Zhang et al. 2016). This suggests that the neural deficits of congenital amusia in Cantonese speakers might partly overlap with the neural network of lexical tone processing.

Second, we did not find neural deficit in the right IFG in either the whole-brain analysis or

ROI analysis. A possible explanation is that different functions of pitch in tonal and non-tonal languages might attune different neural circuitries of pitch processing in the amusics. Pitch, especially refined pitch differences, has a high functional load in Cantonese, systematically distinguishing word meanings in the lexicon. It is possible that the high linguistic functional load of pitch leads to more fragile STG in Cantonese-speaking amusics, which is prone to neural dysfunction. On the other hand, pitch processing in non-tonal languages is less linguistically relevant, and relies on neural circuitries of higher-level musical processing such as the right IFG (Zatorre et al. 1994; Holcomb et al. 1998; Griffiths et al. 1999). Thus congenital amusia in non-tonal language speakers is manifested neurobiologically as deficiency in the right IFG (Loui et al. 2009; Hyde et al. 2006; 2007; 2011; Albouy et al. 2013).

Consistent with this finding of deficient pitch processing in the right STG, previous studies have reported impaired auditory processing of pitch in tonal language speakers with congenital amusia (Nan et al. 2010, 2016; Jiang et al. 2012b; Wang and Peng, 2014; Liu et al. 2013, 2015; Huang et al. 2015a,b). Nan et al. (2016) found that pre-attentive auditory processing of lexical tone is impaired in Mandarin-speaking amusics, as indexed by the reduced mismatch negativity (MMN) response to lexical tone changes. The MMN is primarily generated by neural sources in bilateral auditory regions in the temporal lobe, in addition to secondary sources related to attention switch in the frontal lobe (Opitz et al. 2002; Rinne et al. 2000). This thus implies that auditory processing of lexical tone primarily sub-served by auditory temporal regions is impaired in Mandarin-speaking amusics. Apart from lexical tone processing, Mandarin-speaking amusics have been found to show abnormal N100 while listening to sentences with illegal prosody (Jiang et al. 2012). The N100 is believed to index auditory processing with neural sources in the primary auditory cortex (Krumbholz et al. 2003; Pantev et al. 1988). Altogether these findings suggest that pitch processing in auditory cortices is very likely to be impaired in tonal language speakers with congenital amusia.

Note that there are alternative explanations for the lack of neural deficits in the right IFG in tonal language speakers. One possibility is the discrepancy in experimental and stimulus design between the current study and the previous study (Hyde et al. 2011). For instance, in the previous study the stimuli were long melody-like sequences comprised of 21 pure tone stimuli. Long stimulus sequences might tax pitch encoding and pitch memory more, thus more likely to reveal potential functional deficit in musical processing in the right IFG. This question awaits further investigation in future studies. Moreover, future studies might examine the neural deficits of congenital amusics in tonal and non-tonal languages using the same experimental paradigm in one study.

4.3 Working memory and/or attention deficits

We found that the amusics demonstrated abnormally strong activation of the right MFG and the right precuneus in the repetition condition. The brain activations typically go down with stimulus repetition. However, the amusics exhibited abnormally strong activation with repeated pitch stimuli. This seems to suggest a deficit in attending to repeated pitch stimuli, or encoding repeated pitch stimuli into working memory in the amusics.

Previous studies have found that bottom-up sensory-driven attention is subserved by a ventral fronto-parietal network in the right hemisphere, including the tempo-parietal junction, MFG, IFG, frontal operculum and anterior insula (Chica et al. 2013). The right MFG and IFG are also involved in reorienting attention (Shulman et al. 2009; Japee et al. 2015). Moreover, volume reduction in the right MFG was associated with deficits in episodic retrieval in older adults (Rajah et al. 2011). As for the precuneus, it is involved in episodic memory (Lundstrom et al. 2003, 2005; Wallentin et al. 2006) and self-consciousness (Kjaer et al. 2002; Lou et al. 2004). Thus the abnormally strong activation of the right MFG and precuneus in the amusics might suggest deficits of attending to repeated pitch stimuli, or effortful encoding of repeated pitch stimuli. This result is also consistent with previous findings that the amusics are impaired in pitch encoding and retrieval in short-term memory (Gosselin et al. 2009; Marin et al. 2012; Albouy et al. 2013; Tillmann et al. 2009, in press).

5. Conclusion

To conclude, the current study is the first to examine the neurobiological bases of congenital amusia in tonal language speakers. Cantonese-speaking amusics exhibited abnormal lack of activation in the cerebellum and right STG, which likely reflects a dysfunctional mechanism of relative pitch processing, and abnormally strong activation of the right MFG and precuneus, which likely reflects working memory and/or attention deficits of repeated pitch stimuli. The neural deficits of congenital amusia in Cantonese speakers appear to differ from those in non-tonal language speakers, and partly overlap with the neural network of lexical tone processing in tonal language speakers (e.g. right STG). These findings provide some evidence for the impact of tonal language experience on the neural bases of congenital amusia, and have implications for the intervention for congenital amusia in tonal language speakers.

The current study also has some limitations that wait to be addressed by future studies. First, the sample size of congenital amusics is quite small (11 amusics). Future studies might further examine the neural deficits of congenital amusia in tonal languages using a larger sample of amusics. Second, in the current study the pitch intervals between two tone/musical stimuli in a pair were 2, 3, or 5 semitones respectively, which might be greater than the pitch threshold of some amusics (e.g. Jiang et al. 2011; Huang et al. 2015). Future studies might examine the neural deficits underlying refined pitch processing (e.g. pitch differences of 1 semitone or smaller) in tonal language speakers with congenital amusia. Third, we cannot rule out the possibility that some amusic participants in this study might have other cognitive deficits. The participants were only tested on the Online Identification Test of Congenital Amusia (Peretz et al. 2008), but not other cognitive tests. This might affect the brain activation results to some extent. Last, the task used in the fMRI experiment (judging whether the heard sequences were speech or music) does not require the participants to explicitly process/detect the pitch changes. It has been found that the amusics, especially those in non-tonal languages, typically show normal or nearly normal pitch processing in implicit pitch processing task (Omigie et al. 2013; Peretz et al. 2005, 2009; Hyde et al. 2011; Moreau et al. 2013). While we have found neural deficits of Cantonese-speaking amusics using the current task (also see Nan et al. 2016, who found impaired pre-attentive auditory processing of lexical tone in Mandarin-speaking amusics), future studies might consider using explicit pitch-processing tasks. This will shed further light on the neural bases of congenital amusia in tonal language speakers. Related to this point, please note that previous studies have generally reported bilateral activations with a left-hemisphere predominance for lexical tone processing in native tonal language speakers (Gandour et al. 2002; 2004; Wong et al. 2004; Li et al. 2010; Gu et al. 2013; van Lancker and Fromkin, 1973, 1978; Wang et al. 2001). Nonetheless, we found a right hemisphere lateralization (i.e. right STG) in the current study. This is probably related to the design of the current study. As mentioned before, the task used in the fMRI experiment does not require the participants to access the linguistic meaning or tone category of the stimuli. While the current design is useful for investigating the processing of relative pitch interval, future studies might consider using explicit tasks such as tone identification, which could reveal more about the linguistic processing of lexical tone in the tonal language speakers with congenital amusia.

Acknowledgements

This work was supported by the National Natural Science Foundation of China [NSFC: 11504400], Early Career Scheme [ECS: 25603916], the PolyU Start-up Fund for New Recruits, and the National Basic Research Program of the Ministry of Science and Technology of China [973 Grant: 2012CB720700].

We thank the anonymous reviewers for the constructive comments. We thank Prof. Patrick Wong and Dr. Gangyi Feng for useful suggestions, and Ms. Xiao Wang, Mr. Steve Ka-Hong Wong, and Mr. Yike Yang for help with data collection.

References

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P.E., et al., 2013. Impaired pitch perception and memory in congenital amusia: The deficit starts in the auditory cortex. Brain 136(5), 1639–1661.

Alexander, J.A., Wong, P.C.M., Bradlow, A.R., 2005. Lexical tone perception in musicians and non-musicians. Interspeech, Lisbon, Portugal.

Ayotte, J., Peretz, I., Hyde, K., 2002. Congenital amusia: A group study of adults afflicted with a music-specific disorder. Brain 125(2), 238–251.

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. J. Cog. Neurosci. 23(2), 425–434.

Bidelman, G.M., Hutka, S., Moreno, S., 2013. Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. PLoS One 8(4), e60676.

Boersma, P., Weenink, D., 2014. Praat: Doing phonetics by computer, version 6.0.12 [Computer program]. http://www.praat.org (access 14.09.04).

Boets, B., Op de Beeck, H.P., Vandermosten, M., Scott, S.K., Gillebert, C.R., Mantini, D., et al., 2013. Intact but less accessible phonetic representations in adults with dyslexia.

Science 342(6163), 1251–1254.

Celsis, P., Boulanouar, K., Doyon, B., Ranjeva, J.P., Berry, I., Nespoulous, J.L., Chollet, F., 1999. Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. Neuroimage 9(1), 135–144.

Chandrasekaran, B., Chan, A.H.D., Wong, P.C.M., 2011. Neural processing of what and who information in speech. J. Cogn. Neurosci. 23(10), 2690–2700.

Chao, Y.R., 1930. A system of tone letters. Le Maître Phonétique 45, 24-27.

Chica, A.B., Bartolomeo, P., Lupiáñez, J., 2013. Two cognitive and neural systems for endogenous and exogenous spatial attention. Behav. Brain Res. 237, 107–123.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29(3), 162–173.

Delogu, F., Lampis, G., Belardinelli, M.O., 2006. Music-to-language transfer effect: May melodic ability improve learning of tonal languages by native nontonal speakers? Cogn. Process. 7(3), 203–207.

Delogu, F., Lampis, G., Belardinelli, M.O., 2010. From melody to lexical tone: musical ability enhances specific aspects of foreign language perception. Eur. J. Cogn. Psychol. 22(1), 46–61.

Deutsch, D., Henthorn, T., Marvin, E., Xu, H., 2006. Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period (L). J. Acoust. Soc. Am. 119(2), 719–722.

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(4), 801–810.

Frost, S.J., Landi, N., Mencl, W.E., Sandak, R., Fulbright, R.K., Tejada, E.T., et al., 2009. Phonological awareness predicts activation patterns for print and speech. Ann. Dyslexia 59(1), 78–97.

Gandour, J.T., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., et al., 2004. Hemispheric roles in the perception of speech prosody. NeuroImage 23, 344–357.

Gandour, J.T., Wong, D., Lowe, M., Dzemidzic, M., Satthamnuwong, N., Tong, Y., et al., 2002. A cross-linguistic fMRI study of spectral and temporal cues underlying phonological processing. J. Cogn. Neurosci. 14(7), 1076–1087.

Gosselin, N., Jolicœur, P., Peretz, I., 2009. Impaired memory for pitch in congenital amusia. Ann. N. Y. Acad. Sci. 1169(1), 270–272.

Griffiths, T.D., Johnsrude, I., Dean, J.L., Green, G.G., 1999. A common neural substrate for the analysis of pitch and duration pattern in segmented sound? Neuroreport 10, 3825–30.

Gu, F., Zhang, C., Hu, A., Zhao, G., 2013. Left hemisphere lateralization for lexical and acoustic pitch processing in Cantonese speakers as revealed by mismatch negativity. NeuroImage 83, 637-645.

Holcomb, H.H., Medoff, D.R., Caudill, P.J., Zhao, Z., Lahti, A.C., Dannals, R.F., et al., 1998. Cerebral blood flow relationships associated with a difficult tone recognition task in trained normal volunteers. Cereb. Cortex 8, 534–42.

Huang, W.T., Liu, C., Dong, Q., Nan, Y., 2015a. Categorical perception of lexical tones in Mandarin-speaking congenital amusics. Front. Psychol. 6, 829.

Huang, W.T., Nan, Y., Dong, Q., Liu, C., 2015b. Just-noticeable difference of tone pitch contour change for Mandarin congenital amusics. J. Acoust. Soc. Am. 138(1), EL99–104.

Hyde, K.L., Lerch, J.P., Zatorre, R.J., Griffiths, T.D., Evans, A.C., Peretz, I., 2007. Cortical thickness in congenital amusia: when less is better than more. J, Neurosci. 27(47), 13028–13032.

Hyde, K.L., Peretz, I., 2003. "Out-of-pitch" but still "in-time." Ann. N. Y. Acad. Sci. 999(1), 173–176.

Hyde, K.L., Peretz, I., 2004. Brains that are out of tune but in time. Psychol. Sci. 15(5), 356–360.

Hyde, K.L., Zatorre, R.J., Griffiths, T.D., Lerch, J.P., Peretz, I., 2006. Morphometry of the amusic brain: A two-site study. Brain 129(10), 2562–2570.

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

IBM Corp, 2011. IBM SPSS Statistics for Macintosh, version 20.0. Armonk, NY, IBM Corp.

Ito, M., 2000. Mechanisms of motor learning in the cerebellum. Brain Res. 886(1–2), 237–245.

Itoh, K., Suwazono, S., Arao, H., Miyazaki, K., Tsutomu, N., 2005. Electrophysiological correlates of absolute pitch and relative pitch. Cereb. Cortex 15, 760–769.

Japee, S., Holiday, K., Satyshur, M., Mukai, I., Ungerleider, L.G. 2015. A role of right middle frontal gyrus in reorienting of attention: A case study. Front. Syst. Neurosci.

9(23).

Jiang, C., Hamm, J., Lim, V., Kirk, I., Yang, Y., 2012b. Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics. Mem. Cognit. 40(7), 1109–1121.

Jiang, C., Hamm, J.P., Lim, V.K., Kirk, I.J., Chen, X., Yang, Y., 2012a. Amusia results in abnormal brain activity following inappropriate intonation during speech comprehension. PLoS ONE 7(7), e41411.

Jiang, C., Hamm, J.P., Lim, V.K., Kirk, I.J., Yang, Y., 2011. Fine-grained pitch discrimination in congenital amusics with Mandarin Chinese. Music Percept. 28(5), 519–526.

Joanisse, M.F., Zevin, J.D., McCandliss, B.D., 2007. Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. Cereb. Cortex 17(9), 2084–2093.

Kjaer, T.W., Nowak, M., Lou, H.C., 2002. Reflective self-awareness and conscious states: PET evidence for a common midline parietofrontal core. NeuroImage 17(2), 1080–6.

Krumbholz, K., Patterson, R.D., Seither-Preisler, A., Lammertmann, C., Lutkenhoner, B., 2003. Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. Cereb. Cortex, 13(7), 765–772.

Lee, C.Y., Hung, T.H., 2008. Identification of Mandarin tones by English-speaking musicians and nonmusicians. J. Acoust. Soc. Am. 124(5), 3235–3248.

Lee, C.Y., Lekich, A., Zhang, Y., 2014. Perception of pitch height in lexical and musical tones by English-speaking musicians and nonmusicians. J. Acoust. Soc. Am. 135(3), 1607–1615.

Li, X., Gandour, J.T., Talavage, T., Wong, D., Hoffa, A., Lowe, M., et al., 2010. Hemispheric asymmetries in phonological processing of tones versus segmental units. Neuroreport 21(10), 690–694.

Liu, F., Jiang, C., Pfordresher, P., Mantell, J., Xu, Y., Yang, Y., Stewart, L., 2013. Individuals with congenital amusia imitate pitches more accurately in singing than in speaking: Implications for music and language processing. Atten. Percept. Psychophys. 75(8), 1783–1798.

Liu, F., Maggu, A.R., Lau, J.C.Y., Wong, P.C.M., 2015. Brainstem encoding of speech and musical stimuli in congenital amusia: Evidence from Cantonese speakers. Front. Hum. Neurosci. 8, 1029.

Liu, F., Patel, A.D., Fourcin, A., Stewart, L., 2010. Intonation processing in congenital amusia: Discrimination, identification and imitation. Brain 133(6), 1682–1693.

Lou, H.C., Luber, B., Crupain, M., Keenan, J.P., Nowak, M., Kjaer, T.W., et al., 2004. Parietal cortex and representation of the mental self. Proc. Natl. Acad. Sci. U. S. A. 101(17), 6827–6832.

Loui, P., Alsop, D., Schlaug, G., 2009. Tone-deafness – a new disconnection syndrome? J. Neurosci. 29(33), 10215–10220.

Lu, X., Ho, H.T., Liu, F., Wu, D., Thompson, W.F., 2015. Intonation processing deficits of emotional words among Mandarin Chinese speakers with congenital amusia: An ERP study. Front. Psychol. 6, 385.

Lundstrom, B.N., Ingvar, M., Petersson, K.M., 2005. The role of precuneus and left inferior frontal cortex during source memory episodic retrieval. NeuroImage 27(4), 824–834.

Lundstrom, B.N., Petersson, K.M., Andersson, J., Johansson, M., Fransson, P., Ingvar, M., 2003. Isolating the retrieval of imagined pictures during episodic memory: Activation of the left precuneus and left prefrontal cortex. NeuroImage 20(4), 1934–1943.

Luo, H., Ni, J.T., Li, Z.H., Li, X.O., Zhang, D.R., Zeng, F.G., et al., 2006. Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. Proc. Natl. Acad. Sci. 103(51), 19558–19563.

Macmillan, N.A., Creelman, C.D., 2005. Detection Theory: A User's Guide, second ed. Mahwah, Lawrence Erlbaum Associates.

Marin, M.M., Gingras, B., Stewart, L., 2012. Perception of musical timbre in congenital amusia: Categorization, discrimination and short-term memory. Neuropsychologia 50(3), 367–378.

Moreau, P., Jolicœur, P., Peretz, I., 2013. Pitch discrimination without awareness in congenital amusia: evidence from event-related potentials. Brain Cogn. 81(3), 337–344.

Nan, Y., Huang, W., Wang, W., Liu, C., Dong, Q., 2016. Subgroup differences in the lexical tone mismatch negativity (MMN) among Mandarin speakers with congenital amusia. Biol. Psychol. 113, 59–67.

Nan, Y., Sun, Y., Peretz, I., 2010. Congenital amusia in speakers of a tone language: Association with lexical tone agnosia. Brain 133(9), 2635–2642.

Nguyen, S., Tillmann, B., Gosselin, N., Peretz, I., 2009. Tonal language processing in congenital amusia. Ann. N. Y. Acad. Sci. 1169(1), 490–493.

Nicolson, R.I., Fawcett, A.J., Dean, P., 2001. Developmental dyslexia: The cerebellar deficit hypothesis. Trends Neurosci. 24(9), 508–511.

Norman-Haignere, S.V., Albouy, P., Caclin, A., McDermott, J.H., Kanwisher, N.G., Tillmann, B., 2016. Pitch-responsive cortical regions in congenital amusia. J. Neurosci. 36(10), 2986–2994.

Omigie, D., Pearce, M.T., Williamson, V.J., Stewart, L., 2013. Electrophysiological correlates of melodic processing in congenital amusia. Neuropsychologia 51(9), 1749–1762.

Opitz, B., Rinne, T., Mecklinger, A., von Cramon, D.Y., Schröger, E., 2002. Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. NeuroImage 15(1), 167–174.

Pantev, C., Hoke, M., Lehnertz, K., Lütkenhöner, B., Anogianakis, G., Wittkowski, W., 1988. Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields. Electroencephalogr. Clin. Neurophysiol. 69(2), 160–170.

Patel, A.D., Foxton, J.M., Griffiths, T.D., 2005. Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. Brain Cogn. 59(3), 310–313.

Peng, G., Deutsch, D., Trevor, H., Su, D., Wang, W.S.Y., 2013. Language experience influences non-linguistic pitch perception. J. Chinese Ling. 41(2), 419–447.

Peng, G., Zhang, C., Zheng, H.Y., Minett, J.W., Wang, W.S.Y., 2012. The effect of intertalker variations on acoustic-perceptual mapping in Cantonese and Mandarin tone systems. J. Speech Lang. Hear. Res. 55(2), 579–595.

Peretz, I., Ayotte, J., Zatorre, R.J., Mehler, J., Ahad, P., Penhune, V.B., et al., 2002. Congenital amusia: A disorder of fine-grained pitch discrimination. Neuron 33(2), 185–191.

Peretz, I., Brattico, E., Järvenpää, M., Tervaniemi, M., 2009. The amusic brain: In tune, out of key, and unaware. Brain 132(5), 1277–1286.

Peretz, I., Brattico, E., Tervaniemi, M., 2005. Abnormal electrical brain responses to pitch in congenital amusia. Ann. Neurol. 58(3), 478–482.

Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L.L., Gagnon, B., Trimmer, C.G., et al., 2008. On-line identification of congenital Amusia. Music Percept. 25(4), 331–343.

Peretz, I., Hyde, K.L., 2003. What is specific to music processing? Insights from congenital amusia. Trends Cogn. Sci. 7(8), 362–367.

Pfordresher, P., Brown, S., 2009. Enhanced production and perception of musical pitch in tone language speakers. Atten. Percept. Psychophys. 71(6), 1385–1398.

Rajah, M.N., Languay, R., Grady, C.L., 2011. Age-related changes in right middle frontal gyrus volume correlate with altered episodic retrieval activity. J. Neurosci. 31(49), 17941–54.

Ramus, F., Marshall, C.R., Rosen, S., Van Der Lely, H.K.J., 2013. Phonological deficits in specific language impairment and developmental dyslexia: Towards a multidimensional model. Brain 136(2), 630–645.

Rinne, T., Alho, K., Ilmoniemi, R.J., Virtanen, J., Näätänen, R., 2000. Separate time behaviors of the temporal and frontal mismatch negativity sources. NeuroImage 12, 14–

19.

Saffran, J.R., Griepentrog, G.J., 2001. Absolute pitch in infant auditory learning: Evidence for developmental reorganization. Dev. Psychol. 37(1), 74–85.

Shulman, G.L., Astafiev, S.V., Franke, D., Pope, D.L.W., Snyder, A.Z., McAvoy, M.P., et al., 2009. Interaction of stimulus-driven reorienting and expectation in ventral and dorsal fronto-parietal and basal ganglia-cortical networks. J. Neurosci. 29(14), 4392–4407.

Sihvonen, A.J., Ripollés, P., Leo, V., Rodríguez-Fornells, A., Soinila, S., Särkämö, T., 2016. Neural basis of acquired amusia and its recovery after stroke. J. Neurosci. 36(34), 8872–8881.

Smayda, K.E., Chandrasekaran, B., Maddox, W.T., 2015. Enhanced cognitive and perceptual processing: A computational basis for the musician advantage in speech learning. Front. Psychol. 6, 682.

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. Front. Psychol. 2(120), 1–15.

Tillmann, B., Lévêque, Y., Fornoni, L., Albouy, P., Caclin, A., in press. Impaired short-term memory for pitch in congenital amusia. Brain Res.

Tillmann, B., Rusconi, E., Traube, C., Butterworth, B., Umiltà, C., Peretz, I., 2011b. Fine-grained pitch processing of music and speech in congenital amusia. J. Acoust. Soc. Am. 130(6), 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 Cogn. 71(3), 259–264.

Tong, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Xu, Y., et al., 2005. Neural circuitry underlying sentence-level linguistic prosody. NeuroImage 28(2), 417–428.

Van Lancker, D., Fromkin, V.A., 1973. Hemispheric specialization for pitch and “tone”: Evidence from Thai. J. Phon. 1, 101–109.

Van Lancker, D., Fromkin, V.A., 1978. Cerebral dominance for pitch contrasts in tone language speakers and in musically untrained and trained English speakers. J. Phon. 6, 19–23.

Wallentin, M., Roepstorff, A., Glover, R., Burgess, N., 2006. Parallel memory systems for talking about location and age in precuneus, caudate and Broca’s region. NeuroImage 32(4), 1850–1864.

Wang, X., Peng, G., 2014. Phonological processing in Mandarin speakers with congenital amusia. J. Acoust. Soc. Am. 136(6), 3360–3370.

Wang, Y., Jongman, A., Sereno, J.A., 2001. Dichotic perception of Mandarin tones by Chinese and American listeners. Brain Lang. 78(3), 332–348.

Wong, P.C.M., Ciocca, V., Chan, A.H.D., Ha, L.Y.Y., Tan, L.H., Peretz, I., 2012. Effects of culture on musical pitch perception. PLoS ONE 7(4), e33424.

Wong, P.C.M., Diehl, R.L., 2003. Perceptual normalization for inter- and intratalker variation in Cantonese level tones. J. Speech Lang. Hear. Res. 46(2), 413–421.

Wong, P.C.M., Parsons, L.M., Martinez, M., Diehl, R.L., 2004. The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. J. Neurosci. 24(41), 9153–9160.

Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex.

Cereb. Cortex 11(10), 946–953.

Zatorre, R.J., Evans, A.C., Meyer, E., 1994. Neural mechanisms underlying melodic perception and memory for pitch. J. Neurosci. 14, 1908–19.

Zatorre, R.J., Evans, A.C., Meyer, E., Gjedde, A., 1992. Lateralization of phonetic and pitch discrimination in speech processing. Science 256(5058), 846–849.

Zendel, B.R., Lagrois, M.É., Robitaille, N., Peretz, I., 2015. Attending to pitch information inhibits processing of pitch information: the curious case of amusia. J. Neurosci. 35(9), 3815-3824.

Zhang, C., Chen, S., 2016. Toward an integrative model of talker normalization. J. Exp. Psychol. Hum. Percept. Perform. 42(8), 1252-1268.

Zhang, C., Peng, G., Wang, W.S.Y., 2012. Unequal effects of speech and nonspeech contexts on the perceptual normalization of Cantonese level tones. J. Acoust. Soc. Am. 132(2), 1088–1099.

Zhang, C., Peng, G., Wang, W.S.Y., 2013. Achieving constancy in spoken word identification: Time course of talker normalization. Brain Lang. 126(2), 193–202.

Zhang, C., Pugh, K.R., Mencl, W.E., Molfese, P.J., Frost, S.J., Magnuson, J.S., et al., 2016. Functionally integrated neural processing of linguistic and talker information: an event-related fMRI and ERP study. NeuroImage 124, 536–549.

Table 1. Demographic characteristics of the participants. The amusic participants were identified using the Online Identification Test of Congenital Amusia (Peretz et al. 2008; http://www.brams.umontreal.ca/onlinetest). All amusic participants scored 73 or lower, whereas the control participants scored 80 or higher in the global score of the test, which is the average of all three subtests – out-of-key, offbeat and mistuned subtests. The results (*p*-value) of t-tests comparing the amusics and controls in age and the scores of the amusia test are reported.

fMRI experiment	Amusics	Controls	*p*-value
No. of participants	11 (2 M, 9 F)	11 (2 M, 9 F)	/
Age (range)	22.0 ± 3.1 years (19.2-28.5 years)	21.7 ± 2.7 years (18.8-28.8 years)	*n.s.*
Online Identification Test of Congenital Amusia			
Out-of-key (SD)	67.5 (6.8)	89.5 (5.6)	$p < 0.001$
Offbeat (SD)	74.3 (11.4)	85.6 (7.8)	$p < 0.05$
Mistuned (SD)	57.5 (5.0)	91.9 (5.9)	$p < 0.001$
Global score (SD)	66.5 (4.5)	89.6 (3.5)	$p < 0.001$

Note: M = male; F = female; n.s. = not significant; SD = standard deviation.

Table 2. The mean F0 (Hz) of six pairs of lexical tone stimuli and the notes of pitch-matched musical stimuli presented in the repetition, fixed interval and varied interval conditions.

	55-33/33-55				33-22/22-33				55-22/22-55			
	Speech		Piano		Speech		Piano		Speech		Piano	
Repetition	277	233	C♯4	A♯3	233	208	A♯3	G♯3	277	208	C♯4	G♯3
Fixed interval	349	294	F4	D4	294	262	D4	C4	349	262	F4	C4
	330	277	E4	C♯4	277	247	C♯4	B3	330	247	E4	B3
	311	262	D♯4	C4	262	233	C4	A♯3	311	233	D♯4	A♯3
	294	247	D4	B3	247	220	B3	A3	294	220	D4	A3
	262	220	C4	A3	220	196	A3	G3	262	196	C4	G3
	247	208	B3	G♯3	208	185	G♯3	F♯3	247	185	B3	F♯3
	233	196	A♯3	G3	196	175	G3	F3	233	175	A♯3	F3
	220	185	A3	F♯3	185	165	F♯3	E3	220	165	A3	E3
Varied interval	349	277	F4	C♯4	294	247	D4	B3	349	247	F4	B3
	330	294	E4	D4	277	262	C♯4	C4	330	262	E4	C4
	311	247	D♯4	B3	262	220	C4	A3	311	220	D♯4	A3
	294	262	D4	C4	247	233	B3	A♯3	294	233	D4	A♯3
	262	208	C4	G♯3	220	185	A3	F♯3	262	185	C4	F♯3
	247	220	B3	A3	208	196	G♯3	G3	247	196	B3	G3
	233	185	A♯3	F♯3	196	165	G3	E3	233	165	A♯3	E3
	220	196	A3	G3	185	175	F♯3	F3	220	175	A3	F3

Table 3. Significantly activated clusters in the whole-brain analysis (cluster-wise FWE corrected p = 0.05, uncorrected p = 0.001). FWE-corrected critical cluster size was 36.2 voxels (NN level 1, faces must touch and 2-sided threshold). MNI coordinates are reported for peak activation in the LPI format.

Condition	Region	x	y	z	Size (cm^3)
Main effect of group (in repetition condition)					
Amusics > Controls	R MFG	38	41	17	1.377
Group × Interval (Varied vs. Repetition) interaction					
Amusics > Controls	R precuneus	29	-77	44	1.296
Within-group analyses of relative pitch processing					
Controls: Varied > Fixed = Repetition	Cerebellum	14	-32	-7	2.484
Controls (Speech): Varied > Fixed = Repetition	R STG	63	-4	2	1.107
Other activations					
Speech > Music	R STG	66	-20	1	2.646
Varied > Repetition	R STG	57	5	-1	0.999
Controls: Varied > Repetition	R STG	60	8	01	1.08
Controls: Varied > Fixed	R SFG	26	2	58	2.322
	Cerebellum	14	-35	-7	2.214
Controls (Speech): Varied > Repetition	R STG	66	-13	1	1.296
Amusics (Speech): Repetition > Varied	L fusiform gyrus	-53	-47	-15	1.242

Note: L = left; R = right; MFG = middle frontal gyrus; STG = superior temporal gyrus; SFG

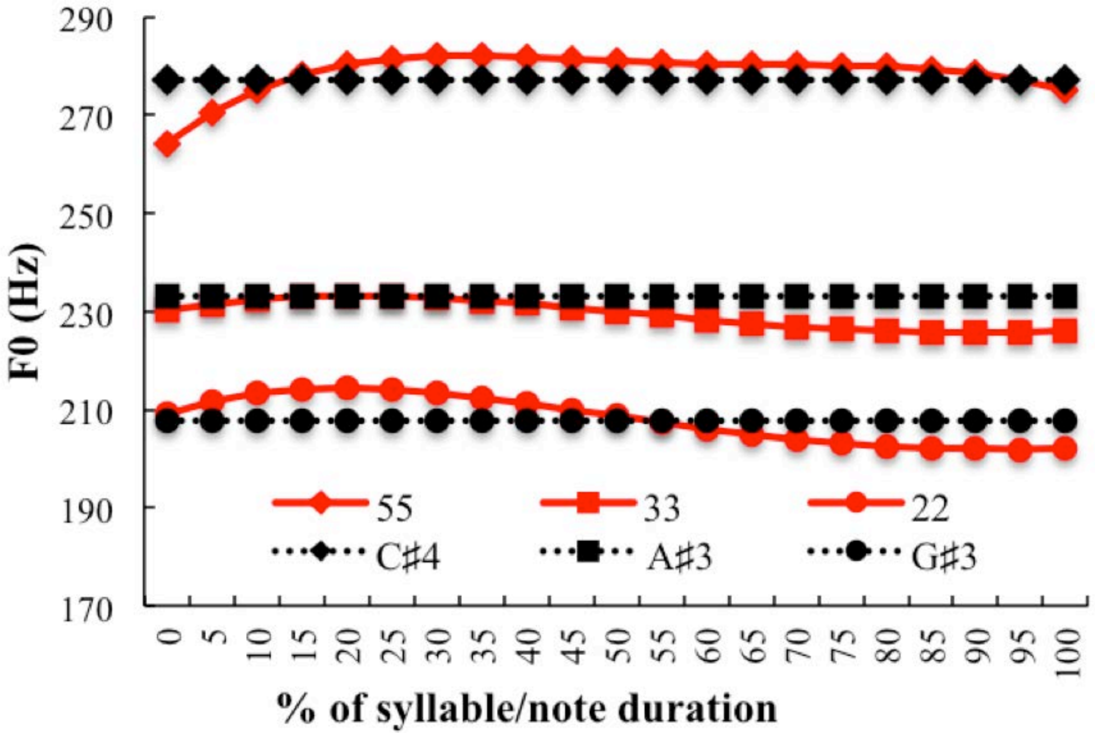
= superior frontal gyrus.

Fig. 1. F0 trajectory of the three level tones and three pitch-matched musical notes. The three words carrying the three level tones (醫 /ji55/ ‘a doctor’; 意 /ji33/ ‘meaning’; 二 /ji22/ ‘second’) were produced by a female Cantonese speaker. The three pitch-matched musical notes were C♯4, A♯3, and G♯3. The three level tones were indicated by red lines, and the musical notes were indicates by black lines.

Fig. 2. Example blocks of musical stimuli in the fMRI experiment. (A) Repetition condition. (B) Fixed interval condition. (C) Varied interval condition.

Fig. 3. Significantly activated brain clusters in the whole-brain analysis (FWE corrected p = 0.05, uncorrected p = 0.001). (A) Significant main and interaction effects of *group*. (B) Significant activations in the control group revealed by the within-group analyses of relative pitch processing (Varied interval > Fixed interval = Repetition). R = right; MFG = middle frontal gyrus; STG = superior temporal gyrus; Rep = repetition condition.

Fig. 4. ROI analysis of the right IFG. A 5 mm sphere centering on the coordinates of the right IFG reported in Hyde et al. (2011) (MNI coordinates: x = 34, y = 32, z = 0) was generated. The mean beta values of all voxels within the sphere were extracted for each condition, and input to the *group* × *domain* × *interval* ANOVA analysis using SPSS. R = right; IFG = inferior frontal gyrus; Rep = repetition condition.


290
270
250
230
210
190
170
F0 (Hz)
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
% of syllable/note duration
55
33
22
C♯4
A♯3
G♯3

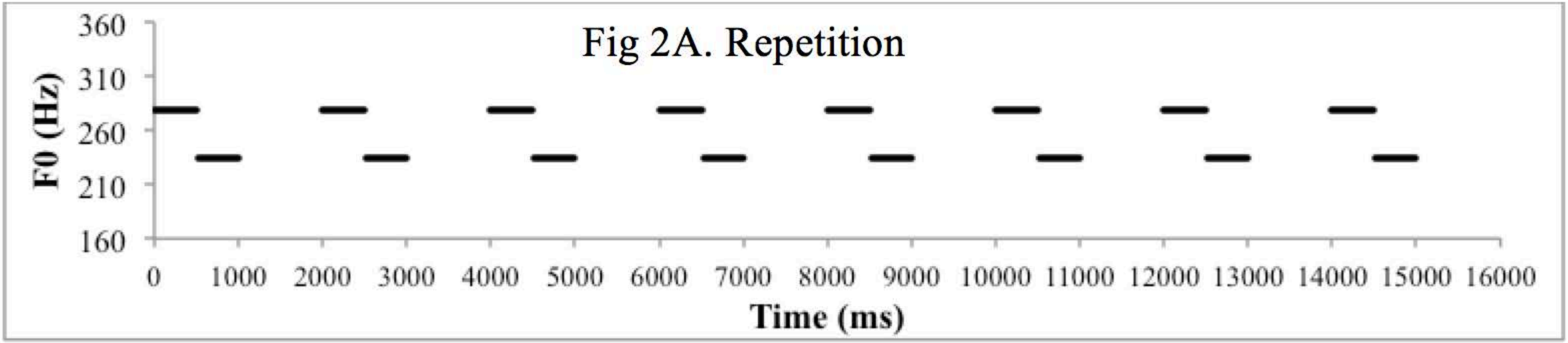


Fig 2A. Repetition

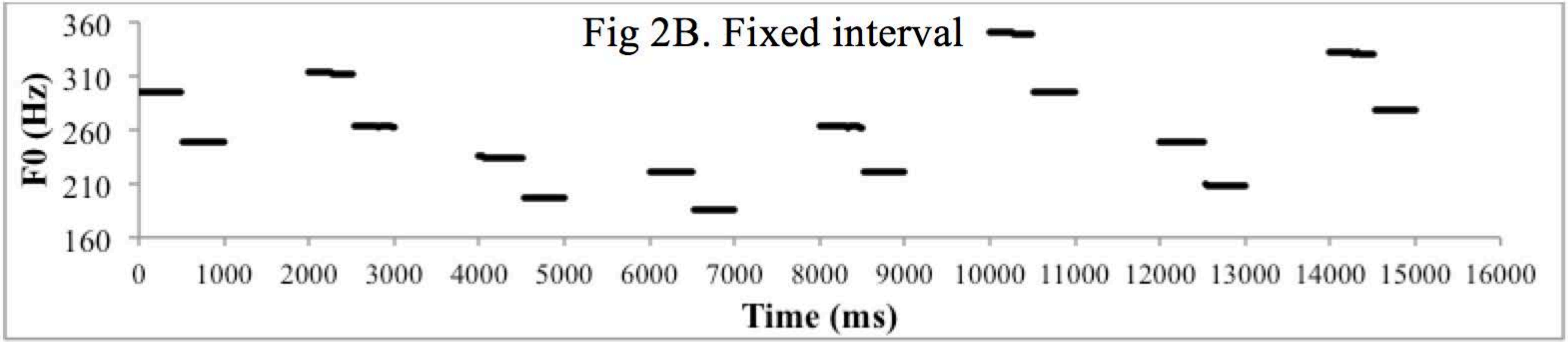


Fig 2B. Fixed interval

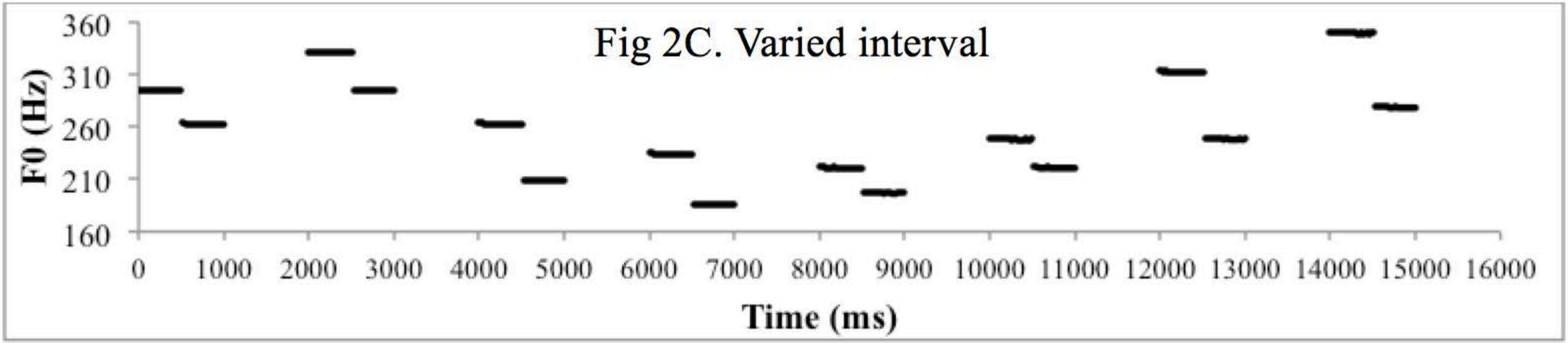


Fig 2C. Varied interval

Fig 3A. Group analyses

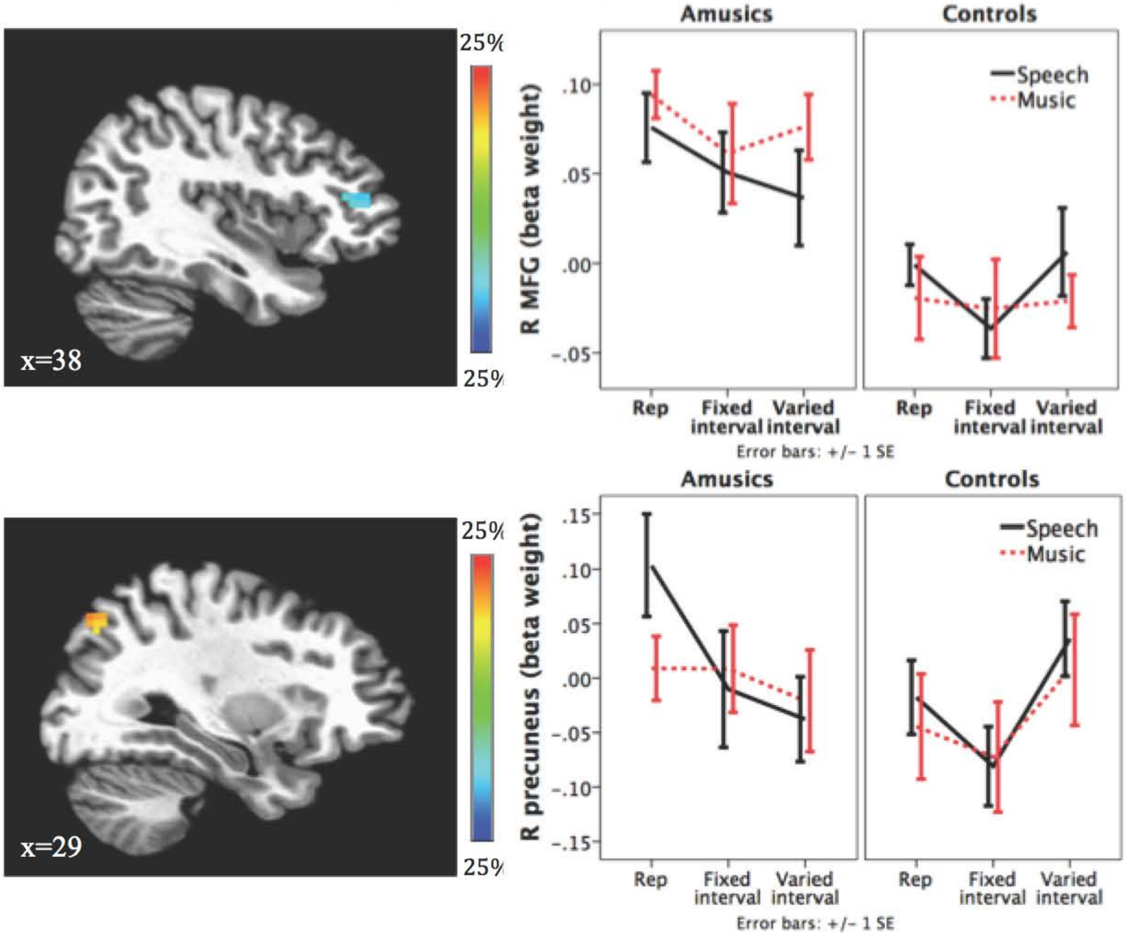


Fig 3B. Within-group analyses of relative pitch processing

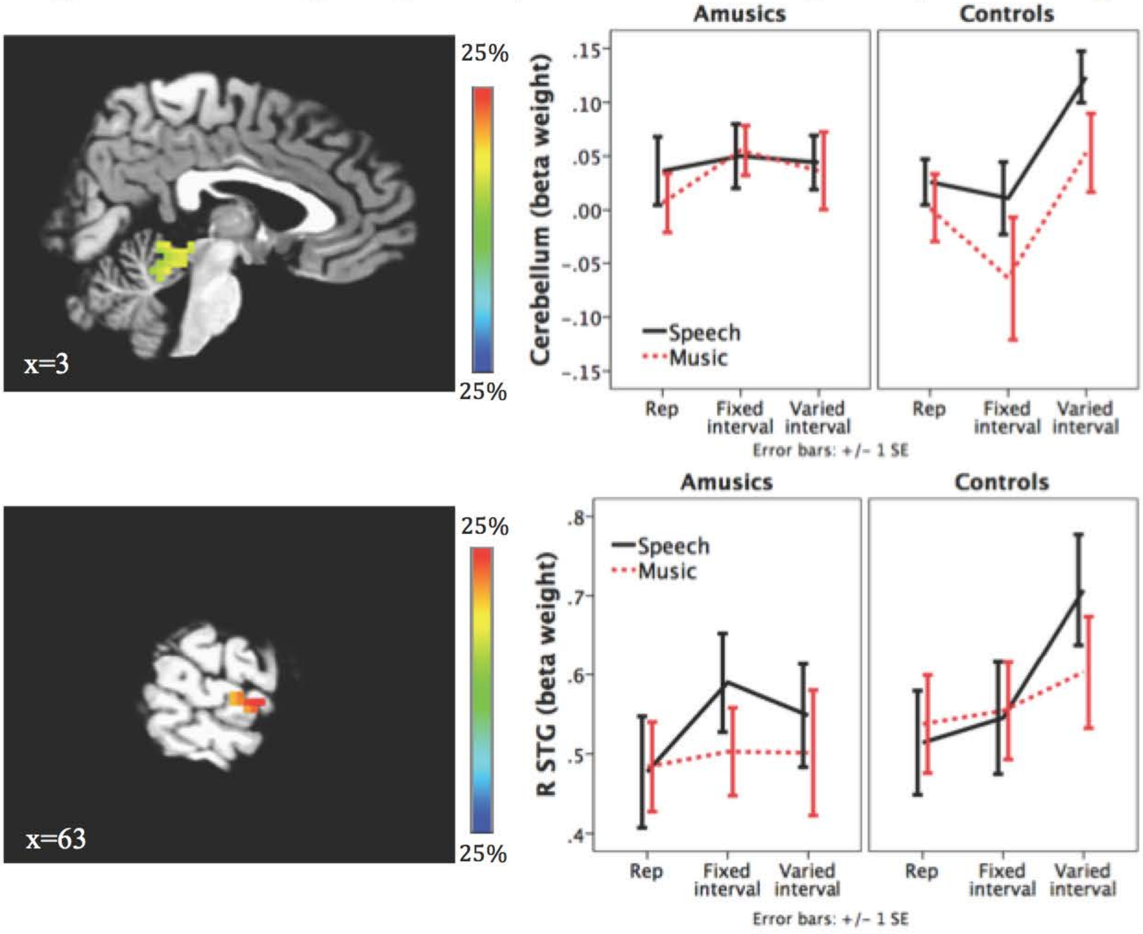


Fig 4. ROI analysis of right inferior frontal gyrus