The neural basis for understanding imitation-induced musical meaning: the role of the human mirror system


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

To link to this article DOI: http://dx.doi.org/10.1016/j.bbr.2018.11.020

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR
The neural basis for understanding imitation-induced musical meaning: The role of the human mirror system

Jun Jiang\textsuperscript{a}, Fang Liu\textsuperscript{b}, Linshu Zhou\textsuperscript{a}, Cunmei Jiang\textsuperscript{a,c,*}

\textsuperscript{a} Music College, Shanghai Normal University, Shanghai, China
\textsuperscript{b} School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK
\textsuperscript{c} Institute of Psychology, Shanghai Normal University, Shanghai, China

Correspondence should be addressed to:
Dr. Cunmei Jiang
Music College
Shanghai Normal University
100 E. Guilin Road
Shanghai, 200234, China
Tel: 0086-21-64322990; Fax: 0086-21-64322935
Electronic mail: cunmeijiang@126.com
**Abstract**

Music can convey meanings by imitating phenomena of the extramusical world, and these imitation-induced musical meanings can be understood by listeners. Although the human mirror system (HMS) is implicated in imitation, little is known about the HMS’s role in making sense of meaning that derives from musical imitation. To answer this question, we used fMRI to examine listeners’ brain activities during the processing of imitation-induced musical meaning with a cross-modal semantic priming paradigm. Eleven normal individuals and 11 individuals with congenital amusia, a neurodevelopmental disorder of musical processing, participated in the experiment. Target pictures with either an upward or downward movement were primed by semantically congruent or incongruent melodic sequences characterized by the direction of pitch change (upward or downward). When contrasting the incongruent with the congruent condition between the two groups, we found greater activations in the left supramarginal gyrus/inferior parietal lobule and inferior frontal gyrus in normals but not in amusics. The implications of these findings in terms of the role of the HMS in understanding imitation-induced musical meaning are discussed.

Keywords: fMRI, musical meaning, inferior frontal gyrus, inferior parietal lobule, congenital amusia
1 Introduction

Like language, music is unique to humans. Before language emerged, ancient human ancestors were assumed to communicate with each other using gestures, facial expressions, and songlike vocalizations [1], which imitated animal movements, animal calls, and sounds of the natural world [2]. Being imitative in nature, musical vocalizations are also thought to precede spoken words in human evolution [2-5]. A similar case has also been observed in infant development, during which prelinguistic infants are able to make music-like sounds that are imitations of maternal utterances [6, 7]. Thus, imitation has been playing a critical role in human communication, evolution, and development.

Similarly, people nowadays also convey intentions and meanings through music [8]. Musical meaning can emerge from the imitation of phenomena in the extramusical world, including the sounds of creatures (e.g., birds), the qualities of objects (e.g., trains), and even the representations of abstract concepts (e.g., wideness) [9-11]. As proposed by the mimesis theory [12], music is an imitation or expression of the phenomena in our environment, and this imitation or expression can be readily captured by listeners [13, 14]. Although understanding the intentions of a composer or performer is a critical component in musical meaning processing [15], little is known about the neural correlates underlying the understanding of musical meaning that derives from the imitation of phenomena.

The human mirror system (HMS) consists of a core parietofrontal network of regions in the inferior parietal lobule (IPL) and inferior frontal gyrus (IFG)/premotor cortex (PMC), and it has been suggested that mirror neurons are fired during imitation of actions [16-19]. Substantial evidence has revealed that the HMS is activated by both action observation and execution [20-24] and motor imagery [22, 25-27]. Listening to human action sounds,
including speech sounds [28-31] and action-related sounds such as tearing paper, screaming and kicking a football [32-35], also causes increased activations in the HMS. More importantly, the HMS is also engaged in processing semantic information of action-related sentences [36]. Taken together, previous studies indicate that imitation leads to the activation of the HMS, where mirror neurons automatically map observed or heard actions onto a motor program [37].

Although the HMS has also been assumed to be associated with imitation in music [38], very few experimental studies have investigated this relationship directly. Previous studies have shown that the HMS can be recruited by musical activities such as watching piano performances [39, 40], tapping to rhythms [41], and making emotional judgments of songs [42]. Furthermore, prior studies have also shown the engagement of a network of activity resembling the HMS during reproducing melodies [43] and rhythmic patterns [44, 45]. However, these musical activities are irrelevant to music (or musical structure) per se.

Music itself can convey meanings by imitating the phenomena in the extramusical world [9-11]. However, none of the previous studies has investigated the HMS’s role in making sense of imitation-induced musical meaning.

To address this issue, we used event-related fMRI techniques to investigate whether the HMS is implicated in processing musical meaning derived from imitation. Using a cross-modal semantic priming paradigm, a widely used design for studies of meaning processing in music [e.g., 46, 47], we presented participants with melodic sequences (primes) that were followed by pictures (targets). The melodic sequences contained a series of notes with either an upward or a downward direction, and the target pictures portrayed either a simple upward or downward movement, with the effect of picture complexity on
brain activity controlled for [48]. Thus, the target pictures were either semantically congruent or incongruent to the musical primes. We hypothesized that, if the listeners were able to process the semantic relationship between the primes and targets by showing a priming effect, it would indicate that they had understood the link between the musical imitation (through melodic sequences) and the natural phenomena (as depicted by the pictures).

Apart from normal participants, we also tested individuals with congenital amusia (hereafter amusia), a neurodevelopmental disorder of musical processing [49]. Individuals with amusia not only have difficulty discriminating fine-grained pitch differences [50-53], but also show deficits in perceiving pitch direction [50, 51, 54, 55] and melodic contour [51, 56], which has led to the “melodic contour deafness hypothesis” by Patel [57]. Given that melodic contour is a building block for musical melodies, deficits in melodic contour perception may affect the understanding of musical meaning in amusia. Indeed, our previous study showed that amusics failed to elicit an N400 effect to musical meaning represented by the direction of pitch change, even though the pitch distances between the first and last notes of the melodic sequences exceeded their pitch direction discrimination thresholds [47]. Thus, by including both normal and amusic individuals as participants, we would be able to reveal the core brain regions that are necessarily required for effective processing of musical meaning, as compared to dysfunctional processing in amusia. Given the role of the HMS in imitation [e.g., 16, 18], we predicted that the parietofrontal regions including the IPL and IFG/PMC would be activated during the understanding of musical meaning in normal but not in amusic individuals.
2 Materials and methods

2.1 Participants

Twelve amusic and 12 normal individuals took part in the experiment. The Montreal Battery of Evaluation of Amusia (MBEA), consisting of six subtests—scale, contour, interval, rhythm, meter, and memory, was used to assess musical abilities of these participants [58, 59]. Participants were identified as amusics if they scored 65 or below on the melodic composite score (sum of the scores on the scale, contour, and interval tests) [54], and below 78% correct on the MBEA global score [59]. Furthermore, participants’ pitch change detection and pitch direction discrimination thresholds were measured using a two-alternative forced choice AXB paradigm [50], in which a sequence of three stimuli (A, X, and B) were presented in each trial and the second one (X) was either the same as the first (A) or last (B). In particular, on each trial of the pitch change detection threshold task, three pure tones (A, X, and B) were presented consecutively, among which tone X was identical to either tone A or B. Participants were required to judge whether tone X had the same pitch as tone A or B. On each trial of the pitch direction discrimination threshold task, three 2-tone pairs (A, X, and B) were presented, where pair X was identical in pitch direction to either pair A or B. Participants were required to indicate whether the pitch direction of pair X was the same as that of pair A or B. Experimental trials were presented with an adaptive staircase procedure, in which the initial pitch difference was four semitones and was then varied with a “two-down, one-up” staircase procedure. The procedure terminated after 16 reversals, and the threshold was computed based on the average pitch difference of the last six reversals. For more details please refer to ref. [50]. After the fMRI experiment, one amusic participant and one matched control participant
were excluded from further analysis based on a head motion criterion of 3 mm of translational and 3° of rotational movement across the session. As a result, the final sample consisted of 11 participants in each group.

As shown in Table 1, the two groups were matched in age, sex, and education, but the amusic group showed worse performance on the MBEA and higher pitch thresholds than the control group. All participants were right-handed as assessed by the Edinburgh Handedness Inventory [60]. All had normal hearing and reported to have normal or corrected-to-normal vision. None had reported history of neurological or psychiatric diseases. None of the controls received formal musical training or played any musical instruments. One amusic participant had one year of singing lessons and another had one year of piano lessons. Ethical approval was granted by Shanghai Normal University in China, and all participants gave written informed consent and were paid for their participation.

Table 1 Characteristics of participants

<table>
<thead>
<tr>
<th>Variable</th>
<th>Amusics</th>
<th>Controls</th>
<th>Group-statistic</th>
<th>d (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in years</td>
<td>22.91 (2.55)</td>
<td>23.09 (0.94)</td>
<td>t(12.69) = -0.22, p = .828</td>
<td>-0.09 (-0.93, 0.74)</td>
</tr>
<tr>
<td>Sex (male: female)</td>
<td>3:8</td>
<td>3:8</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Handedness</td>
<td>11R</td>
<td>11R</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Education in years</td>
<td>15.64 (2.11)</td>
<td>16.36 (1.12)</td>
<td>t(15.22) = -1.01, p = .329</td>
<td>-0.43 (-1.27, 0.43)</td>
</tr>
</tbody>
</table>

Mean scores of MBEA

<p>| Scale                   | 18.27 (2.72)  | 27.91 (1.22)  | t(13.86) = -10.71, p &lt; .001 | -4.57 (-6.43, -2.67) |
| Contour                 | 19.00 (3.29)  | 28.27 (0.90)  | t(11.51) = -9.02, p &lt; .001  | -3.85 (-5.59, -2.07) |</p>
<table>
<thead>
<tr>
<th>Interval</th>
<th>18.00 (3.90)</th>
<th>27.64 (1.50)</th>
<th>t(12.90) = -7.65, p &lt; .001</th>
<th>-3.26 (-4.74, -1.74)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhythm</td>
<td>19.73 (3.88)</td>
<td>26.91 (1.64)</td>
<td>t(13.47) = -5.66, p &lt; .001</td>
<td>-2.41 (-3.62, -1.16)</td>
</tr>
<tr>
<td>Meter</td>
<td>18.45 (5.05)</td>
<td>27.18 (2.79)</td>
<td>t(20) = -5.02, p &lt; .001</td>
<td>-2.14 (-3.19, -1.06)</td>
</tr>
<tr>
<td>Memory</td>
<td>20.36 (5.08)</td>
<td>28.00 (1.48)</td>
<td>t(11.69) = -4.78, p &lt; .001</td>
<td>-2.04 (-3.19, -0.85)</td>
</tr>
<tr>
<td>Melodic score</td>
<td>55.27 (7.21)</td>
<td>83.82 (2.86)</td>
<td>t(13.06) = -12.20, p &lt; .001</td>
<td>-5.20 (-7.33, -3.05)</td>
</tr>
<tr>
<td>Global score (%)</td>
<td>63.23 (5.90)</td>
<td>92.17 (3.10)</td>
<td>t(15.12) = -14.40, p &lt; .001</td>
<td>-6.14 (-8.45, -3.81)</td>
</tr>
</tbody>
</table>

**Pitch perception threshold**

| Pitch change detection | 1.40 (1.04) | 0.27 (0.12) | t(10.27) = 3.57, p = .005 | 1.52 (0.44, 2.56) |
| Pitch direction discrimination | 4.02 (1.32) | 0.48 (0.22) | t(10.57) = 8.76, p < .001 | 3.74 (1.94, 5.49) |

**Note.** R = right-handed; MBEA = Montreal Battery of Evaluation of Amusia. MBEA scores are expressed as the number of correct responses out of 30. Pitch perception thresholds are in semitones. Standard deviation values are shown in parentheses.

### 2.2 Stimuli

A total of 84 short melodic sequences were created as potential prime stimuli, which consisted of 15 quarter notes with either an upward or a downward pitch direction. Half of the sequences were constructed with a small pitch distance no more than two semitones between two consecutive tones, whereas the other half were constructed with a large pitch distance of more than four semitones. Because all sequences had an ascending or a descending direction, the pitch distances between the first and last tones ranged from 12 to 28 semitones. Each sequence lasted four seconds. All notes were generated by Overture 4.1 (GenieSoft Inc.) and played with a grand piano sound using Pianissimo 1.0 (Acoustica Inc.). Using Adobe Audition CS6 (Adobe Systems Inc.), these sequences were normalized to have the same sound intensity level (-3 dB) and 400 ms fade-out ramps.
Target pictures were two schematic representations, one with an upward movement and the other with a downward movement. Each prime was paired with a target picture to form semantically congruent or incongruent conditions (see Fig. 1), thus resulting in 168 prime-target pairs.

A behavioral pretest was conducted in order to select paired stimuli for the formal fMRI experiment. An independent group of 17 musically untrained participants were asked to rate each prime-target pair for their congruency on a 9-point scale from 1 = strongly incongruent in meaning to 9 = strongly congruent in meaning, resulting in two ratings for each prime. The pairs with ratings below 4 and above 6 were considered to be incongruent and congruent, respectively. For each prime, paired t-tests were conducted to determine whether the two ratings were significantly different from each other. In the end, 96 prime-target pairs (congruency ratings on 48 congruent pairs: $M = 7.28, SD = 0.67$; congruency ratings on 48 incongruent pairs: $M = 2.84, SD = 0.60$; $t(47) = 31.58, p < .001$) were selected for the fMRI experiment. Among these pairs, primes comprised of 24 upward and 24 downward melodic sequences with 12 small and 12 large pitch distances, respectively. As in the pretest, each prime was used twice: in one trial, the prime was semantically congruent to the target picture; and in the other trial, the prime and target were incongruent.

### 2.3 Procedure

Before performing the fMRI experiment, participants were first trained on the task using a personal computer. They repeatedly practiced six trials outside the scanner until
they became fully familiarized with the task. Inside the scanner, foam pads were used, and explicit verbal instructions were given to the participants in order to limit their head motion. Using the E-prime 2.0 software (Psychology Software Tools Inc.), visual stimuli were presented and projected onto a screen at the front of the scanner. Participants viewed these stimuli through a mirror attached to the head coil. Auditory stimuli were presented via MR-compatible headphones. Participants wore foam earplugs to reduce the effect of ambient scanner noise. Before scanning, the sound volume was adjusted to a comfortable listening level for each participant.

During the scanning session, participants were asked to keep their body still and their eyes open, directing toward the middle of the screen. They were first given the same six trials for practice. Each trial started with a black fixation cross (with a random duration of 2 s, 4 s, or 6 s) at the center of the white screen. When the fixation cross disappeared, the prime was presented with a black screen background (4 s), followed by the visual presentation of a target picture (2 s). After the target picture had disappeared, a response screen (2 s) appeared and participants were asked to indicate whether the meaning expressed by the prime-target pair was congruent or incongruent by pressing one of the two response buttons with their thumbs. Half of the participants responded to the congruent trials with the right thumb and the incongruent trials with the left thumb, and the other half did the opposite. The next trial began 2 s after the previous trial, no matter whether the participants made their response within 2 s or not. The 96 trials were presented in pseudo-randomized order with a maximum of three congruent or incongruent trials in a row. These trials were divided into two sessions with a break in between. The order of presentation was counterbalanced across participants.
2.4 Data acquisition and analysis

Participants were scanned on a 3T Siemens Magnetom Trio Tim System (Siemens, Erlangen, Germany) with a 12-channel head coil. Functional images were acquired using a gradient echo-planar imaging sequence with the following parameters: repetition time = 2 s, echo time = 30 ms, field of view = 192 mm, flip angle = 90°, matrix size = 64 × 64 mm, voxel size = 3 × 3 × 3 mm³, 37 axial slices, slice thickness = 3 mm, and inter-slice gap = 0.6 mm.

The analysis of the fMRI data was performed using the DPABI 3.0 [61] and SPM 12 toolboxes (Wellcome Trust Center for Neuroimaging, London, UK) run in Matlab R2016a (MathWorks Inc, Natick, MA, USA). The two toolboxes offer complementary functions. Compared to SPM, DPABI requires less manual operations, making preprocessing less time-consuming with a reduced risk of inadvertent mistakes [61, 62]. However, SPM, but not DPABI, provides further statistical analyses (e.g., GLM analysis) for task-based fMRI data [61]. Since DPABI also calls SPM functions in preprocessing data [62], it is convenient to perform the GLM analysis using SPM based on the preprocessed data by DPABI [61]. Therefore, the functional imaging data were first preprocessed in DPABI. For preprocessing, blood oxygen level dependent (BOLD) images were realigned to the first functional image in order to correct for head movement. The realigned volumes were normalized to the EPI template in the Montreal Neurological Institute (MNI) space. Each normalized image was then smoothed with a 6-mm full-width at half-maximum isotropic Gaussian kernel. Statistical analyses based on the general linear model were then carried out in SPM. At the first level, neural activity during the presentation of the target was modeled by a canonical hemodynamic response function and its temporal derivative. For each participant, four
conditions were modeled separately: congruent trials with correct responses, congruent
trials with incorrect and omitted responses, incongruent trials with correct responses,
incongruent trials with incorrect and omitted responses. Additionally, we included six
regressors modeling head movement parameters. To reduce low frequency drifts, a high-
pass filter was set at 128 s. An autoregressive AR(1) model was applied to account for
temporal autocorrelation intrinsic to the fMRI time-series. Individual contrast images for the
conditions “incongruent against congruent trials with correct responses” and “congruent
against incongruent trials with correct responses” were generated. One-sample t-tests were
performed to evaluate whether the observed differences were significantly different from
zero within each group separately. Regions consisting of a cluster size of at least 10 voxels
that exceeded an uncorrected voxel threshold of $p < .001$ were considered reliable. At the
second level, two-sample independent t-tests were used to investigate the effect of group.
Considering the role of the HMS in imitation, we included the combined regions of interest
(ROIs; i.e., the bilateral IPL, supramarginal gyrus, IFG, and precentral gyrus) for small
volume correction analysis. All anatomical ROIs were constructed using the Talairach
 Daemon option of the WFU PickAtlas 3.0.5 toolbox [63]. Statistical significance was set at
voxel-level uncorrected $p < .001$ and cluster-level false discover rate (FDR) corrected $p
< .05$ with $k > 10$ voxels. Anatomical labels for activated regions were provided by the SPM
Anatomy toolbox version 2.2c [64, 65] and if unavailable, then provided by the xjView 9.6
toolbox (http://www.alivelearn.net/xjview/).
Bivariate correlations were computed between MBEA scores, behavioral data, and beta
values of the ROIs. For each ROI demonstrating a significant group effect, the MarsBaR
0.44 toolbox (http://marsbar.sourceforge.net/index.html) was used to extract each
participant’s mean BOLD parameter estimate value (beta value) from a 6 mm radius sphere around peak activations.

3 Results

3.1 Behavioral results

To eliminate the contribution of response bias, we adopted the sensitivity measure from the signal detection theory [66] as an accuracy index. A “yes” response to a congruent trial was defined as a hit whereas a “yes” response to an incongruent trial was defined as a false alarm. To avoid the biasing effect of extreme hit or false-alarm proportions of zero or one on $d'$ values, we adopted the strategy of adding 0.5 to all data cells in the two-by-two contingency table [67, 68].

The values of $d'$ were submitted to a two-way repeated-measures ANOVA with group (amusics, controls) as a between-subjects factor and pitch direction (upward, downward) as a within-subjects factor. Neither the main effect of pitch direction, $F(1, 20) = 0.92, p = .349$, $\eta_p^2 = .04$, nor the interaction between group and pitch direction, $F(1, 20) = 1.15, p = .296$, $\eta_p^2 = .05$, was significant. Therefore, the data were collapsed across pitch direction. The two-sample $t$-test showed that controls ($M = 2.84, SD = 0.99$) performed better than amusics ($M = 0.85, SD = 0.77$), $t(20) = -5.26, p < .001, d = -2.24, 95\% CI [-3.31, -1.14]$, on judging semantic congruency between the primes and targets.

3.2 Neuroimaging results

As can be seen in Table 2, when comparing incongruent with congruent trials, amusics showed more activation in the left extra-nuclear and right sub-gyral, caudate nucleus, and the brainstem. Controls, by contrast, only showed increased activation in the left supplementary motor area (Boersma, #3817). SMA is a relevant area with regard to
sensorimotor interactions. A possible interpretation for this activation is that it may be a
stimulus driven motor representation rather than an intention driven representation.

Table 2  Differential brain activations from the within-group and between-group analyses
(incongruent > congruent)

<table>
<thead>
<tr>
<th>Region of activation</th>
<th>BA</th>
<th>No. of voxels</th>
<th>MNI coordinates</th>
<th>z score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>y</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Amusics</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R Sub-gyral</td>
<td>20</td>
<td>24</td>
<td>-18</td>
<td>27</td>
<td>3.91</td>
<td></td>
</tr>
<tr>
<td>R Caudate nucleus</td>
<td>18</td>
<td>-18</td>
<td>21</td>
<td>3.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L Extra-nuclear</td>
<td>10</td>
<td>-6</td>
<td>-36</td>
<td>3</td>
<td>3.69</td>
<td></td>
</tr>
<tr>
<td>R Brainstem</td>
<td>3</td>
<td>-33</td>
<td>0</td>
<td>3.49</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Controls</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>L Supplementary motor area</td>
<td>6</td>
<td>11</td>
<td>-3</td>
<td>3</td>
<td>57</td>
<td>4.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Controls &gt; Amusics</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>L Supramarginal gyrus*</td>
<td>40</td>
<td>19</td>
<td>-48</td>
<td>-45</td>
<td>30</td>
<td>4.05</td>
</tr>
<tr>
<td>L Inferior frontal gyrus*</td>
<td>45</td>
<td>21</td>
<td>-48</td>
<td>27</td>
<td>15</td>
<td>3.70</td>
</tr>
<tr>
<td>L Inferior frontal gyrus*</td>
<td></td>
<td></td>
<td>-51</td>
<td>33</td>
<td>12</td>
<td>3.50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Amusics &gt; Controls</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>no significant clusters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. L = left; R = right. * Small volume-corrected ps = .036 (FDR corrected).

When comparing the incongruent > congruent contrast between groups, we found that
controls yielded a greater activation in the left supramarginal gyrus/IPL and IFG than
amusics (see Fig. 2). No activated region was found in amusics relative to controls.
The results from the opposite comparison of congruent > incongruent trials are presented in Table 3. For the within-group analysis, neither amusics nor controls showed more clusters for congruent than incongruent trials. These results may be explained by fMRI adaptation, according to which unrelated stimuli elicit enhanced neural activation compared to related stimuli while related stimuli reduce BOLD responses relative to unrelated or novel stimuli [69, 70]. For the between-group analysis, amusics showed more clusters in the left IPL and IFG for congruent than incongruent trials compared to controls. In other words, amusics showed reduced BOLD responses to incongruent trials in the left IPL and IFG, while controls demonstrated increased responses to incongruent trials in these regions. These results were consistent with those obtained by comparing incongruent with congruent trials in Table 2.

Table 3 Differential brain activations from the within-group and between-group analyses (congruent > incongruent)

<table>
<thead>
<tr>
<th>Region of activation</th>
<th>MNI coordinates</th>
<th>BA</th>
<th>No. of voxels</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>z score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Amusics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no significant clusters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Controls</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no significant clusters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Controls &gt; Amusics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no significant clusters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Amusics > Controls**

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>L Supramarginal gyrus*</td>
<td>40</td>
<td>19</td>
<td>-48</td>
<td>-45</td>
<td>30</td>
</tr>
<tr>
<td>L Inferior frontal gyrus*</td>
<td>45</td>
<td>21</td>
<td>-48</td>
<td>27</td>
<td>15</td>
</tr>
<tr>
<td>L Inferior frontal gyrus*</td>
<td>-51</td>
<td>33</td>
<td>12</td>
<td>3.50</td>
<td></td>
</tr>
</tbody>
</table>

* Note. L = left; R = right. * Small volume-corrected $ps = .036$ (FDR corrected).

We also analyzed the correlations among MBEA scores, accuracy in congruency judgment and the associated brain activations, but not the correlations between pitch thresholds and these measures, because the pitch distances between the first and last notes of the melodic sequences exceeded participants’ thresholds for pitch perception. For amusics, neither the MBEA scores nor the behavioral measures of performance were significantly correlated with the beta values of the IPL ($ps > .555$) or IFG ($ps > .504$).

Similarly, none of the correlations was significant for controls (all $ps > .200$). The lack of correlation between these measures might be because we had a limited sample of participants with homogeneous characteristics, which might have underestimated the true degree of relationship between these variables [71, 72]. When the data from the two groups were pooled, all correlations were significantly positive ($r(20) > .44$, $ps < .042$), due to the between-group differences.

**4 Discussion**

Using fMRI, we investigated the neural correlates underlying the understanding of musical meaning that derives from the imitation of natural phenomena in our environment. While amusic individuals showed a dysfunctional HMS in the understanding of musical meaning, normal individuals showed significantly increased activations in the left IPL and
the left IFG during comprehension of musical meaning. These findings reveal for the first
time the HMS as the prime agent for understanding musical meaning that derives from
imitation and extend its role to abstract representation of meaning in music.

The first main finding of our study is that the left HMS is activated in normal
individuals during the understanding of musical meaning that derives from the imitation of
phenomena. In the HMS, the IPL is mostly implicated in motoric descriptions of actions [19,
20, 73], while the IFG is more concerned with the goal or intention of actions [74-76]. In
our study, when listening to an upward or a downward melody, the listeners needed to map
the kinematic pattern of the heard pitch movement onto an internal motor representation of
the same movement. Consistent with the IPL’s role in sensorimotor transformation [77-79],
we observed stronger BOLD responses to incongruent than congruent trials in the IPL. In
addition, the ascending and descending melodic sequences in the present study cued the
intentions of upward and downward movements (e.g., going upstairs or downstairs). To
judge the semantic congruency between the picture and the melody, participants had to infer
the intentions of the movements of the primes and targets. Therefore, the activation of the
IFG in normal individuals reflected the understanding of these intentions.

Although the PMC is often activated during imitation [23, 80-82], no significant
activation in the PMC was observed in our study. However, after using a more liberal
threshold of uncorrected \( p < .001 \) and cluster size of \( k = 0 \) voxels, we observed activity
changes in the middle frontal gyrus \( (x = -48, y = 6, z = 54, z \text{ score } = 3.37, p < .001, k = 1, \)
BA 6) and supplementary motor area \( (x = -3, y = 6, z = 57, z \text{ score } = 3.26, p = .001, k = 2, \)
BA 6). There are two possible interpretations for the less robust activation of the PMC in
the present study. One possibility may lie in the fact that our participants were not familiar
with the experimental stimuli and task. Indeed, previous studies have reported that the PMC and supplementary motor area are significantly more active for familiar than unfamiliar stimuli, such as music [83-86] and visual pictures [87, 88]. In the present study, although walking up or down the stairs as depicted in the target pictures could be considered as familiar actions, the melodic sequences that were composed specifically for the current experiment could be unfamiliar to our participants. This might have resulted in the nonsignificant activation of the PMC in our study. Future studies are required to further explore this possibility. Another possibility is related to the modality of stimulus presentation. It has been suggested that the premotor activity is based on unimodal (visual, auditory, or tactile) attention shifts [89-92], which are tightly coupled to the motor program [93]. However, our study involved crossmodal attention shifts from auditory cues to visual targets, which resulted in decreased activation of the PMC. Indeed, behavioral research has confirmed that the processing of unimodal auditory cues and targets is coupled to the motor program, which is not the case for crossmodal auditory cues and visual targets [94].

The activation of the HMS is confined to the left hemisphere in our study, which is consistent with previous evidence that action sounds are associated with the activation of the left HMS [32, 95, 96]. However, it is contradictory to the finding that bilateral HMS is involved during action observation and imitation [21, 97]. This discrepancy may be attributed to the difference in stimuli across different studies. The stimuli used in the present study were both auditory and visual stimuli (i.e., the musical primes and target pictures), whereas the stimuli in those previous studies [21, 97] were visual stimuli only. It has been shown that the left hemisphere of the human brain has a multimodal (visual, auditory) HMS [19] and may code actions at a more abstract level or in an amodal manner [98], while the
right hemisphere only has a visual HMS [19]. Thus, the activation of the left HMS in our study may reflect the multimodal (visual, auditory) stimuli used.

The second main finding of our study is that the HMS is involved in the abstract representation of musical meaning. Previous fMRI data have indicated that the HMS is implicated in the processing of action-related information, such as executing actions [22], perceiving actions [e.g., 23, 28, 35], and imagining actions [e.g., 22, 25]. Some evidence also suggests that a mirror-like system is implicated in producing musical stimuli [43-45]. Our study extends these results by suggesting the role of the HMS in understanding meaning in music. Particularly, in the present study, each melody with an upward or a downward pitch direction was paired with a picture with an upward movement and another with a downward movement. In order to judge whether the meaning of the target picture was congruent or incongruent with the meaning of the prime, the participants had to extract the meanings from both the prime and the target. Thus, the activation of the HMS reflected the extraction of abstract meaning. On the other hand, what we manipulated was the semantic relations between musical primes and target pictures. This design ruled out the possibility that the HMS was activated by musical primes or target pictures alone. Indeed, the IPL and IFG share cognitive functions, especially in semantic integration [99-101], possibly due to the anatomical connections between the two [102, 103]. Consequently, the activation of the HMS reflected semantic integration processes. In short, the HMS in the present study is involved in the understanding of abstract meaning in music.

The third main finding of our study is that unlike normals, amusics demonstrated a dysfunctional HMS in the understanding of musical meaning, which is in agreement with our prior EEG study that reported the absence of an N400 effect in processing musical
meaning in amusia [47]. The failure to derive musical meaning from pitch change direction confirmed that amusia is indeed a pitch-related disorder, including pitch perception [e.g., 50, 51, 104], tonal functional cognition [105-107], and emotion perception [108-111]. Although impaired pitch discrimination and memory in amusia is related to abnormal activity changes in the right IFG as well as decreased functional connectivity between the right IFG and the right auditory cortex [112-114], our study further indicates that impaired musical meaning understanding in amusia is associated with aberrant brain activity in the left parietofrontal cortices. These results imply that functional anomalies in the amusic brain may depend on task demands. Moreover, given that the HMS is an innate endowment [e.g., 115, 116], our findings confirm the assumption that the pitch-based deficit of amusia is likely to be present at birth [117, 118]. Indeed, it has been suggested that amusia has a genetic basis [119, 120]. From this perspective, one could speculate that the brain function for understanding musical imitation may be hardwired in the brain of neurotypical individuals, but not in that of amusic individuals.

In sum, the present data show that the HMS plays an important role in understanding musical meaning that derives from the imitation of phenomena, providing new insight into the neural representation of musical meaning in the HMS. It is worth noting that we used the direction of pitch change to imitate an upward or a downward movement, which reflects a creative intention. Interestingly, Steinbeis and Koelsch [15] manipulated experimental instructions (that is, whether participants were informed that the piece of music was written by a composer or generated by a computer), and found that understanding the composer’s intentions activated the anterior medial frontal cortex, superior temporal sulcus/middle temporal gyrus/IPL, temporal pole, IFG, and middle occipital gyrus in listeners. Therefore,
it is reasonable to assume that the HMS may be involved in the understanding of musical meaning, regardless of whether they are derived from imitation or not. Taken together, our findings provide functional imaging evidence in support of the mimesis theory that music is an imitation of phenomena, and imitation-induced musical meaning can be readily understood by normal listeners [12-14].
Acknowledgements

The authors thank Dr. Lisa Aziz-Zadeh, Xiaochen Tang, Dr. Haijiang Li, and Dr. Shengchuang Feng for their valuable suggestions on an earlier version of this manuscript. This work was supported by the National Natural Science Foundation of China (grant number 31470972 to CJ and FL, and grant number 31500876 to LZ and JJ.), and the European Research Council Starting Grant (CAASD, No. 678733) to FL and CJ.
1 Declarations of interest

2 None.
1 References


[99] T.-L. Chou, J.R. Booth, D.D. Burman, T. Bitan, J.D. Bigio, D. Lu, N.E. Cone,
Developmental changes in the neural correlates of semantic processing, Neuroimage

[100] T.-L. Chou, C.-W. Chen, M.-Y. Wu, J.R. Booth, The role of inferior frontal gyrus and
inferior parietal lobule in semantic processing of Chinese characters, Exp. Brain Res.

evidence for bilateral fronto-parietal involvement in language processing, Brain Lang.
131 (2014) 56-64.


[103] J. Wang, L. Fan, Y. Zhang, Y. Liu, D. Jiang, Y. Zhang, C. Yu, T. Jiang,
Tractography-based parcellation of the human left inferior parietal lobule, Neuroimage

[104] X. Lu, Y. Sun, H.T. Ho, W.F. Thompson, Pitch contour impairment in congenital
amusia: new insights from the Self-paced Audio-visual Contour Task (SACT), PLoS
One 12(6) (2017) e0179252.

[105] C. Jiang, F. Liu, W.F. Thompson, Impaired explicit processing of musical syntax and
tonality in a group of Mandarin-speaking congenital amusics, Music Percept. 33(4)
(2016) 401-413.

[106] D. Omigie, M.T. Pearce, L. Stewart, Tracking of pitch probabilities in congenital


Figure captions

**Fig. 1.** Design of the semantic priming paradigm. Melodic sequences with either upward or downward pitch direction changes were used as primes and pictures depicting upward or downward movements were used as targets.

**Fig. 2.** Significant activation areas in controls compared to amusics. The color scale reflects the $t$ value for the contrast between incongruent and congruent conditions for voxels above the FDR-corrected threshold of $p < .05$; the more yellow a pixel is, the higher the $t$ value is.