

Music-experience-related and musical-error-dependent activations in the brain

Kuniyoshi L. Sakai^{1,*}, Yoshiaki Oshiba¹, Reiya Horisawa¹, Takeaki Miyamae^{2,3}, Ryugo Hayano^{3,4}

¹Department of Basic Science, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo 153-8902, Japan,

²Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA,

³Suzuki School of Music, The Talent Education Research Institute, Matsumoto-shi 390-8511, Japan,

⁴Department of Physics, School of Science, The University of Tokyo, Tokyo 113-0033, Japan

^{*}Address correspondence to Kuniyoshi L. Sakai, Department of Basic Science, Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan. Email: sakai@sakai-lab.jp

Although music is one of human-unique traits such as language, its neural basis for cortical organization has not been well understood. In the present functional magnetic resonance imaging study, we tested an error-detection task with different types of musical error (pitch, tempo, stress, and articulation conditions) and examined three groups of secondary school students having different levels of music experience. First, we observed distinct activation patterns under these music conditions, such that specific activations under the pitch condition were consistently replicated for all tested groups in the auditory areas, as well as in the left language areas under the articulation condition. Second, music-experience-related activations were observed in multiple regions, including the right sensorimotor area under the pitch condition, as well as in the right premotor cortex under the articulation condition. Indeed, the right homologs of the language areas were specifically activated under the stress and articulation conditions. Third, activations specific to the group with the highest proficiency in music were observed under the tempo condition mostly in the right regions. These results demonstrate the existence of music-related signatures in the brain activations, including both universal and experience-related mechanisms.

Key words: training; syntax; right frontal regions; music; auditory areas.

Introduction

Music performance is a highly complex task that requires precise regulation and integration of the sensorimotor system under the control of high-order cognitive functions of the human brain (Herholz and Zatorre 2012; Miyamae 2018). Moreover, active processes of music performance and perception may also involve predictive coding, which makes predictions about what could possibly follow next (Koelsch et al. 2019). Accumulating evidence suggests that musical training not only enhances various musical skills but also facilitates the interaction with other cognitive abilities, such as intelligence (Schellenberg 2004, 2006), working memory (Fujioka et al. 2006), language-related skills including literacy (Forgeard et al. 2008; Seither-Preisler et al. 2014), and speech prosody (Thompson et al. 2004). The most fundamental regions would be the auditory areas, that is, the bilateral superior temporal gyri and Heschl's gyri (STG/HG), which are naturally involved in processing tonal pitch (Hall and Plack 2009). By using songs with lyrics as stimuli, we showed that the STG/HG activations were right dominant during the detection of errors in melodies, while they were left dominant during the detection of errors in lyrics

(Yasui et al. 2009), which claim has been supported by recent works (Flinker et al. 2019; Albouy et al. 2020). Another study reported enhanced activations in both the auditory and motor regions, even when the auditory and motor tasks were separately tested in professional pianists (Bangert et al. 2006). Another potentially music-related region is the right lateral premotor cortex (LPMC) and inferior frontal gyrus (IFG), that is, the right homologs of the left language areas (Sakai 2005). The critical role of these right regions in prosodic processing has been suggested by lesion studies (Ross and Mesulam 1979), and it has been confirmed by imaging studies as well (Plante et al. 2002; Strelnikov et al. 2006). It is thus essential to focus on subcomponents of music, including pitch recognition and prosodic processing.

In the present study with functional magnetic resonance imaging (fMRI), we targeted students taking private violin lessons for about 10 years (range, 5–14 years) with the Suzuki Method (<https://internationalsuzuki.org>), which was initiated by Shin'ichi Suzuki. The Suzuki Method is a series of music education courses that has spread throughout the world, inspired by the acquisition of the mother tongue (Suzuki 2013). This method particu-

Received: September 16, 2021. Revised: November 20, 2021. Accepted: November 21, 2021

© The Author(s) 2021. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permission@oup.com.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

larly emphasizes musical training by ear and playing by heart, through listening to the recorded performances of virtuosi and outstanding musicians (Steinschaden and Zehetmair 1985). In general, notation reading and theory are introduced after establishment of a routine of listening and practicing by ear. Suzuki educator Winifred Crock stated that “Tone was an incredibly important issue for Dr Suzuki. He spoke about tone and sound at every lecture, at every lesson, and in many writings. He told of listening to Kreisler’s recordings again and again and being captivated with the sound” (Hendricks 2011, p. 143). This is why we focused on Suzuki students, who would represent ideal “native speakers” of music. We recruited two control groups of age-matched students: one with experiences mostly of the piano or keyboards, and the other with limited or no music experience at all (Supplementary Table 1). Some participants of the former control group experienced other instruments as well, because they were in a secondary school orchestra. The inclusion of students trained on the Suzuki Method, as well as an age- and exposure-controlled group not trained on the same method, is a strong and unique feature of the present study.

Based on an error-detection task in language to study underlying neural foundation of syntactic and phonological processes (Suzuki and Sakai 2003), we developed a similar error-detection task in music for the present study, with which participants detected errors or unnatural phrases of musical excerpts. We tested both natural and unnatural stimuli to realize execution of right musical judgment while measuring brain activations. By using melodies of western classical music, we tested four main conditions: pitch, tempo, stress, and articulation (Fig. 1). We anticipate differential activation patterns for each of these musical elements, because their features and aspects are clearly distinct from the following reasons. First, the “pitch” of individual tones is a fundamental element of music, in which the auditory areas would likely be involved as noted above. Second, the “tempo” (i.e., *Agogik*) plays a basic role of speed control in music, reflecting performing styles and/or emotional states. Indeed, the tempo of a musical piece is typically associated with its musical motifs and feelings, by which various regions for cognitive functions may be activated, especially for well-trained musicians who can play instruments (or sing) with creative expression and imagination. Third, the “stress” (i.e., *Dynamik*) of individual tones is similar to accent in English, in that those sounds require certain forces to be produced, and the sensorimotor regions may be activated by mimicking these forces even without producing real sounds. Finally, “articulation” further depends on phrase or syntactic structures and their interpretation in music.

Generally speaking, musical phrase structure and its interpretation play critical roles to determine not only tempo (Timmers et al. 2000) and stress but also articulation (Berry 1985) (e.g., “rhythm” as the articulation of time). Therefore, these three factors would mutually interact to produce enhancement or suppression.

Under the articulation condition, activations in the language areas would be enhanced, while some regions activated under either tempo or stress condition may be suppressed under the articulation condition. The present study primarily aimed to identify brain activation patterns under each of these music conditions and to examine how such activation patterns would be replicated between the tested groups. Any consistent and crucial activations would be eliminated by direct comparisons among groups; group-specific activation patterns were subsidiary for our purposes. It is also possible that multiple conditions share identical cognitive processes to search for a type of musical errors. Using this novel setup of experiments, we attempted to elucidate the localization of music-related systems in the brain.

Materials and Methods

Participants

A total of 107 secondary school students mostly at the age of 15 years (range, 12–17 years; see Table 1) participated in this study. First, volunteers taking violin lessons with the Suzuki Method for several years (in the middle class or upper) were recruited in the areas around Tokyo. We focused on this adolescent age, because 10 years of experience would be ideal for the acquisition of expert performance (Ericsson et al. 1993), where the typical age of acquisition (AOA; the earliest age if learned multiple instruments) was 5 years old for the Suzuki students we recruited (Table 1). The age effect, if any, was removed from MRI data analyses by adopting it as a nuisance factor (see below). The 37 Suzuki students were all right handed according to the Edinburgh Handedness Inventory (Oldfield 1971) and had no history of neurological disorders. Seventy age-matched students were also recruited from a secondary school attached to the Faculty of Education, The University of Tokyo. Due to excessive head movements during MRI data acquisition, four Suzuki students and five secondary school students were dropped. The remaining Suzuki students, designated the “Suzuki” (S) group, had early AOAs in music, that is, they had started their musical training at relatively early years ranging from 2 to 9 years old. We divided the remaining secondary school students into two groups according to their AOAs: an “early” AOA (E) group of 36 students, who started their musical training at 2–8 years of age, and a “late” AOA (L) group of 29 students, who started their musical training at 9–13 years of age or had no music experience. This division of the secondary school students was justified by the absence of a significant difference in AOAs ($t[67] = 0.8$, $P = 0.4$) between the S and E groups (see Table 1).

All participants were administered a questionnaire regarding their musical (instrument and/or vocal) training, including both private/group lessons and self-education. For each participant, the period between his/her AOA and the latest age with musical training was defined as the duration of exposure (DOE in years) in music (without duplicating an overlapped period

A Normal**B Pitch****C Tempo****D Stress****E Articulation**

Fig. 1. Examples of music stimuli taken from a piece and tested under the four main conditions. (A) An initial section from Minuet in G major (BWV Anh. 116) composed by J. S. Bach. We also used two more French pieces. From the original pieces, we changed one portion (asterisk) to make an unnatural stimulus, which was played as a whole excerpt. (B) Under the pitch condition, the pitch of a note was changed (e.g., from D to E). (C) Under the tempo condition, the tempo of a phrase was changed (e.g., faster than other phrases). (D) Under the stress condition, the stress of a phrase was changed (e.g., weaker or stronger). (E) Under the articulation condition, the articulation of a phrase was changed. For example, a music phrase was played “monotonously” without proper articulation or without a structural combination of individual musical notes.

for multiple instruments); if there was an absence from practice of more than 6 months, the period of absence was subtracted to obtain the DOE (see Table 1). Musical training as a part of the school curriculum was not included in the DOE, because it was only 40 h each year and thus had little effect on age-matched groups, except for one Suzuki student who was in a secondary

school specialized in music. Moreover, the total time spent practicing music (practice in hours) was also estimated for each participant by adding together the number of hours of lessons and practices for each of multiple instruments, if any (see Supplementary Table 1). Participants with DOE of less than 1 year were regarded as having no music experience; there were 16 such

Table 1. Three participant groups related with music experience.

Groups	N	Age (years)	LQ	AOA (years)	DOE (years)	Practice (hours)
Suzuki (S)	33 (18 f)	15 ± 0.2	84 ± 3.5	4.5 ± 0.3	10 ± 0.4	3900 ± 430
Early (E)	36 (27 f)	15 ± 0.3	82 ± 3.5	4.8 ± 0.2	7.8 ± 0.6	2400 ± 520
Late (L)	29 (14 f)	14 ± 0.3	86 ± 2.9	10 ± 0.2	1.8 ± 0.4	720 ± 280

Notes: Data are shown with the mean ± SEM. f, female; LQ, laterality quotients of handedness; AOA, age of acquisition; DOE, duration of exposure; Practice, approximate total time spent practicing music.

participants in the L group. We acknowledge that the difference in trained instruments (the violin for the S group, and the piano/keyboards for the E group) would be a potential confound, but the use of “flute” playing as stimuli, instead of violin or piano playing, was suitable to examine effects of musical training in general, because no students in the S or E group experienced playing the flute (see [Supplementary Table 1](#)).

All participants, as well as their legal guardians for those younger than 18 years of age, provided their written informed consent to participate in this study, after the nature and possible consequences of the study were explained. Approval for these experiments was obtained from the institutional review board of the University of Tokyo, Komaba Campus (No. 497-2 and 497-3). All research studies were performed in accordance with the Declaration of Helsinki, Singapore Statement on Research Integrity, and relevant guidelines/regulations in Japan (Science Council of Japan, and Japan Society for the Promotion of Science).

Stimuli

The music stimuli consisted of short excerpts from three popular pieces of classical music arranged for the flute: Minuet in G major (BWV Anh. 116) composed in 1725 by Johann Sebastian Bach, Sicillienne in G minor (Op. 78) composed in 1893 by Gabriel Fauré, and Violin Sonata in A major (FWV 8) composed in 1886 by César Franck. These pieces were played by a professional flutist, Takeaki Miyamae (an author of this article), and digitally recorded. All of the Suzuki students knew this Minuet through violin playing, but other pieces were not included in the standard Suzuki course. To reduce the group difference in familiarity level, we instructed all participants to listen to these original excerpts three times a day for 7 days to become familiarized with those pieces by the time of individual fMRI experiments. Such memory factors or familiarization effects do not interfere with the goal of the study; we expected that the brain regions required for memory retrieval, including the hippocampus, would be additionally involved.

We separately prepared “unnatural” stimuli, which consisted of the same short excerpts played and recorded as described above but with an error in pitch, tempo, stress, or articulation. [Figure 1](#) shows examples of each type of error; these errors were created by three experienced teachers (a flutist, a violinist, and a pianist) of the Suzuki Method, with consideration for various aspects of performance errors in general. For each of the

four error types, we made 10 unnatural stimuli, each containing a single error. The errors in pitch were either lower or higher notes (mean: 0.9 semitones; interval: −2 to +6 semitones), always changing a single note. The errors in tempo were either slower or faster notes, spanning a phrase of multiple notes (mean: 4.7 notes; range: 3–9 notes or one half to two bars). The errors in stress were either weaker (*p*) or stronger (*mf* or *f*) notes, spanning a phrase of multiple notes (mean: 2.2 notes; range: 1–6 notes within a bar). The errors in articulation included “staccato” instead of “legato”, “decrescendo” instead of “crescendo”, and monotonously without intonation, all of which spanned multiple notes (mean: 6.9 notes; range: 2–21 notes or within four bars). The inclusion of an error of one of these types corresponded to the four “main conditions”: the pitch, tempo, stress, and articulation conditions.

We edited the auditory stimuli with WaveLab 8 software (Steinberg Media Technologies GmbH) and extracted 10 natural and 10 unnatural stimuli of 15 s each for each condition. If the excerpt had a break point at either end, we added a 2-s fade-in and/or 2-s fade-out. The onset of each error occurred at 7.1 ± 2.5 s (averaged for all conditions), which was normally distributed within a 15-s segment. Under one main condition, we performed two scanning runs, each of which included five natural and five unnatural stimuli in a random order. The same excerpt’s natural and unnatural versions, that is, equally sampled from the original pieces, were presented to each participant also in a random order.

As a control condition for the error-detection task, we also tested the connection condition, under which we presented either a single natural excerpt or two excerpts from different pieces. The latter two excerpts were artificially “connected” at similar notes (without obvious break points) just like a single piece (for 15 s in total), in which musical flow or context was clearly disrupted at the break points. The connection condition required familiarization with the musical pieces used, thus controlling memory retrieval demand of the pieces under the main conditions for each individual. This condition also controlled the basic auditory processes for the musical pieces, as well as error detection and associated decision-making, which enabled a more focused contrast between each of the main conditions and the connection condition used for fMRI analyses. In our previous fMRI study ([Suzuki and Sakai 2003](#)), we clarified that the activations selective to syntactic judgments were observed equally for normal and anomalous (i.e., grammatical

and ungrammatical) sentences. Therefore, we focused on “musical judgments” themselves by utilizing this control condition, rather than comparing the responses to errors with those to normal excerpts.

During the scans, the participants wore an MRI-compatible headphone, VisuaStim Digital (Resonance Technology Inc.), a pair of earmuffs (3M Peltor), and a pair of earplugs (Earasers, Persona Medical) to reduce the high-frequency noises (>1 kHz) of the scanner. The stimulus presentation was controlled by the Presentation software package (Neurobehavioral Systems), with which behavioral data (accuracy and response times [RTs]) were obtained. Before scanning, we appropriately adjusted the sound level for each participant by presenting the first 20 s of the Minuet.

Task

Under the five conditions (pitch, tempo, stress, articulation, and connection) described above, we used an error-detection task to require participants to detect errors as unnatural stimuli, or as artificially connected excerpts. This task actually entails the correct identification of original musical pieces as well as correct judgment of musical expressions under each of the main conditions. In each trial, participants listened to an excerpt for 15 s and judged whether there was an unnatural portion in the excerpt. After a brief beep (0.2 s, at the pitch of A4) following the excerpt, participants pushed one of two buttons (one for natural and the other for unnatural) within 2 s. The three pieces of music were presented randomly, and the same stimuli never appeared twice under any condition.

Before each scanning, we informed participants of the name of a “main condition” (e.g., pitch) to be tested in the next run. The rationale for this information was to set specific “musical judgments” to the participants, and the effect of an expectation to hear a specific type of error was to enhance specific cognitive and attentional processes, just like syntactic or phonological judgments (cf. Suzuki and Sakai 2003), in the specified cortical/subcortical regions. One scanning run consisted of 10 trials under a main condition as well as of 3 trials under the connection condition (e.g., in the order of connection–connection–pitch–pitch...pitch–connection), where the connection trials included one or two artificially connected excerpts at the same rate. To distinguish the connection trials from the main trials, we presented a brief beep (0.2 s; at the pitch of G5) before presenting an excerpt only under the connection condition. Following four runs, each with one of the four main conditions, the participants took a rest for about 10 min before proceeding to four more runs in a day. To exclude any priming or order effects of conditions, we tested the main conditions in one of two orders: tempo–articulation–stress–pitch–articulation–tempo–pitch–stress or stress–pitch–tempo–articulation–pitch–stress–articulation–tempo.

MRI Data Acquisition and Analyses

The following methods followed the procedures published previously by our team (Ohta et al. 2017; Tanaka et al. 2017, 2019). For the MRI data acquisition, the participant was in a supine position, and his or her head was immobilized inside the radiofrequency coil. The MRI scans were conducted on a 3.0T system, GE Signa HDxt 3.0 T (GE Healthcare). We scanned 30 axial slices, each having a thickness of 3 mm and a gap of 0.5 mm, covering the volume range of -38.5 to $+66$ mm from the anterior to posterior commissure (AC–PC) line in the vertical direction, using a gradient-echo echo-planar imaging (EPI) sequence (repetition time [TR] = 2 s, echo time [TE] = 30 ms, flip angle [FA] = 78° , field of view [FOV] = 192×192 mm², resolution = 3×3 mm²). In a single run, we obtained 112 volumes following four dummy images, which allowed for the rise of the MR signals. After completion of the fMRI session, high-resolution T_1 -weighted images of the whole brain (136 axial slices, $1.0 \times 1.0 \times 1.0$ mm³) were acquired from all participants with a three-dimensional fast spoiled gradient recalled acquisition in the steady state (3D FSPGR) sequence (TR = 8.5 ms, TE = 2.6 ms, FA = 25° , FOV = 256×256 mm²). These structural images were used for normalizing fMRI data.

The fMRI data were analyzed in a standard manner using SPM12 statistical parametric mapping software (Wellcome Trust Center for Neuroimaging, <http://www.fil.ion.ucl.ac.uk/spm/>) (Friston et al. 1995), implemented on MATLAB (Math Works). The acquisition timing of each slice was corrected using the middle slice (the 15th slice chronologically) as a reference for the EPI data. We realigned the time-series data in multiple runs to the first volume in all runs. The realigned data were resliced every 3 mm using seventh-degree B-spline interpolation, so that each voxel of each functional image matched that of the first volume. We removed runs that included data with a translation of >2 mm in any of the three directions and with a rotation of $>1.4^\circ$ around any of the three axes; we dropped nine participants (see above) whose head movement exceeded these thresholds for both runs of a condition.

After alignment to the AC–PC line, each participant's T_1 -weighted structural image was coregistered to the mean functional image generated during realignment. The coregistered structural image was spatially normalized to the standard brain space as defined by the Montreal Neurological Institute (MNI), using the “unified segmentation” algorithm with light regularization, which is a generative model that combines tissue segmentation, bias correction, and spatial normalization in the inversion of a single unified model (Ashburner and Friston 2005). After spatial normalization, the resultant deformation field was applied to the realigned functional imaging data. All normalized functional images were then smoothed by using an isotropic Gaussian kernel of 9 mm full width at half maximum. Low-frequency noise was removed by high-pass filtering at 1/128 Hz.

In a first-level analysis (i.e., the fixed-effects analysis), each participant's hemodynamic responses induced by the four main conditions (pitch, tempo, stress, and articulation) as well as the connection condition in each run were modeled with a boxcar function with a duration of 13 s, excluding 1 s each from both ends of one trial. The interval between trial events was 4 s, minimally allowing for separation of the hemodynamic responses to the different trials. The boxcar function was then convolved with a hemodynamic response function. To minimize the effects of head movement, the six realignment parameters obtained from preprocessing were included as a nuisance factor in a general linear model. The images under each of the five conditions were then generated in the general linear model for each participant and used for the intersubject comparison in a second-level analysis (i.e., the random-effects analysis) with a flexible factorial option. Other nuisance factors were age, laterality quotients of handedness (LQ), and gender. To examine activated regions in an unbiased manner, we adopted whole-brain analyses.

A repeated-measures analysis of variance (rANOVA) with *t* tests was performed with two factors (participant groups \times conditions), the results of which were thresholded at uncorrected $P < 0.001$ for the voxel level, and at corrected $P < 0.05$ for the cluster level, with family-wise error (FWE) correction across the whole brain. We assumed equal variances among participants (independent factor), participant groups (independent factor), and conditions. For the anatomical identification of activated regions, we basically used the Anatomical Automatic Labeling method (<http://www.gin.cnrs.fr/en/tools/aal/>) (Tzourio-Mazoyer et al. 2002). We extracted the mean percent signal changes of each participant from the local maxima of each region, using the MarsBaR-toolbox (<http://marsbar.sourceforge.net/>).

To justify sample sizes in relation to what has been previously published, we performed a power analysis (Mumford 2012) by using our previous work with a similar error-detection task (grammatical or ungrammatical) with speech sounds (Umejima et al. 2021). We used the whole-brain data from a group of 15 participants (Bilingual group, G2 condition, "Sentence – Lexical list" contrast; see Umejima et al. 2021), according to a previous pilot study (Durnez et al. 2016). By setting a specific set of values (80% power for a two-tailed $\alpha = 0.05$ test) (Lehr 1992), we used an online estimator of Neuropower (<http://www.neuropowertools.org/neuropower/neuropowerstart/>), and obtained 23 subjects for this value of average power (FWE corrected $P < 0.05$). All of the three participant groups in the present study satisfied this requirement (see Table 1).

Results

Behavioral Data

Figure 2 shows the population distribution regarding AOAs. We tested three age-matched groups with distinct

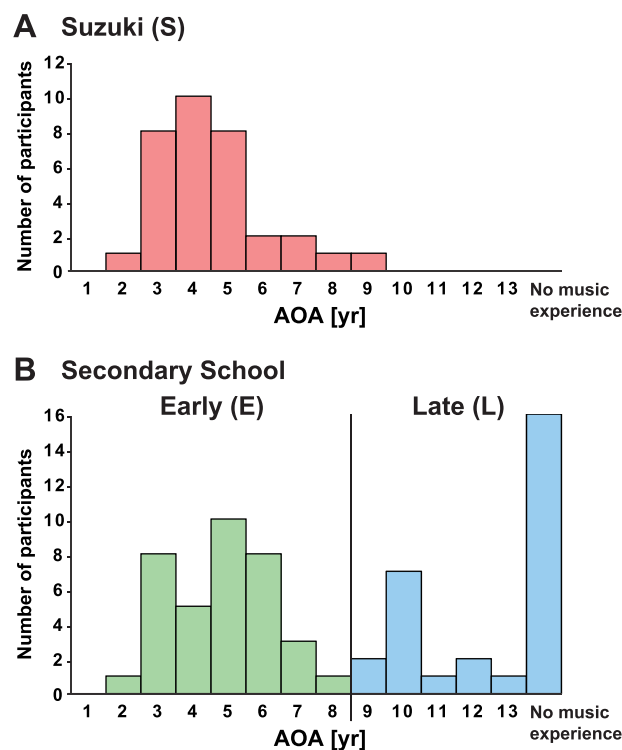


Fig. 2. Population distribution regarding the ages of acquisition (AOAs). Histograms of AOAs are separately shown for the Suzuki (S, in red) (A) and Secondary School students (B). The latter students were divided into the Early AOA (E, in green) and Late AOA (L, in blue) groups.

levels of music experience: the Suzuki (S), early AOA (E), and late AOA (L) groups (Table 1). In our survey before all the testing, the participants had reported levels of familiarity with the musical pieces by rating whether each piece was known (= 2), familiar (= 1), or unfamiliar (= 0). The combined familiarity ratings of the three pieces (max = 6, min = 0) for the S, E, and L groups were 2.9 ± 0.2 , 1.8 ± 0.2 , and 1.2 ± 0.3 (mean \pm standard error of the mean [SEM]), respectively, indicating that the ratings for the S group were slightly (about the score of 1), but significantly, higher than the other groups ($P < 0.0001$), while those for the E and L groups were not significantly different ($P > 0.1$). With respect to the total time spent practicing music, the S group was trained most well, while the E group was less trained and the L group had limited or no experience in music. Moreover, the DOE (in years) of the S group was significantly longer than that of the E group ($t[67] = 3.3$, $P < 0.05$), even when the AOAs of the S and E groups were arranged to be equal to one another (see Fig. 2) and were equally earlier than the AOA of the L group ($P < 0.0001$). The group differences with regard to behavioral data and brain activations were thus considered to reflect primarily the amounts of music experience or musical training methods.

To examine any differences in the groups or those under the conditions at the behavioral level, we compared the accuracy and RTs for the error-detection task (Fig. 3). Even if an error occurred in the middle of the

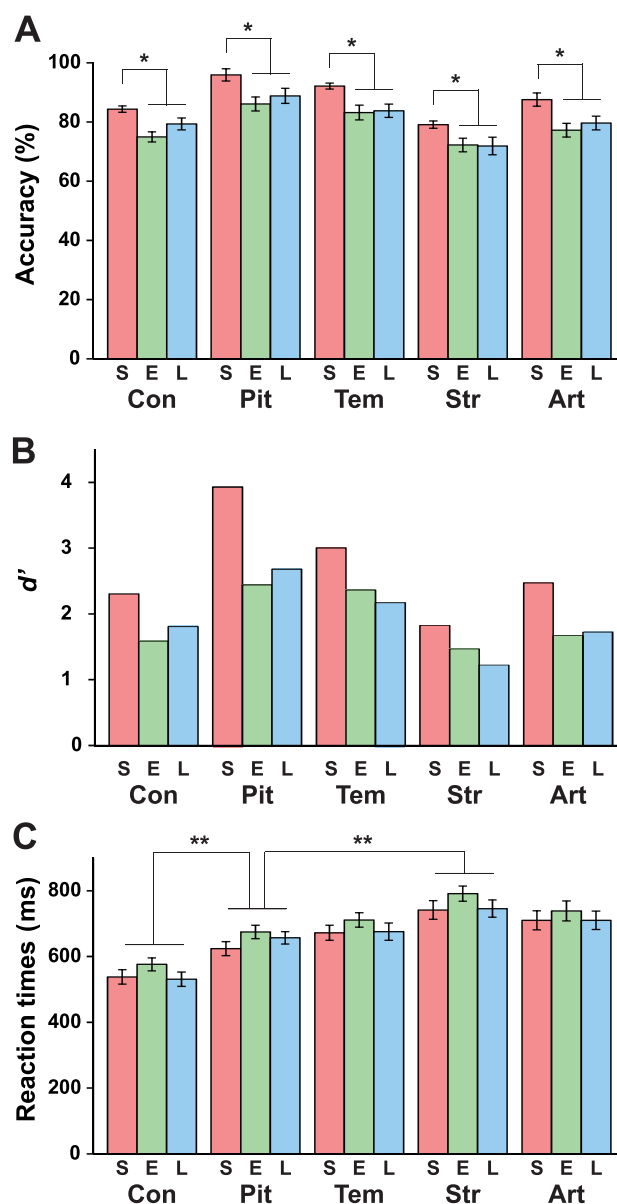


Fig. 3. Behavioral results for the connection and four main conditions. Histograms are shown for the accuracy (A), d' (B), and response times (RTs) (C). Note that the S group showed significantly higher accuracy and d' than the E and L groups under all conditions. Error bars indicate the SEM for the participants. *Holm-corrected $P < 0.05$; **Corrected $P < 0.0001$.

excerpts, some difficult ones required more time to decide presence or absence of an error, further increasing RTs; indeed, we observed clear condition differences (Fig. 3C). For the main fMRI analyses, we included all trials, regardless of successful detection of errors or not, thus reflecting differences in behavioral data, but we also examined the runs with higher accuracy ($\geq 75\%$) in subsequent analyses.

We performed two-way rANOVA (group \times condition), and the accuracy data showed both significant main effects of group ($F[2, 380] = 8.7$, $P < 0.001$) and condition ($F[4, 380] = 40$, $P < 0.0001$) without an interaction of group by condition ($F[8, 380] = 0.4$, $P = 0.9$) (Fig. 3A). Under all conditions, the S group showed significantly higher

accuracy than the E and L groups (connection: $t[96] = 2.6$, Holm corrected $P < 0.05$; pitch: $t[96] = 3.4$, corrected $P < 0.005$; tempo: $t[96] = 3.1$, corrected $P < 0.05$; stress: $t[96] = 2.5$, corrected $P < 0.05$; articulation: $t[96] = 3.7$, corrected $P < 0.005$). These results demonstrate that the S group had the highest proficiency in music. Because the pitch and tempo conditions almost reached ceiling effects, we obtained d' from a Z value of the hit rate (correct detection of “unnatural” stimuli) minus that of the false-alarm rate (incorrect responses to “natural” stimuli), thereby removing no-response trials under each condition. The resultant d' for each group closely replicated the accuracy data with more distinct dominance of the Suzuki group (Fig. 3B), and all of the d' values were higher than 1 (i.e., significantly above chance).

Regarding the RTs, there was a significant main effect of condition ($F[4, 380] = 98$, $P < 0.0001$), without a main effect of group ($F[2, 380] = 1.1$, $P = 0.4$) or their interaction ($F[8, 380] = 0.4$, $P = 0.9$) (Fig. 3C). Among the four main conditions, the RTs were shortest under the pitch condition and longest under the stress condition ($t[97] = 11$, corrected $P < 0.0001$), consistent with the difference in accuracy under these two conditions ($t[97] = 11$, corrected $P < 0.0001$). These results suggest that the task was easiest under the pitch condition, probably due to the obvious changes in melody, whereas under the stress condition, the auditory stimuli were more susceptible to scanner noises during MRI data acquisition.

Under the connection condition, the RTs were significantly shorter than under the easiest pitch condition ($t[97] = 9.0$, corrected $P < 0.0001$), while the accuracy was significantly lower under the connection condition than that under the pitch condition ($t[97] = 8.4$, corrected $P < 0.0001$). This result may have been due to participants having difficulty distinguishing the two French pieces, which were less familiar than the Minuet for the participants. Nevertheless, the shortest RTs under the connection condition for all three groups were consistent with its most fundamental nature.

Brain Activation Patterns Under Each Music Condition

We first contrasted each of the four main conditions (pitch, tempo, stress, and articulation) with the connection condition (i.e., pitch – connection, etc.) separately for the S, E, and L groups (Fig. 4). Among the main conditions, overall brain activations were most extensive under the pitch condition for the S and E groups, in spite of the fact that this condition was the easiest one, thus excluding the possible involvement of task difficulty or other domain-general factors from the activated regions. Figure 4A shows the bilaterally enhanced activations for the S and E groups, including the left LPMC (L. LPMC) and right precentral/postcentral gyrus (R. PrCG/PoCG; the sensorimotor area), as well as the left supramarginal gyrus (L. SMG); the R. PrCG/PoCG region is located just posterior to the L. LPMC. Moreover, the L. IFG activation

Table 2. Regions identified by consistent clusters of activation for all groups (S/E/L) or S/E.

Brain regions	BA	Side	S group				E group				L group			
			x	y	z	Z	x	y	z	Z	x	y	z	Z
Articulation-connection														
LPMC	6/8	L	-48	-4	38	4.0					-57	8	32	3.4
			-54	5	20	4.3	-48	5	17	5.1				
F3op	44	L	-39	8	29	4.2	-39	2	26	5.5	-48	2	23	4.0
F3t	45	L	-54	29	20	5.5					-48	23	20	4.1
			-42	35	2	4.4	-48	32	11	5.2	-45	32	11	4.8
LPMC	6/8	R	54	8	41	5.1	48	5	29	5.2				
Stress-connection														
LPMC	6/8	R	57	8	38	4.5	57	2	29	4.1	60	8	20	5.5
F3op	44	R	57	8	20	3.8								
PrCG	4	R	48	-1	47	4.0	45	-16	44	3.8	48	-4	41	5.3
Insula		R									36	-4	8	4.0
PoCG	43	R					63	-7	23	4.0	66	-4	20	4.5
											66	-16	29	4.0
SMG	40	R	63	-22	26	4.5	66	-19	32	3.9	57	-25	26	4.3
STG	22	R					66	-31	17	3.3				
Pitch-connection														
SMA	6	M	9	-22	56	4.8	12	-28	50	4.2				
PrCG	4/6	L	-51	-4	35	5.4	-45	-7	44	4.9				
F3op	44	L					-36	2	29	4.4				
			-60	5	17	4.5	-57	8	11	3.7				
SMG	40	L	-54	-40	23	4.9	-60	-37	32	3.8				
			-63	-31	29	3.8								
HG	41/42	L	-39	-37	17	5.8	-39	-37	11	6.0				
STG/HG	22/42	L	-48	-19	5	6.7	-51	-16	2	7.2	-51	-19	5	6.6
MTG	37	L	-48	-61	2	3.3	-48	-61	2	4.6				
SOG	19	L	-18	-85	23	4.2	-21	-79	26	3.7				
MOG/LiG	18/19	L	-48	-76	11	3.7								
			-21	-82	8	3.8	-21	-70	2	3.8				
PrCG	4	R	30	-16	50	4.9	27	-22	53	3.5				
			45	-7	47	5.7	45	-10	50	4.7				
Insula		R	36	-13	20	3.9								
PoCG	3/43	R	24	-28	56	4.0	36	-34	56	3.3				
			66	-10	26	3.6	54	-4	20	3.3				
SMG	40	R					66	-22	29	4.0				
STG/HG	22/42	R	54	-13	5	7.3	51	-10	2	7.7	51	-13	2	6.2
			45	-31	11	6.2								

Notes: Stereotactic coordinates (x, y, z) in the MNI space are shown for activation peaks of Z values (see Figs 4 and 5). Uncorrected $P < 0.001$ for the voxel level, FDR corrected $P < 0.05$ for the cluster level. For the results of the tempo – connection contrast, see [Supplementary Table 2](#). BA, Brodmann's area; L, left; M, medial; R, right; F3op/F3t, opercular/triangular parts of the inferior frontal gyrus; HG, Heschl's gyrus; LiG, lingual gyrus; SOG/MOG, superior/middle occipital gyrus; PoCG, postcentral gyrus; PrCG, precentral gyrus; SMA, supplementary motor area; SMG, supramarginal gyrus; STG/MTG, superior/middle temporal gyrus.

was observed for the E group alone. Although bilateral activations were observed for the S group in the post middle temporal gyrus (post-MTG) and middle occipital gyrus/calcarine sulcus (MOG/Calc.), that is, the visual areas, even without any visual stimuli, these activations were left lateralized for the E group.

It is most striking to note that the bilateral STG/HG, that is, the auditory areas, were activated consistently for all three groups in the pitch – connection contrast ([Table 2](#), [Supplementary Fig. 1](#)), and that the L group, the group with the least musical training, showed activations in these regions alone. Moreover, the supplementary motor area (SMA) showed similarly significant activations under the pitch condition for the S and E groups, and those activations were more enhanced for the S group ([Fig. 5A](#), [Table 2](#)).

Under the tempo condition (the tempo – connection contrast), in marked contrast, “only” the S group showed

significant activations. The activated regions were mostly the left auditory areas (L. STG/HG) and right lateral regions: R. PrCG/PoCG, R. STG/MTG, R. MOG/Calc., and right insula ([Fig. 4B](#), [Supplementary Table 2](#)). In addition, some more medial regions were included: the SMA, right putamen, and right amygdala/hippocampus (R. Amyg/HC) ([Fig. 5B](#)).

Under the stress condition (the stress – connection contrast), activations were consistently observed in the R. PrCG/PoCG and R. LPMC for all groups ([Fig. 4C](#), [Table 2](#)). The L. LPMC was activated for the E and L groups, while the L. IFG was activated for the E group alone, which matched the activation patterns observed under the pitch condition ([Supplementary Table 3](#)). The R. MOG/Calc. was activated for the E group, while the bilateral post-MTG regions were activated for the L group alone. It is interesting to note that activations were limited to the right frontal/parietal regions for the S group.

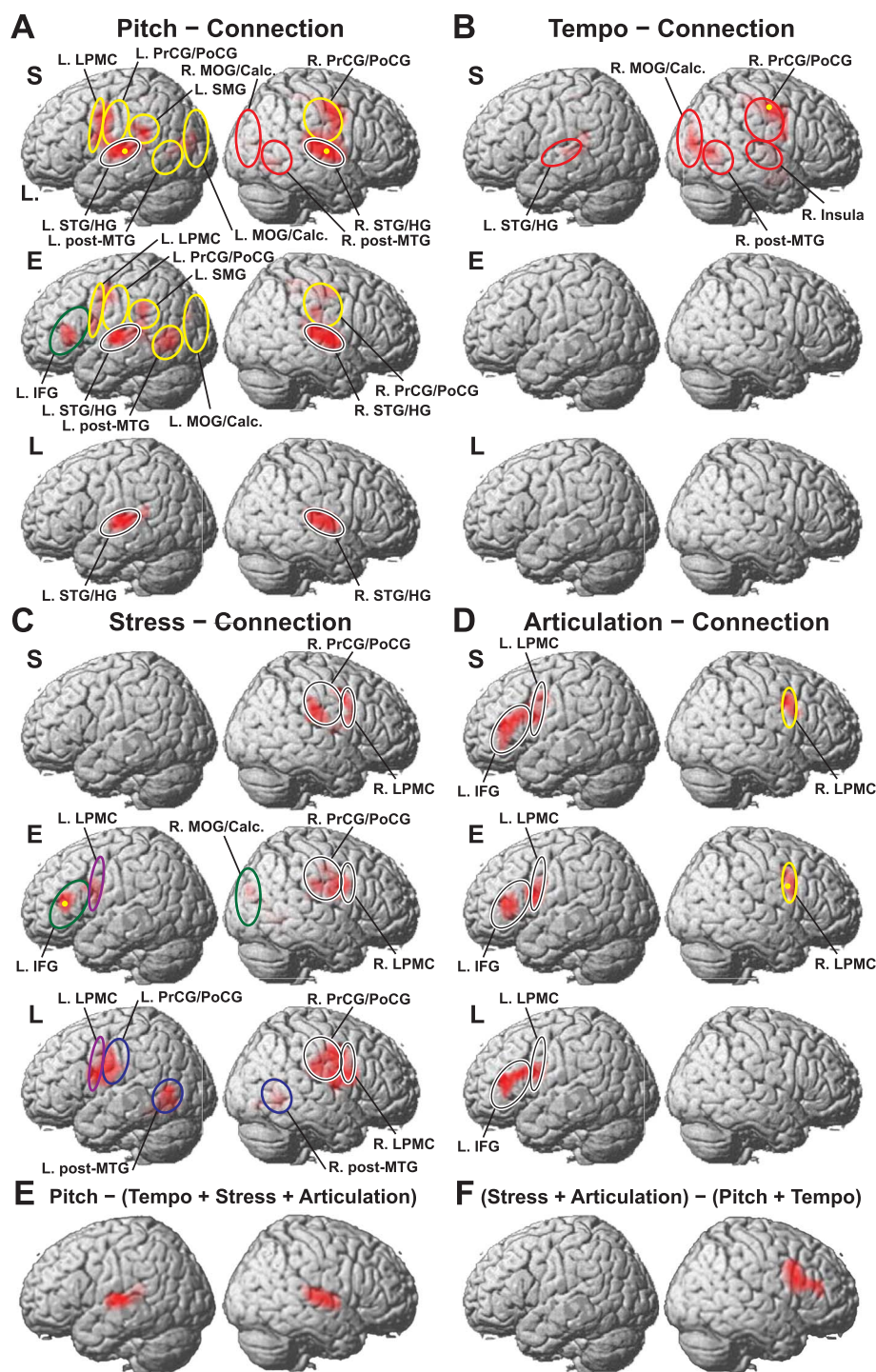


Fig. 4. Brain activation patterns under each music condition. Cortical regions identified by pitch (A), tempo (B), stress (C), and articulation (D) conditions, each of which was contrasted with the connection condition. Activations were projected onto the left (L.) and right lateral surfaces of a standard brain (FWE-corrected $P < 0.05$). Group-specific regions are circled in red, green, and blue for the S, E, and L groups, respectively. Regions circled in black, yellow, and purple are for the S/E/L, S/E, E/L combinations for consistent activation, respectively (see [Supplementary Fig. 1](#)). Direct contrasts of pitch – (tempo + stress + articulation) (E) and (stress + articulation) – (pitch + tempo) (F) are also shown for the S/E/L groups (S, E, and L combined) and S/E groups (S and E combined), respectively (FWE-corrected $P < 0.05$). For the former contrast, an exclusive mask of negative activation was applied (uncorrected $P < 0.001$).

Under the articulation condition (the articulation – connection contrast), the L. LPMC and L. IFG, that is, the language areas in the left frontal cortex, were consistently activated for all groups (Fig. 4D, Table 2).

Additional activations were observed in the R. LPMC consistently for the S and E groups, but not for the L group (Table 2). Under any conditions, there was no region activated for the S and L groups (i.e., two extremes) alone.

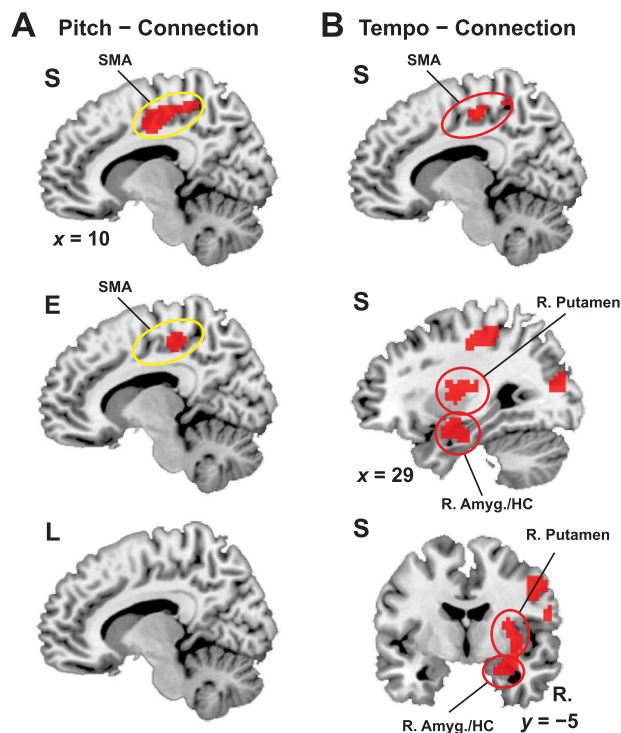


Fig. 5. Condition-specific activations in medial regions of the brain. (A) Activations under the pitch condition were projected onto a parasagittal section ($x = 10$) of the standard brain. The SMA was activated for the S and E groups, but not for the L group (FWE-corrected $P < 0.05$). (B) Activations under the tempo condition were projected onto the parasagittal sections ($x = 10$ and $x = 29$) and a coronal section ($y = -5$), showing significant activations in the SMA, R. Putamen, and R. Amyg./HC for the S group alone.

Consistency and Differences in Cortical Activations across Groups and Conditions

Given these patterns of activations under the main conditions, we further examined condition-specific activations without using the connection condition. A direct contrasts of pitch – (tempo + stress + articulation) for the S, E, and L groups clearly located significant activations in the bilateral STG/HG (stereotactic coordinates [x, y, z] of the local maxima: [$-48, -19, 5$] and [$51, -13, 2$]) alone (Fig. 4E). Moreover, another direct contrasts of (stress + articulation) — (pitch + tempo) for the S and E groups revealed the R. LPMC activations (the local maximum: [$45, 8, 29$]), which extended anteriorly to the R. IFG (F3op/F3t; the local maximum: [$46, 32, 17$]), demonstrating the distinct functional role of these regions (Fig. 4F).

Figure 6 shows the percent signal changes across groups and main conditions, for each local maximum of the L. STG/HG, R. STG/HG, R. PrCG/PoCG, L. IFG, and R. LPMC (see Fig. 4). In the L. STG/HG (Fig. 6A), activations under the pitch condition were significantly higher than those under the other conditions ($t[390] = 9.9$, $P < 0.01$), which was replicated by the R. STG/HG (Fig. 6B; $t[390] = 7.8$, $P < 0.01$). In the R. PrCG/PoCG (Fig. 6C), activations for the S and E groups were higher than those for the L group under the pitch condition ($t[96] = 2.9$, $P < 0.01$), and activations for the S group were also higher

than those for other groups under the tempo condition ($t[96] = 2.5$, $P < 0.05$). In contrast, the L. IFG (Fig. 6D) was activated in the E group alone under the pitch condition ($t[67] = 2.2$, $P < 0.05$, a comparison with the S group), which was consistent with the activations under the stress and articulation conditions ($t[321] = 2.9$, $P < 0.01$, a comparison indicated in the figure), both of which were observed in the E and L groups. In contrast, the R. LPMC (Fig. 6E) was activated for the S and E groups under the articulation condition ($t[96] = 2.1$, $P < 0.05$, a comparison with the L group), which was consistent with activations under the stress condition ($t[361] = 4.8$, $P < 0.01$, a comparison indicated in the figure). These results indicate that activations in the R. PrCG/PoCG, L. IFG, and R. LPMC (Fig. 6B–E) were actually shared by the multiple main conditions, whereas those in the bilateral STG/HG (Fig. 6A,B) were specific to the pitch condition.

Finally, we examined whether the activations observed in the S group, who showed significantly higher accuracy under all conditions (Fig. 3A,B), indeed reflected musically valid judgment. For this purpose, we additionally analyzed the runs with higher accuracy ($\geq 75\%$) for the S group, after omitting several trials with music stimuli in which the S group scored less than 60%. As a result, we had to drop 5 out of 33 participants who either had less than 75% accuracy for both runs of a single condition or who exhibited excessive head movements. The results were basically the same as those shown in Figs 4 and 5 (see Supplementary Fig. 2). The bilateral STG/HG, L. SMG, R. PrCG/PoCG, and SMA were activated under the pitch condition, together with an additional region of the left and the medial superior frontal gyrus. Under the tempo condition, the L. STG/HG, R. PrCG/PoCG, R. Insula, R. Putamen, and R. Amyg./HC were significantly activated (Supplementary Table 4). Note that activation in the visual areas disappeared under the pitch or tempo condition. Under the stress and articulation conditions, the activation patterns replicated those shown above, confirming the reliability of our observation of these activations. Moreover, under the articulation condition, the L. F3op/F3t activation extended ventrally to the orbital part of the IFG (F3O, Brodmann's area 47).

Discussion

By examining three groups with different levels of music experience, we obtained several striking results. First, we observed distinct activation patterns under the main conditions, such that specific activations under the pitch condition were consistently replicated for all tested groups in the bilateral STG/HG, that is, the auditory areas (Fig. 4E), as well as in the L. IFG and L. LPMC under the articulation condition (Fig. 4A,D). Secondly, music-experience-related activations for the S and E groups were observed in multiple regions, including the R. PrCG/PoCG, under the pitch condition (Fig. 4A), as well as in the R. LPMC under the articulation condition (Fig. 4D). Indeed, the R. LPMC/IFG, that is, the right homologs of

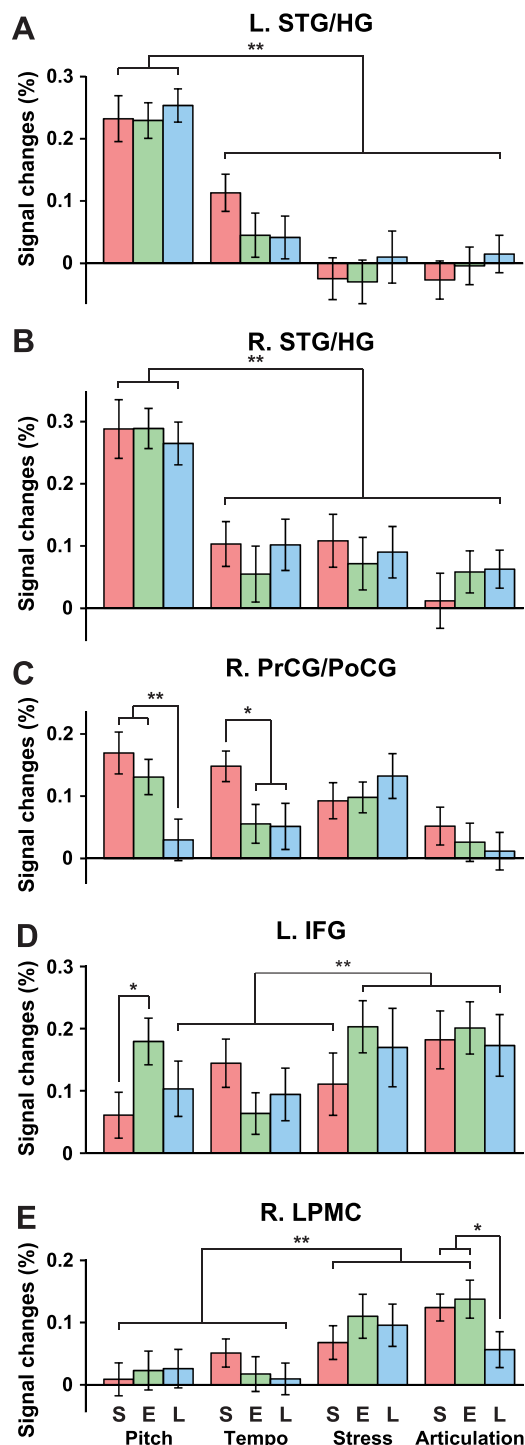


Fig. 6. Consistency and differences in cortical activations across groups and conditions. Histograms for the percent signal changes are shown for each of the representative local maxima (yellow dots) in Figure 4: the R. STG/HG (A), L. STG/HG (B), R. PrCG/PoCG (C), L. IFG (D), and R. LPMC (E). The signal changes are shown for each of the four main conditions with reference to the connection condition. Error bars indicate the SEM for the participants. * $P < 0.05$; ** $P < 0.01$.

the language areas, was specifically activated under the stress and articulation conditions (Fig. 4F). Thirdly, activations specific to the S group were observed under the tempo condition mostly in the right regions (Figs 4B and 5B). This group-specific pattern may be related with

the processes of monitoring the right tempo and/or special training effects for Suzuki students. These results demonstrate the existence of music-related signatures in the brain activations, including both universal and experience-related mechanisms. Because the nuisance factors for analyzing activations included age, IQ, and gender, our results were free from these factors. Using this focused and careful setup of experiments, the localization of music-related systems in the brain, including the bilateral STG/HG and R. LPMC/IFG, was clearly elucidated, as discussed in detail below.

The Suzuki group detected errors at a higher accuracy (see Fig. 3), because they had better understanding about the musical pieces and/or they could memorize them better. Under the tempo condition, even for the runs with higher accuracy (Supplementary Fig. 2B), we observed activations specific to the S group in the left auditory cortex (L. STG/HG) and right sensorimotor cortex (R. PrCG/PoCG), together with the R. Putamen, and R. Amyg./HC, that is, the regions related with learning and memory (Fig. 5B). On the other hand, the R. PrCG/PoCG, which may represent “hand areas” for playing instruments, was consistently activated for the S group under the pitch, tempo, and stress conditions (see Fig. 6C). The unique activation patterns under the tempo condition may be associated with their musical training and/or certain properties of the Suzuki Method, which emphasizes tempo controls in lessons, thereby imitating the playing styles of great performers. Shin’ichi Suzuki stated that “For learning musical beat and correct musical tempo I made tapes with piano accompaniment for all the pieces of Suzuki School Vol. 1 . . . This method has had a marvelous effect on the students’ sensitivity for musical tempo and beat” (Suzuki 1998, p. 10). In other words, the ability to keep time in a musically meaningful manner, that is with irregular but consistent variations that match the expressive content of a piece, is considered one of the most important skills in playing instruments. For example, “melody leads” have been shown to reveal the melodic intentions of performers, with the performance of events interpreted as “melody” preceding other events involving chords by 20–50 ms (Palmer 1996). Moreover, Palmer (1996) reported that melody leads were larger in the performances of experts than those of students, and the melody leads of students increased with practice. Further detailed studies are required to determine which of these factors were responsible for the activations we observed. Nevertheless, these findings suggest that the understanding of intricate and subtle cues in an expressive music performance, one of the important factors that highlight musical experience, could account for the marked activations in the identified regions of the S group.

Language and music share some critical properties, that is, both consist of temporal sequences of sounds that produce phrase or syntactic structures as well. There are an increasing number of studies on the common neural foundation for language and music (Maess et al. 2001;

Besson and Schön 2003; Levitin and Menon 2003; Patel 2003; Fadiga et al. 2009; Dehaene-Lambertz et al. 2010; Koelsch et al. 2013; Tervaniemi et al. 2021). We have previously shown that the most critical regions in syntactic processes are the left opercular/triangular parts of the IFG (L. F3op/F3t) and the L. LPMC (Hashimoto and Sakai 2002; Kinno et al. 2008). A possible “universal” function of the L. F3op/F3t could be a syntactic process called Merge, which is a simple and primitive combinatory operation to create a phrase or clause (Chomsky 1995, 2000). Activations in this region actually predicted “the Degree of Merger,” that is, the maximum depth of merged subtrees (called Mergers), in a sentence; the more binary branches (i.e., syntactic nodes) the sentence had, the more active the region became (Ohta et al. 2013a, 2013b). If musical pieces are constructed with hierarchical structures, these regions, that is, grammar centers, would play a crucial role as well, when structured phrases, rather than individual notes, are evaluated as the tonal structure (Jackendoff and Lerdahl 2006). On the other hand, these regions may not be recruited when other aspects of music are examined, because these regions were not significantly activated during semantic or phonological error detection (Suzuki and Sakai 2003). In the present study, activations in the L. IFG were significantly greater under the articulation and stress conditions than those under the pitch and tempo conditions (see Fig. 6D, $P < 0.01$ with some group mixing), indicating that the articulatory processes were more effective for the L. IFG activation than simpler pitch or tempo processing. While another previous fMRI study of music with lyrics reported an interaction between syntactic processes of sentences and harmony in the L. IFG (Kunert et al. 2015), the present study extends the role of the L. IFG to music without lyrics. We recently showed that the grammar centers play a pivotal role among the syntax-related networks (Kinno et al. 2014; Tanaka et al. 2019), further proposing that the R. LPMC/IFG, that is, the right homologs of the grammar centers, have supportive role for the function of the L. IFG and L. LPMC. Indeed, the R. LPMC/IFG was involved under the articulation and stress conditions consistently for the S and E groups (Fig. 4F). In contrast, the R. LPMC activations were significantly lower under the articulation condition for the L group (see Fig. 6E, $P < 0.05$), whose members had the least music experience. Monotonous speech production, that is, speech lacking proper speech articulation or lacking the structural combination of individual syllables, which is similar to the example of our stimuli used under the articulation condition (see Fig. 1E), is actually caused by a lesion in the R. LPMC/IFG (Ross and Mesulam 1979).

For the runs with higher accuracy for the S group, we observed significant activation in the L. F3O, in addition to the L. F3op/F3t and L. LPMC, under the articulation condition (Supplementary Fig. 2D), which is consistent with the previous proposal that a functional

role of this language-related region is to process musical structures (Levitin and Menon 2003). Because the L. F3O subserves sentence comprehension (Sakai 2005), this intriguing result fits nicely with the idea that articulation is deeply involved in interpreting musical phrases. When experienced performers in music were asked to play a piece “unmusically” (i.e., without any musical style), their approaches to articulation changed significantly from those in their musical performances, suggesting that the rules of performance timing (e.g., the use of rubato patterns) may reflect the mechanisms for parsing acoustic material into phrases and articulated melodic lines (Palmer 1989). On real performances and creative processes, Gerald Klickstein stated that “Like a storyteller, you lead an audience through the narrative of a piece. Without appreciation for compositional design, a performer might overlay grandiose inflections that have little to do with a piece’s syntax” (Klickstein 2009, p. 34). As Louis Moyse (1912–2007), a modern composer and flutist, put it, “Articulation is the word of the composer” (personal communication).

Supplementary Material

Supplementary material can be found at *Cerebral Cortex* online.

Funding

This study received partial funding from Suzuki School of Music, the Talent Education Research Institute. The funder was not involved in the collection, analysis, interpretation of data, or the decision to submit it for publication.

Notes

We would like to thank the teachers and students who have supported our research, especially Kayono Nagata, Etsuko Suehiro, Seizo Azuma, Wakana Miyachi, Mio Noguchi, and Sakiko Ishikawa for coordinating the Suzuki-Brain project, Hayate Tada, Keita Umejima, Naoko Komoro for technical assistance, and Hiromi Matsuda for administrative assistance. *Conflict of Interest:* We declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

K.L.S., Y.O., T.M., and Ryu.H. planned the research; Y.O. and Rei.H. performed the experiments; Y.O., Rei.H., and K.L.S. analyzed the data; K.L.S., Y.O., T.M., and Ryu.H. wrote the paper.

Data Availability

The authors may make the stimuli and/or experimental paradigm available upon request. The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

References

- Albouy P, Benjamin L, Morillon B, Zatorre RJ. 2020. Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science*. 367:1043–1047. <https://doi.org/10.1126/science.aaz3468>.
- Ashburner J, Friston KJ. 2005. Unified segmentation. *NeuroImage*. 26: 839–851. <https://doi.org/10.1016/j.neuroimage.2005.02.018>.
- Bangert M, Peschel T, Schlaug G, Rotte M, Drescher D, Hinrichs H, Heinze H-J, Altenmüller E. 2006. Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *NeuroImage*. 30:917–926. <https://doi.org/10.1016/j.neuroimage.2005.10.044>.
- Berry W. 1985. Metric and rhythmic articulation in music. *Music Theory Spectr*. 7:7–33. <https://doi.org/10.2307/745879>.
- Besson M, Schön D. 2003. Comparison between language and music. In: Peretz I, Zatorre RJ, editors. *The cognitive neuroscience of music*. Oxford: Oxford University Press, pp. 269–293.
- Chomsky N. 1995. *The minimalist program*. Cambridge (MA): The MIT Press.
- Chomsky N. 2000. Minimalist inquiries: The framework. In: Martin R, Michaels D, Uriagereka J, editors. *Step by step: essays on minimalist syntax in honor of Howard Lasnik*. Cambridge (MA): The MIT Press, pp. 89–155.
- Dehaene-Lambertz G, Montavont A, Jobert A, Alliol L, Dubois J, Hertz-Pannier L, Dehaene S. 2010. Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain Lang*. 114:53–65. <https://doi.org/10.1016/j.bandl.2009.09.003>.
- Durnez J, Degryse J, Moerkerke B, Seurinck R, Sochat V, Poldrack RA, Nichols TE. 2016. Power and sample size calculations for fMRI studies based on the prevalence of active peaks. *Preprint* 1–23. <https://doi.org/10.1101/049429>.
- Ericsson KA, Krampe RT, Tesch-Römer C. 1993. The role of deliberate practice in the acquisition of expert performance. *Psychol Rev*. 100: 363–406. <https://doi.org/10.1037/0033-295x.100.3.363>.
- Fadiga L, Craighero L, D'Ausilio A. 2009. Broca's area in language, action, and music. *Ann NY Acad Sci*. 1169:448–458. <https://doi.org/10.1111/j.1749-6632.2009.04582.x>.
- Flinker A, Doyle WK, Mehta AD, Devinsky O, Poeppel D. 2019. Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. *Nat Hum Behav*. 3:393–405. <https://doi.org/10.1038/s41562-019-0548-z>.
- Forgeard M, Winner E, Norton A, Schlaug G. 2008. Practicing a musical instrument in childhood is associated with enhanced verbal ability and nonverbal reasoning. *PLoS One*. 3(e3566):1–8. <https://doi.org/10.1371/journal.pone.0003566>.
- Friston KJ, Holmes AP, Worsley KJ, Poline J-P, Frith CD, Frackowiak RSJ. 1995. Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapp*. 2:189–210. <https://doi.org/10.1002/hbm.460020402>.
- Fujioka T, Ross B, Kakigi R, Pantev C, Trainor LJ. 2006. One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain*. 129:2593–2608. <https://doi.org/10.1093/brain/awl247>.
- Hall DA, Plack CJ. 2009. Pitch processing sites in the human auditory brain. *Cereb Cortex*. 19:576–585. <https://doi.org/10.1093/cercor/bhn108>.
- Hashimoto R, Sakai KL. 2002. Specialization in the left prefrontal cortex for sentence comprehension. *Neuron*. 35:589–597. [https://doi.org/10.1016/s0896-6273\(02\)00788-2](https://doi.org/10.1016/s0896-6273(02)00788-2).
- Hendricks KS. 2011. The philosophy of Shinichi Suzuki: "Music education as love education". *Phil Music Educ Rev*. 19:136–154. <https://doi.org/10.2979/philmusieducrevi.19.2.136>.
- Herholz SC, Zatorre RJ. 2012. Musical training as a framework for brain plasticity: behavior, function, and structure. *Neuron*. 76: 486–502. <https://doi.org/10.1016/j.neuron.2012.10.011>.
- Jackendoff R, Lerdahl F. 2006. The capacity for music: what is it, and what's special about it? *Cognition*. 100:33–72. <https://doi.org/10.1016/j.cognition.2005.11.005>.
- Kinno R, Kawamura M, Shioda S, Sakai KL. 2008. Neural correlates of noncanonical syntactic processing revealed by a picture-sentence matching task. *Hum Brain Mapp*. 29:1015–1027. <https://doi.org/10.1002/hbm.20441>.
- Kinno R, Ohta S, Muragaki Y, Maruyama T, Sakai KL. 2014. Differential reorganization of three syntax-related networks induced by a left frontal glioma. *Brain*. 137:1193–1212. <https://doi.org/10.1093/brain/bwt013>.
- Klickstein G. 2009. *The musician's way: a guide to practice, performance, and wellness*. Oxford: Oxford University Press.
- Koelsch S, Rohrmeier M, Torrecuso R, Jentschke S. 2013. Processing of hierarchical syntactic structure in music. *Proc Natl Acad Sci U S A*. 110:15443–15448. <https://doi.org/10.1073/pnas.1300272110>.
- Koelsch S, Vuust P, Friston K. 2019. Predictive processes and the peculiar case of music. *Trends Cogn Sci*. 23:63–77. <https://doi.org/10.1016/j.tics.2018.10.006>.
- Kunert R, Willems RM, Casasanto D, Patel AD, Hagoort P. 2015. Music and language syntax interact in Broca's area: an fMRI study. *PLoS One*. 10(e0141069):1–16. <https://doi.org/10.1371/journal.pone.0141069>.
- Lehr R. 1992. Sixteen S-squared over D-squared: A relation for crude sample size estimates. *Stat Med*. 11:1099–1102. <https://doi.org/10.1002/sim.4780110811>.
- Levitin DJ, Menon V. 2003. Musical structure is processed in "language" areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *NeuroImage*. 20:2142–2152. <https://doi.org/10.1016/j.neuroimage.2003.08.016>.
- Maess B, Koelsch S, Gunter TC, Friederici AD. 2001. Musical syntax is processed in Broca's area: an MEG study. *Nat Neurosci*. 4:540–545. <https://doi.org/10.1038/87502>.
- Miyamae T. 2018. Musical experience and brain plasticity: Does musical training change brain structure and function? (Japanese). *Brain Nerve*. 70:617–631. <https://doi.org/10.11477/mf.1416201054>.
- Mumford JA. 2012. A power calculation guide for fMRI studies. *Soc Cogn Affect Neurosci*. 7:738–742. <https://doi.org/10.1093/scan/nss059>.
- Ohta S, Fukui N, Sakai KL. 2013a. Computational principles of syntax in the regions specialized for language: integrating theoretical linguistics and functional neuroimaging. *Front Behav Neurosci*. 7(204):1–13. <https://doi.org/10.3389/fnbeh.2013.00204>.
- Ohta S, Fukui N, Sakai KL. 2013b. Syntactic computation in the human brain: the degree of merger as a key factor. *PLoS One*. 8(e56230):1–16. <https://doi.org/10.1371/journal.pone.0056230>.
- Ohta S, Koizumi M, Sakai KL. 2017. Dissociating effects of scrambling and topicalization within the left frontal and temporal language areas: An fMRI study in Kaqchikel Maya. *Front Psychol*. 8(748):1–14. <https://doi.org/10.3389/fpsyg.2017.00748>.

- Oldfield RC. 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*. 9:97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4).
- Palmer C. 1989. Mapping musical thought to musical performance. *J Exp Psychol Hum Percept Perform*. 15:331–346. <https://doi.org/10.1037/0096-1523.15.2.331>.
- Palmer C. 1996. On the assignment of structure in music performance. *Music Percept*. 14:23–56. <https://doi.org/10.2307/40285708>.
- Patel AD. 2003. Language, music, syntax and the brain. *Nat Neurosci*. 6:674–681. <https://doi.org/10.1038/nn1082>.
- Plante E, Creusere M, Sabin C. 2002. Dissociating sentential prosody from sentence processing: activation interacts with task demands. *NeuroImage*. 17:401–410. <https://doi.org/10.1006/nimg.2002.1182>.
- Ross ED, Mesulam M-M. 1979. Dominant language functions of the right hemisphere? Prosody and emotional gesturing. *Arch Neurol*. 36:144–148. <https://doi.org/10.1001/archneur.1979.00500390062006>.
- Sakai KL. 2005. Language acquisition and brain development. *Science*. 310:815–819. <https://doi.org/10.1126/science.1113530>.
- Schellenberg EG. 2004. Music lessons enhance IQ. *Psychol Sci*. 15:511–514. <https://doi.org/10.1111/j.0956-7976.2004.00711.x>.
- Schellenberg EG. 2006. Long-term positive associations between music lessons and IQ. *J Educ Psychol*. 98:457–468. <https://doi.org/10.1037/0022-0663.98.2.457>.
- Seither-Preisler A, Parncutt R, Schneider P. 2014. Size and synchronization of auditory cortex promotes musical, literacy, and attentional skills in children. *J Neurosci*. 34:10937–10949. <https://doi.org/10.1523/JNEUROSCI.5315-13.2014>.
- Steinschaden B, Zehetmair H. 1985. *Ear training and violin playing*. Van Nuys (CA): Summy-Birchard Inc., Alfred Publishing Co. Inc.
- Strelnikov KN, Vorobyev VA, Chernigovskaya TV, Medvedev SV. 2006. Prosodic clues to syntactic processing—A PET and ERP study. *NeuroImage*. 29:1127–1134. <https://doi.org/10.1016/j.neuroimage.2005.08.021>.
- Suzuki S. 1998. *Shinichi Suzuki: his speeches and essays*. Van Nuys (CA): Summy-Birchard Inc., Alfred Publishing Co. Inc.
- Suzuki S. 2013. *Nurtured by love*. Revised ed. Van Nuys (CA): Alfred Publishing Co. Inc.
- Suzuki K, Sakai KL. 2003. An event-related fMRI study of explicit syntactic processing of normal/anomalous sentences in contrast to implicit syntactic processing. *Cereb Cortex*. 13:517–526. <https://doi.org/10.1093/cercor/13.5.517>.
- Tanaka K, Ohta S, Kinno R, Sakai KL. 2017. Activation changes of the left inferior frontal gyrus for the factors of construction and scrambling in a sentence. *Proc Jpn Acad Ser B Phys Biol Sci*. 93:511–522. <https://doi.org/10.2183/pjab.93.031>.
- Tanaka K, Nakamura I, Ohta S, Fukui N, Zushi M, Narita H, Sakai KL. 2019. Merge-generability as the key concept of human language: evidence from neuroscience. *Front Psychol*. 10(2673):1–16. <https://doi.org/10.3389/fpsyg.2019.02673>.
- Tervaniemi M, Putkinen V, Nie P, Wang C, Du B, Lu J, Li S, Cowley BU, Tammi T, Tao S. 2021. Improved auditory function caused by music versus foreign language training at school age: is there a difference? *Cereb Cortex*. 32:63–75. <https://doi.org/10.1093/cercor/bhab194>.
- Thompson WF, Schellenberg EG, Husain G. 2004. Decoding speech prosody: do music lessons help? *Emotion*. 4:46–64. <https://doi.org/10.1037/1528-3542.4.1.46>.
- Timmers R, Ashley R, Desain P, Heijink H. 2000. The influence of musical context on tempo rubato. *J New Music Res*. 29:131–158. <https://doi.org/10.1076/jnmr.29.2.131.3095>.
- Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*. 15:273–289. <https://doi.org/10.1006/nimg.2001.0978>.
- Umejima K, Flynn S, Sakai KL. 2021. Enhanced activations in syntax-related regions for multilinguals while acquiring a new language. *Sci Rep*. 11:7296. <https://doi.org/10.1038/s41598-021-86710-4>.
- Yasui T, Kaga K, Sakai KL. 2009. Language and music: differential hemispheric dominance in detecting unexpected errors in the lyrics and melody of memorized songs. *Hum Brain Mapp*. 30:588–601. <https://doi.org/10.1002/hbm.20529>.