Behavioral/Cognitive

Neural Correlates of Music Listening and Recall in the Human Brain

Yue Ding,[1] Yang Zhang,[1] Wenjing Zhou,[2] Zhipei Ling,[3] Juan Huang,[4] ◉Bo Hong,[1] and ◉Xiaoqin Wang[1,4]

[1]Tsinghua Laboratory of Brain and Intelligence (THBI) and Department of Biomedical Engineering, Tsinghua University, Beijing 100084, P.R. China, [2]Department of Epilepsy Center, Tsinghua University Yuquan Hospital, Beijing 100040, P.R. China, [3]Department of Neurosurgery, General Hospital of People's Liberty Army, Beijing 100853, P.R. China, and [4]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland 21205

Previous neuroimaging studies have identified various brain regions that are activated by music listening or recall. However, little is known about how these brain regions represent the time course and temporal features of music during listening and recall. Here we analyzed neural activity in different brain regions associated with music listening and recall using electrocorticography recordings obtained from 10 epilepsy patients of both genders implanted with subdural electrodes. Electrocorticography signals were recorded while subjects were listening to familiar instrumental music or recalling the same music pieces by imagery. During the onset phase (0–500 ms), music listening initiated cortical activity in high-gamma band in the temporal lobe and supramarginal gyrus, followed by the precentral gyrus and the inferior frontal gyrus. In contrast, during music recall, the high-gamma band activity first appeared in the inferior frontal gyrus and precentral gyrus, and then spread to the temporal lobe, showing a reversed temporal sequential order. During the sustained phase (after 500 ms), delta band and high-gamma band responses in the supramarginal gyrus, temporal and frontal lobes dynamically tracked the intensity envelope of the music during listening or recall with distinct temporal delays. During music listening, the neural tracking by the frontal lobe lagged behind that of the temporal lobe; whereas during music recall, the neural tracking by the frontal lobe preceded that of the temporal lobe. These findings demonstrate bottom-up and top-down processes in the cerebral cortex during music listening and recall and provide important insights into music processing by the human brain.

Key words: auditory; ECoG; hearing; memory; music; recall

Significance Statement

Understanding how the brain analyzes, stores, and retrieves music remains one of the most challenging problems in neuroscience. By analyzing direct neural recordings obtained from the human brain, we observed dispersed and overlapping brain regions associated with music listening and recall. Music listening initiated cortical activity in high-gamma band starting from the temporal lobe and ending at the inferior frontal gyrus. A reversed temporal flow was observed in high-gamma response during music recall. Neural responses of frontal and temporal lobes dynamically tracked the intensity envelope of music that was presented or imagined during listening or recall. These findings demonstrate bottom-up and top-down processes in the cerebral cortex during music listening and recall.

Introduction

Understanding how the brain analyzes, stores, and retrieves auditory information in speech and music remains one of the most

challenging problems in neuroscience. Compared with brain mechanisms for speech processing (Hickok and Poeppel, 2007; Leonard and Chang, 2014), we know much less on how the human brain processes music. Electrocorticography (ECoG) recordings from the human brain provide an opportunity to directly explore temporal dynamics of cortical activation between brain regions in sound perception and imagery, which refers to the process when sound perception is internally generated in the absence of external stimuli. Previous ECoG studies on speech processing revealed serial progression of the activation between

auditory cortex on superior temporal gyrus (STG) and regions on the frontal lobe (Edwards et al., 2010). There is accumulating evidence that imagery and perception share similar neural representations in overlapping cortical regions (Rosen et al., 2000; Palmer et al., 2001; Aziz-Zadeh et al., 2005). When studying covert speech, which involves only speech imagery, and overt speech, which contains both speech perception and imagery, researchers found that the models built from an overt speech dataset using high-gamma band activity could be used to reconstruct a covert speech, suggesting a partially share neural substrate (Pasley et al., 2012; Martin et al., 2014). However, imagery-related brain activation is thought to result from top-down induction mechanisms (Tian and Poeppel, 2012).

Previous neuroimaging studies on perception of tones and tone patterns have revealed the recruitment of brain regions, including the secondary auditory cortex (Griffiths et al., 1999; Zatorre and Belin, 2001; Hall et al., 2002; Hart et al., 2003), the bilateral inferior frontal gyrus (IFG) (Koelsch et al., 2002; Tillmann et al., 2003), as well as the cerebellum, basal ganglia, supplementary motor area, premotor cortex, and parietal cortex (Janata and Grafton, 2003; Doelling and Poeppel, 2015; Fujioka et al., 2015; for review, see Zatorre et al., 2002). It has been observed that the high-gamma band ECoG activities recorded from posterior STG and the precentral gyrus are significantly correlated with the intensity of the music heard by subjects (Potes et al., 2012; Sturm et al., 2014). Furthermore, neuroimaging studies suggest a considerable overlap in the brain activation between music perception and imagination, including the STG, middle temporal gyrus, IFG, middle frontal gyrus, parietal lobe, supramarginal gyrus, supplementary motor area, and premotor cortex (Zatorre et al., 1996, 2007; Halpern and Zatorre, 1999; Janata, 2001; Schürmann et al., 2002; Halpern et al., 2004; Kaiser and Lutzenberger, 2005; Herholz et al., 2008, 2012; Leaver et al., 2009; Hubbard, 2010). Martin et al. (2018) recorded ECoG responses while a musician played music pieces with or without auditory feedback, and found that auditory cortex areas showed similar spectral and temporal tuning properties between perception and imagery conditions. Encoding models could be built to predict high gamma neural activity (70–150 Hz) from the spectrogram representation of the recorded music sounds in both perception and imagery conditions.

However, how the cortical regions involved in music perception and imagination are temporally engaged during music listening and recall has remained largely unknown. Identifying temporal sequences of neural activity across different cortical regions during music recall is essential to elucidate mechanisms underlying the formation and retrieval of music memory. In the present study, we aimed to identify the cortical regions that are involved during music listening and recall, and characterize the temporal dynamics of neural activity during these two processes. We hypothesize that music listening and recall produce similar neural activity within overlapping cortical regions that reflect particular features of music (heard or imagined) but with different temporal orders of neural activity among these regions. To address these questions, we took the advantage of high temporal and spatial resolutions of ECoG signals to investigate the neural correlates of music listening and recall in the same subjects.

Materials and Methods

The present study was approved by the Institutional Review Boards of Tsinghua University, Yuquan Hospital, and Chinese PLA General Hospital. Informed consent was obtained from each subject before the experiment.

Table 1. Subject demographics[a]

Subject ID	Age (yr)	Gender	Seizure focus location	Electrode coverage (contact number)
S1	26	Female	Right inferior precentral gyrus	Right frontal lobe (12)
				Right parietal lobe (21)
				Right temporal lobe (28)
S2	28	Female	Left frontal lobe	Left frontal lobe (2)
				Left parietal lobe (23)
				Left temporal lobe (16)
S3	24	Male	Right inferior precentral gyrus and hippocampus	Right frontal lobe (41)
				Right parietal lobe (9)
				Right temporal lobe (2)
S4	15	Male	Left frontal lobe and middle temporal gyrus	Left frontal lobe (31)
				Left parietal lobe (9)
				Left temporal lobe (13)
S5	21	Female	Right inferior frontal lobe	Right frontal lobe (12)
				Right parietal lobe (8)
				Right temporal lobe (8)
S6	33	Male	Left occipital lobe and hippocampus	Left parietal lobe (11)
				Left temporal lobe (26)
				Left occipital lobe (6)
S7	19	Female	Right inferior precentral gyrus	Right frontal lobe (37)
				Right parietal lobe (4)
				Right temporal lobe (14)
S8	13	Male	Left middle temporal gyrus, temporal-occipital junction, hippocampus	Left parietal lobe (31)
				Left temporal lobe (12)
				Left occipital lobe (16)
S9	17	Male	Right frontal pole	Right frontal lobe (34)
				Right parietal lobe (5)
				Right temporal lobe (11)
S10	45	Male	Right anterior temporal lobe and medial temporal lobe	Right frontal lobe (34)
				Right parietal lobe (10)
				Right temporal lobe (18)

[a]All the subjects were right handed.

Experimental subjects
A total of 10 subjects were tested in this study: 9 from Tsinghua University affiliated Yuquan Hospital and 1 from Chinese PLA General Hospital, including 6 males and 4 females (Table 1). Their ages ranged between 13 and 45 years (mean 24.1 years, SD 9.6 years, median 22.5 years). All subjects were patients diagnosed as having medically intractable epilepsy and underwent ECoG recording to identify the seizure foci. The subjects' ECoG signals were chronically monitored with subdural electrodes for a period of 1–4 weeks, during which time they were available to participate in the experiments reported here. The ECoG electrode coverage of every subject was purely determined by clinical needs. All 10 subjects were right-handed. No detailed information on musical functioning level was available other than there was no formal music training for any of the subjects.

Data acquisition
Subjects were implanted with clinical subdural electrodes (platinum grids, 4 mm diameter of each electrode and 1 cm interelectrode center-to-center distance). Four subjects received left hemisphere implants, and the other 6 received right hemisphere implants (Table 1). In all subjects (except for Subject 10), ECoG signals were recorded from implanted electrodes using a 96-channel amplifier/digitizer system (G.tec). The amplifier sampled signals at 1200 Hz, using a high-pass filter with a 0.1 Hz cutoff frequency and a notch filter at 50 Hz to remove power line noise. Subject 10's data were recorded by a clinical amplifier (Nicolet) with the sampling rate of 1024 Hz. Four electrodes placed on the external surface of the skull with the contacts of the electrodes facing away from the skull were used as ground and reference (two as ground and two as reference, for redundancy).

Sound stimuli
Sound stimuli consisted of 16 pieces of well-known instrumental music without lyrics (7- to 14.3-s-long), including 6 Chinese and 10 Western

Table 2. Information of the database of music pieces

Music name	Stimulus duration (s)	Chosen times
Military Anthem of the People's Liberation Army	7.0	0
Butterfly lovers	7.5	2
Opening theme of TV series "Journal to the west"	7.0	1
Alla Turca Piano Sonata No. 11	7.5	0
Für Elise	7.6	1
Wedding March (Richard Wagner)	14.3	2
Méditation	10.0	0
Opening theme of weather broadcast	12.5	1
Opening theme of daily CCTV news hour	14.3	4
Summer (Joe Hisaishi)	11.5	0
The Blue Danube	8.5	0
The Magnificent Seven	9.7	0
Carmen (Bizet)	8.0	0
Doraemon	8.0	1
Galloping Horses	10.5	0
Welcome march	10.5	5

music pieces (Table 2). The music pieces were edited using MATLAB (The MathWorks) to equal their mean intensity. The stimuli were presented to the subjects through the free-field loudspeakers (AX510 Multimedia Speaker, Zylux Acoustic) placed beneath a computer screen in front of a subject, ~50 cm away. If the condition of subjects allowed, a pair of insert earphones were used instead (ER2, Etyotic Research). Three subjects received sound stimuli via inserted earphones. The volume was adjusted to a comfortable level for both free-field loudspeakers and inserted earphones, at ~65 dB SPL. At least 1 h before an ECoG recording session, each subject was presented with all musical pieces in the stimuli and asked to rate the familiarity of each piece with a score between 1 and 5 (1: least familiar; 5: most familiar). The music pieces with a rating ≥4 by each subject were chosen as the stimuli for that subject. Six of 10 subjects were tested using one music piece, and the other 4 subjects were tested by 2–4 music pieces. Stimulus presentations were controlled by MATLAB using Psychophysics Toolbox 3.0 extension (Brainard, 1997).

Electrode localization
The location of the electrodes relative to the cortical surface was determined using Freesurfer software (Dale et al., 1999; Fischl et al., 1999) and custom MATLAB programs. MRI data were acquired from each subject on an Achieva 3.0T TX scanner (Philips) before surgery. Following the placement of the subdural grids, 3D head CT images were obtained. CT images were registered to the presurgical structural MRI images with Statistical Parametric Mapping implementation of the mutual information-based transform algorithm (Wells et al., 1996) in Freesurfer (Fig. 1A). The registration was then visually inspected and manually adjusted if necessary. The locations and the anatomical labels of electrodes relative to the individual brains were obtained using the Talairach coordinates (Reuter et al., 2012). For the data analysis in this study, the individual brains were coregistered to the Fsaverage Standard Brain in Freesurfer, and all electrodes were displayed on an inflated standard brain for visualization (Fig. 1B). The brain region labels were determined according to Desikan-Killiany Atlas (Desikan et al., 2006).

Experimental design and statistical analysis
Experiment procedure. We conducted two experiments in each subject: an imagery experiment and a control experiment as explained below. The imagery experiment consisted of a short cue and a long cue condition, the comparison between which would be used as a manipulation check of music recalling. In both conditions, subjects were presented with the initial portion of a music piece and instructed to recall the rest of the piece by imagery. Subjects pressed a button (space key on a computer keyboard) as soon as they mentally completed the recall with the hand ipsilateral to their electrode coverage. This allowed us to estimate the recall duration for further analyses. In the short cue condition, a subject listened to the initial ~5 s of a music piece and recalled the rest portion of the music piece (~5 s). In the long cue condition, a subject listened to the

initial ~8 s of a music piece and recalled the rest portion of the music piece (~2 s) (Fig. 1C). The onset of the recall was always at the same time point within one stimulus across subjects for a given experimental condition. The imagery experiment therefore provided ECoG signals corresponding to the listening of the initial (cue) portion of a music piece and the recall of the unpresented portion of the music piece. The two conditions (long cue or short cue) were tested in separate blocks. In each condition, several stimuli were presented to subjects randomly, with each music piece repeated for 20 times. The duration of each block depended on the number of stimuli but was not longer than 20 min. Subjects were allowed to pause the testing whenever they needed to take a rest.

A control experiment was conducted in each subject before the imagery experiment to obtain ECoG signals to compare with ECoG signals obtained in the imagery experiment. The control experiment was designed to compare the differences in neural responses between recall and nonrecall conditions (Fig. 1C). The interval time between the control and imagery experiments was longer than 1 h. The music pieces of the control experiment were identical to those used in the imagery experiment for each subject. To ensure that subjects attended to the playback of the initial portion of a music piece, but not to recall the rest of the music piece after the playback ended, they were instructed to focus their attention on a 1000 Hz pure tone, which occurred after the end of the music cue, and to press a button with the hand ipsilateral to their electrode coverage as soon as they heard it. As in the imagery experiment, each music piece was repeated for 20 times in the control experiment, in which 11%–20% were catch trials. In catch trials, the pure tone was played at a random time after the music cue ended. In noncatch trials, the pure tone was played at the time the music piece would have ended (10 s after music cue onset). Only data from noncatch trials in the control experiment were used to compare with the data from the imagery experiment. Each subject's reaction time in the control experiment was recorded (0.47 ± 0.24 s, N = 10 subjects). The onset phase activity in the imagery and control experiments was taken from the time windows after the offset of the music cue and compared.

Behavioral data analysis. Data analyses were performed using custom MATLAB programs. For behavioral data, each subject's reaction time was estimated by the average response time of the control experiment. The reaction time was subtracted from the raw response time in the imagery experiment to obtain the recall duration that presumably reflected the duration of the mental imagery. In the imagery experiment, the trials with recall duration longer than 3 SDs from the mean recall duration of all trials in short or long cue condition were excluded from further analyses, which resulted in 14.1 ± 4.2 trials for each condition. Because recalling an unpresented portion of a music piece is a variable process, we only analyzed the ECoG data during the recall period for the 8 subjects, except for Subjects S3 and S4 who showed significantly different recall durations between short and long cue conditions (two-tailed two-sample t test, $p < 0.05$, Fig. 1D; Table 3). The ECoG data during the listening period were analyzed for all 10 subjects. In the control experiment, the catch trials were used to control the attention of subjects and were excluded from the further data analysis.

ECoG data analysis. Data from a total of 505 electrodes were included in the analysis, excluding the electrodes clinically identified within the ictogenic zone or considered as corrupted (e.g., showing frequent epileptiform activities) during recording (Fig. 1A). The raw ECoG data were filtered by a notch filter to remove 50 Hz noise and its second and third harmonics, using Fieldtrip toolbox in MATLAB (Oostenveld et al., 2011). The filtered data were then segmented into trials starting from 2 s before stimulus onset to 2 s after the button press. The 2 s prestimulus and poststimulus windows were included to avoid edge effects in subsequent analyses. The segmented data were manually examined in Fieldtrip software to identify and remove noisy trials to generate the preprocessed data for the further analyses. There were at least 15 trials analyzed in each stimulus condition in every subject (ranging from 15 to 40 trials in each subject).

High-gamma response and latency of the onset phase. The onset phase was defined as the first 500 ms after the start of a music piece for the listening condition or after the end of the initial portion of a music piece for the music recall condition. The onset phase of the long cue and short
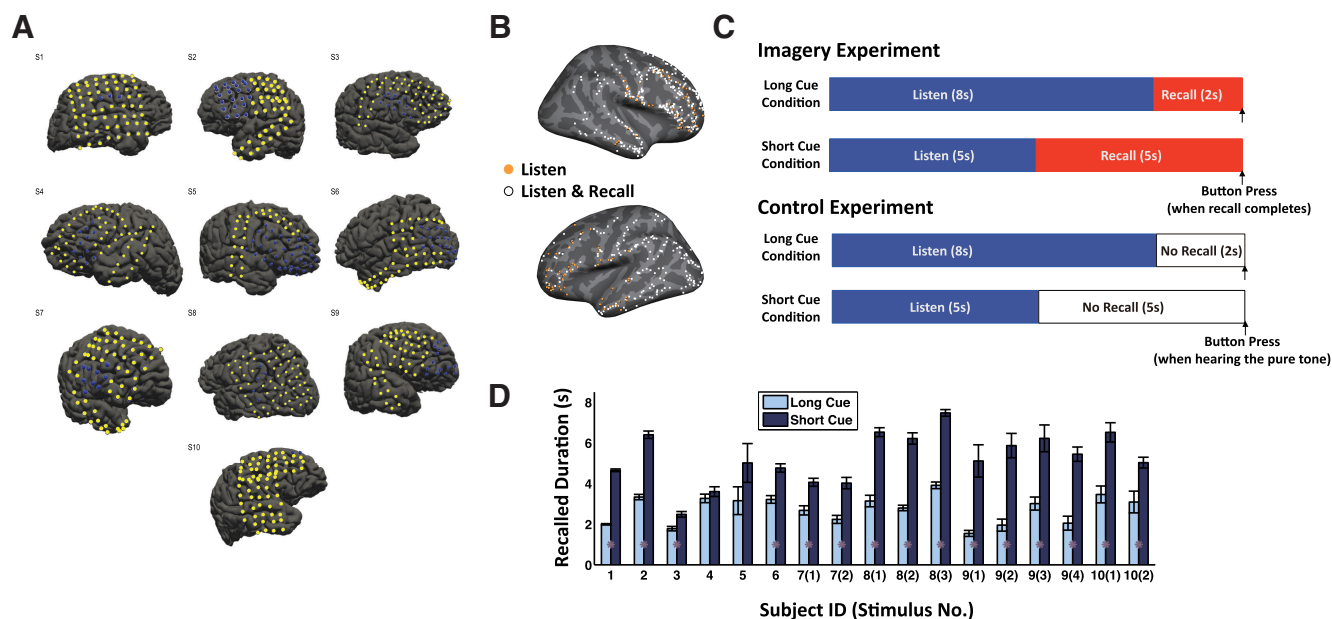
Figure 1. Experiment paradigm and behavioral result. **A**, Individual electrode coverage. Yellow dots represent electrodes included in the analysis. Dark blue dots represent electrodes clinically identified within the ictogenic zone or considered as corrupted during recording. **B**, Location of all the electrodes. Orange dots represent electrodes analyzed only in music listening. White dots represent electrodes analyzed in both music listening and music recall. **C**, Experiment paradigms of the imagery experiment (top two bars) and the control experiment (bottom two bars). Blue bars represent music listening. Red bars represent music recall. White bars represent no music playback and no recall. Arrows indicate the timing of button press in the imagery experiment (when a subject finished music recall) and the timing of pure tone and button press in the control experiment (when a subject heard the pure tone). **D**, Behavioral results of all the subjects. Dark blue bars represent recall duration of short cue condition. Light blue bars represent recall duration of long cue condition. Purple stars represent significant difference between long cue and short cue conditions (unpaired Student's t test, $p < 0.05$). Error bars indicate SEM across trials. Subjects 1–6 were tested by one stimulus, and Subjects 7–10 were tested by more than one stimulus. Data from each stimulus are plotted separately.

Table 3. Statistical result of behavioral data[a]

	S1(1)	S2(1)	S3(1)	S4(1)	S5(1)	S6(1)	S7(1)	S7(2)
t	36.93	13.68	3.93	1.05	1.57	5.55	4.61	5.20
p	<0.001	<0.001	<0.001	0.30	0.13	<0.001	<0.001	<0.001

	S8(1)	S8(2)	S8(3)	S9(1)	S9(2)	S9(3)	S9(4)	S10(1)	S10(2)
t	9.34	11.04	15.45	4.18	6.03	4.34	6.86	4.85	3.17
p	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.004

[a]p values were calculated by two-tailed two-sample t test between long cue and short cue condition for each subject.

cue conditions were collapsed in the analyses. High-gamma response (60–140 Hz) was used to quantify the onset phase of ECoG signals, which was extracted by the following steps. To correct the 1/frequency decay of power in spectra of ECoG signals to enhance lower-frequency components: (1) the ECoG signal was first filtered by a series of narrow-band zero-phase filters (bandwidth 10 Hz) between 60 and 140 Hz (60–70, 70–80, . . . 130–140 Hz); (2) the output signal of each filter was then Hilbert-transformed to obtain the envelope; (3) for each filtering channel, a baseline ECoG signal was taken from a 500 ms time window before the music onset; (4) the envelope of each filtering channel was then normalized to the mean of the envelope during baseline, expressed as a percentage change of the mean; and (5) the normalized envelopes of all filtering channels were averaged to produce an overall envelope for each trial, defined as the high-gamma response.

To characterize the temporal sequence of the onset high-gamma response across cortical regions, we estimated the latency of high-gamma response during the onset phase in music listening and music recall, respectively. For music listening, the threshold of significant high-gamma response was set as the response exceeding 99.7% CI of the baseline. A high-gamma response was defined as significant only if it exceeded the threshold for at least 100 ms (Nourski et al., 2014). The time point when the high-gamma response first exceeded the threshold was defined as the onset latency of music listening. For music recall, high-gamma response in the imagery experiment had to be significantly larger

than the control experiment (two-sample t test across trials, $p < 0.05$) and lasted for at least 100 ms to be taken as a significant recall response. The time point when the high-gamma response first showed a significant difference between the imagery experiment that the control experiment (two-sample t test across trials, $p < 0.05$) was defined as the onset latency of music recall.

Cortical response analysis of the sustained phase. The sustained phase of the ECoG signal during music listening was defined as the time window from 500 ms after the start of a music piece to the end of the music piece. For music recall, the sustained phase was defined as the time window from 500 ms after the end of the initial portion a music piece (cue) to the time when subjects completed the imagery, which was supposed to be near the end of the whole music piece. The cortical regions responded during the sustained phase were analyzed by the intertrial coherence (ITC) measure (Golumbic et al., 2013), using two types of ITC measures (phase-ITC and power-ITC).

For phase-ITC, the ECoG signal of a trial was filtered into six frequency bands (delta 1–3 Hz, theta 4–7 Hz, alpha 8–12 Hz, beta 12–30 Hz, low gamma 30–40 Hz, high-gamma 60–140 Hz) to generate the phase feature.

For power-ITC, the ECoG signal of a trial was analyzed by a multitaper-based spectrum method on Slepian sequences (Slepian, 1978) (window: 400 ms; step: 10 ms). Center frequencies ranged from 1 to 140 Hz in 1 Hz increments. For each center frequency, the power of the

ECoG signal was log-transformed and normalized to the mean power of the baseline (a 500 ms time window before the music onset) on a trial-by-trial basis, expressed as a percentage change of the mean, and then averaged within each frequency band. The frequency bands we chose were identical as phase-ITC.

For both phase-ITC and power-ITC, correlation coefficients were then calculated between ITC measures of pairs of trials obtained from each electrode. Due to the different lengths of recall, the long cue and short cue conditions were analyzed separately for the sustained phase. We calculated the correlation coefficients for the trials of the same stimulus (within-stimulus) as well as the trials of different stimuli (cross-stimulus). For example, if Stimulus A had M trials and Stimulus B had N trials, there were $(C_M^2 + C_N^2)$ correlation coefficients of within-stimulus trials ($R|within$), and M × N correlation coefficients of cross-stimulus trials ($R|across$), calculated as follows:

$$R|within = \{corr(Sig_A)(i_A), Sig_A((j_A)), corr(Sig_B(i_B), Sig_B(j_B))\} \quad (1)$$

$$where\ i_A = 1, 2, \ldots, M;\ j_A = 1, 2, \ldots, M;\ i_A \neq j_A;$$

$$i_B = 1, 2, \ldots, N;\ j_B = 1, 2, \ldots, N;\ i_B \neq j_B$$

$$R|across = \{corr(Sig_A(m), Sig_B(n))\},\ where\ m = 1, 2, \ldots, M,$$

$$n = 1, 2, \ldots, N \quad (2)$$

The correlation coefficients were Fisher z-transformed (Meng et al., 1992) to meet the normal distribution for the statistical analysis. The significance of ITC measures was determined by comparing the within-stimulus correlation coefficients and the cross-stimulus correlation coefficients with a two-sample t test. Significant larger within-stimulus correlation coefficients ($p < 0.05$, two-sample t test) indicated consistent responses across the trials of the same stimulus.

ITC analysis was conducted for music listening and music recall, respectively. For music recall, the recall duration may vary. We hypothesized that subjects recalled the entire unpresented portion of the music pieces but at a slightly different pace for each trial. To adjust the recall duration to the same length for each music piece, the recall response was linearly compressed or expanded to the actual unpresented duration on a trial-by-trial basis after calculating ITC measures with the temporal structure preserved.

Correlation analysis. For the electrodes revealed by ITC showing a consistent response across the trials of the same stimulus, a correlation analysis was further applied to assess the degree to which neural responses were correlated with the intensity profile of music (heard or recalled). For the electrodes with significant phase-ITC, the neural response was defined as the averaged bandpass filtered (delta band) signal across trials; whereas for the electrodes with significant power-ITC, the neural response was defined as the averaged high-gamma responses across trials. An intensity profile was obtained for each music piece by calculating the running average of the squared amplitude of short segments of the piece (window width: 0.01 s; overlap: 50%), and down-sampled to the same sampling rate as the neural response (100 Hz). Pearson correlation was calculated between the neural response and the intensity profile of the music piece. To quantify the accuracy and time lag of this "tracking" property, we computed the Pearson correlation coefficients between the music intensity profile and the corresponding neural response at different lag times, ranging from 200 ms preceding to 300 ms lagging a stimulus. The statistical significance of the Pearson correlation was evaluated by a permutation test, where the null distribution consisted of the correlation coefficients between the music piece in shuffled temporal orders and the neural response in the original temporal order. A significantly larger correlation coefficient than the null distribution (corrected for multiple comparisons; false discovery rate) (Benjamini and Yekutieli, 2001) indicated that the neural response significantly correlated with the intensity profile of the music piece at a particular lag time.

Results

To identify the involvement of specific cortical regions during music listening and recall and characterize temporal dynamics of neural activity during these two processes, we analyzed ECoG data obtained while subjects listened to and recalled highly familiar music pieces.

Cortical regions activated during music listening and recall

We first analyzed cortical regions that were activated during music listening or recall. ECoG signals were analyzed in two time windows, referred to as onset (0–500 ms) and sustained (after 500 ms) phases of the responses.

Onset phase

Figure 2A shows cortical regions with significant high-gamma responses during the onset phase of music listening. The color indicates the high-gamma response strength (measured as the percentage of change in high-gamma power over the prestimulus baseline within the period of significant high-gamma response). During music listening, posterior STG (pSTG) showed the strongest high-gamma response in the right hemisphere (Fig. 2A). In the left hemisphere, pSTG also showed stronger high-gamma response than the other cortical regions. In addition, bilateral supramarginal gyrus, temporal lobe, superior frontal gyrus, and precentral gyrus, as well as the IFG of the right hemisphere showed high-gamma response during the onset phase of music listening (Fig. 2A). Figure 2B shows cortical regions with significant high-gamma responses during the onset phase of music recall. Significant responses were only found in the right hemisphere, including STG, precentral area, IFG, and superior frontal gyrus (Fig. 2B).

In Figure 2C, we plot high-gamma responses to music listening from an example electrode located in the right pSTG region as shown in Figure 2A (right). This electrode showed similarly strong responses to music stimuli during listening in both imagery and control experiments. Figure 2D shows high-gamma response during music recall from an example electrode located in the right precentral gyrus. On the same figure, we also plot the high-gamma response of the same electrode from the control experiment while the subject listened to the same portion of the music piece. This electrode showed a stronger response in the imagery experiment (recall) than in the control experiment (no recall).

Sustained phase

The ECoG signals during music listening or recall lasted for several seconds after the initial onset phase (0–500 ms). We refer to this portion of neural responses as the sustained phase. Figure 3A plots multiple trials of ECoG signals recorded from one electrode while a subject listened to the same piece of music. To determine whether the ECoG signals were consistent across multiple trials in a particular experimental condition (listening or recall), ITC was calculated using either phase (phase-ITC) or power (power-ITC) signals in one of the six frequency bands (see Materials and Methods).

Figure 3B plots the number of electrodes with significant within-stimulus ITC at each frequency band for both phase-ITC and power-ITC in two experimental conditions (listening and recall). In the delta band, 6 electrodes showed significant within-stimulus phase-ITC during listening and 4 electrodes during recall (Fig. 3B) ($p < 0.05$, unpaired t test). The delta-band phase-ITC values of all electrodes are displayed on the brain surface in
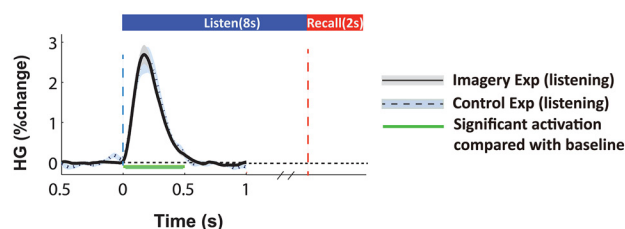
A Listening (onset phase)

Mean HG (% change)

C
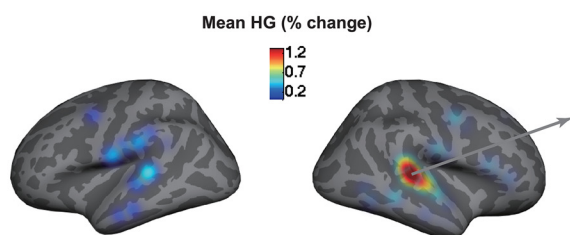
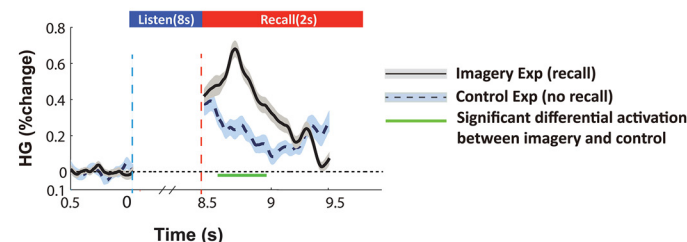## B Recall (onset phase)

Mean HG (% change)
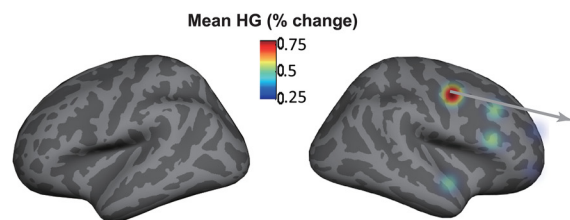
D

Figure 2. Brain regions with neural responses during the onset phase of music listening and recall. ***A***, ***B***, Group-averaged brain surface maps showing regions with significant high-gamma responses in music listening (***A***) or recall (***B***) condition. Color represents the high-gamma response strength expressed as the percentage of high-gamma power change relative to the baseline. There was no significant high-gamma response on the left hemisphere during music recall. ***C***, An example of high-gamma responses, expressed as the percentage of high-gamma power change relative to the baseline, recorded from an electrode on the right pSTG during music listening in the imagery experiment (solid line with gray shadow) and the control experiment (dashed line with blue shadow). Time 0 indicates the onset of music playback. Horizontal green bar represents the period of significant high-gamma response compared with the baseline high-gamma signal. Shadows are SEM across trials. ***D***, High-gamma responses of an example electrode on the right precentral gyrus during music recall in the imagery experiment (solid line with gray shadow) and no-recall condition in the control experiment (dashed line with blue shadow). Two vertical dashed lines indicate the beginning and end, respectively, of the music cue. Horizontal green bar represents the period of significant differential high-gamma response between the imagery experiment and the control experiment. Shadows are SEM across trials.

Figure 3*C*. Electrodes with the highest delta-band phase-ITC values during music listening were located mainly in the right pSTG. In contrast, during music recall, electrodes with the highest delta-band phase-ITC values were located on the right supramarginal gyrus. Significant power-ITC was found mainly in the high-gamma band, with 7 and 36 electrodes showing significant within-stimulus power-ITC during listening and recall, respectively (Fig. 3*B*). Electrodes with the highest high-gamma band power-ITC values during music listening were located mainly in the bilateral pSTG with the right pSTG having the highest ITC values (Fig. 3*D*, top row). The right precentral gyrus and IFG also had significant high-gamma band power-ITC. As for music recall, significant high-gamma band power-ITC was found in bilateral frontal areas, with higher values on the right prefrontal areas (Fig. 3*D*, bottom row).

Sequential high-gamma responses cross cortical regions during the onset phase of music listening and recall
As shown in Figure 2, several cortical regions were activated during the onset phase of music listening and recall with different high-gamma response strengths. To study the temporal dynamics of high-gamma responses during these two processes, we calculated the latencies of high-gamma responses within these cortical regions. To increase the statistical power of the latency analysis, we grouped the activated cortical regions into five general regions according to the region labels in the Desikan-Killiany Atlas (Desikan et al., 2006), which were as follows: (1) TG, temporal gyri; (2) SG, supramarginal gyrus; (3) PG, precentral gyrus; (4) IFG; and (5) SFG, superior frontal gyrus (Fig. 4*A*, left). The mean latency of the neural activity during music listening was calculated for each of these regions in both hemispheres (Fig. 4*A*, right). On the left hemisphere, the PG region had the shortest latency, followed by SG, TG, and SFG regions. However, the la-

tency differences among these five regions were not significant (Wilcoxon rank sum test, $p > 0.05$). On the right hemisphere, TG and SG regions had similar latencies (Wilcoxon rank sum test, $p = 0.3766$). These two regions were activated significantly earlier than the PG region (Wilcoxon rank sum test, $p < 0.01$). The IFG region had significantly longer latency than the PG region (Wilcoxon rank sum test, $p < 0.05$). The SFG region has a similar latency as the IFG region. The latency differences between right and left hemispheres were not significant (Wilcoxon rank sum test, $p > 0.05$). There was also no significant difference between the latencies of listening condition in imagery experiment and control experiment (paired t test, $p = 0.8426$) for the cortical regions that were activated in both experiments.

The latency data in Figure 4*A* show that there was a sequential delay from the posterior to anterior cortical regions (TG/SG to IFG/SFG) on the right hemisphere during the onset phase of music listening. In Figure 4*B* (left), we compare average latencies during the onset phase of music listening and recall in four brain regions on the right hemisphere those were activated by both listening and recall conditions. In contrast to music listening where the latency increased from TG to PG and IFG regions, the latency decreased from TG to PG and IFG regions during music recall (Fig. 4*B*, left). The SFG region had the longest latency among all other regions in both listening and recall conditions. The opposite sequential orders of high-gamma response during the onset phase of listening and recall were confirmed by the linear regression of the latencies across these activated regions (Fig. 4*B*, left), which showed that there was a trend of increasing latency from the TG to IFG during music listening (slope > 0, $r = 0.4416$, $p < 0.01$) and decreasing latency during music recall (slope < 0, $r = 0.5865$, $p < 0.05$). Figure 4*B* (right) illustrates the opposite sequential orders of high-gamma response in the right hemisphere.
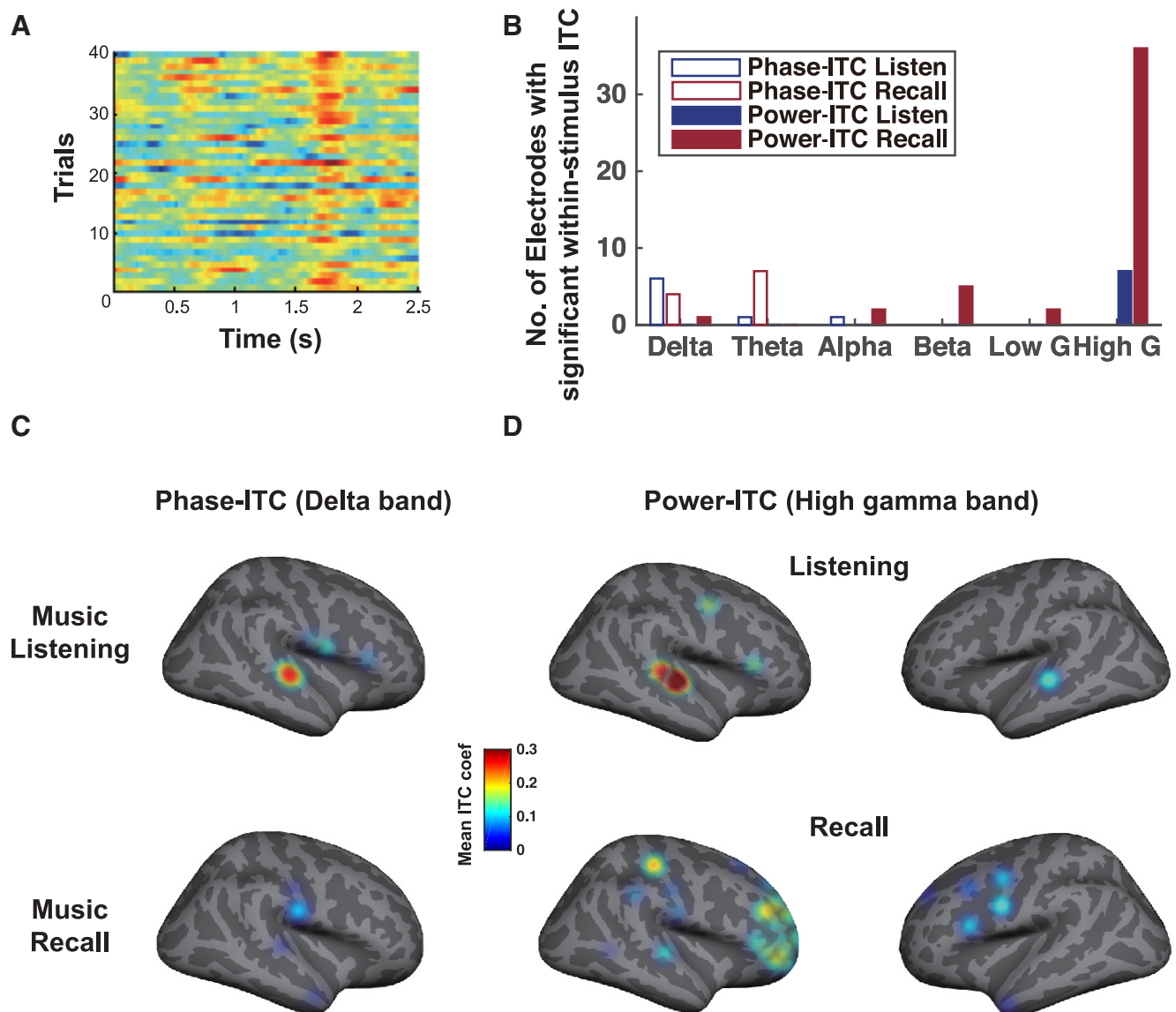
Figure 3. Brain regions with neural responses during the sustained phase of music listening and recall. **A**, Traces of ECoG signals during music listening from an example electrode with high-power-ITC value, filtered between 60 and 140 Hz, and arranged by trial number. **B**, Numbers of electrodes with significant within-stimulus phase-ITC (open bars) or power-ITC (solid bars) in the music listening (blue) and music recall (red) conditions at each frequency band. Significant phase-ITC was found dominantly in the low-frequency range (delta, theta bands), whereas significant power-ITC was found in high-gamma band. **C**, Locations of electrode sites with significant delta phase-ITC in music listening (top) and recall (bottom) conditions on the group-averaged brain. Color represents the ITC value at each site. **D**, Locations of electrode sites with significant high-gamma power-ITC in music listening (upper row) and recall (bottom row) on the group-averaged brain. Color represents the ITC value at each site.

Neural response tracks intensity profile of music during music listening and recall

The consistent neural responses across trials in a particular experimental condition (listening or recall) suggested that ECoG signals may reflect particular features of the music pieces, such as intensity profile. Based on the result in Figure 3B as well as findings of a previous study (Golumbic et al., 2013), we chose delta band phase-ITC and high-gamma band power-ITC to analyze the relationship between neural response and music intensity. For the electrodes with significant phase-ITC (Fig. 3C), the neural response was defined as the averaged bandpass filtered (delta band) signal across trials; whereas for the electrodes with significant power-ITC (Fig. 3D), the neural response was defined as the averaged high-gamma responses across trials.

Figure 5A showed an example of a music intensity profile (black curve) overlaying with its corresponding neural response

(blue curve) during music listening, recorded from one representative electrode. Pearson correlation was calculated between the neural response and the intensity profile of the music piece, with different lag times, ranging from 200 ms preceding 300 ms lagging a stimulus. Figure 5B plots the correlation coefficients at different lag times for the same electrode shown in Figure 5A, where a positive lag time indicates the neural activity lagging the music intensity profile and a negative lag time indicates the neural activity proceeding the music intensity profile. The correlation coefficient reached the peak at the lag time of 90 ms, indicating that the neural activity was lagging behind the music intensity profile. The peak correlation coefficients of the electrodes showing significant correlation during music listening were plotted on the average brain in Figure 5C, D. Figure 5C shows the electrodes from all subjects with significant correlation coefficients using phase-ITC features, located on the right pSTG,
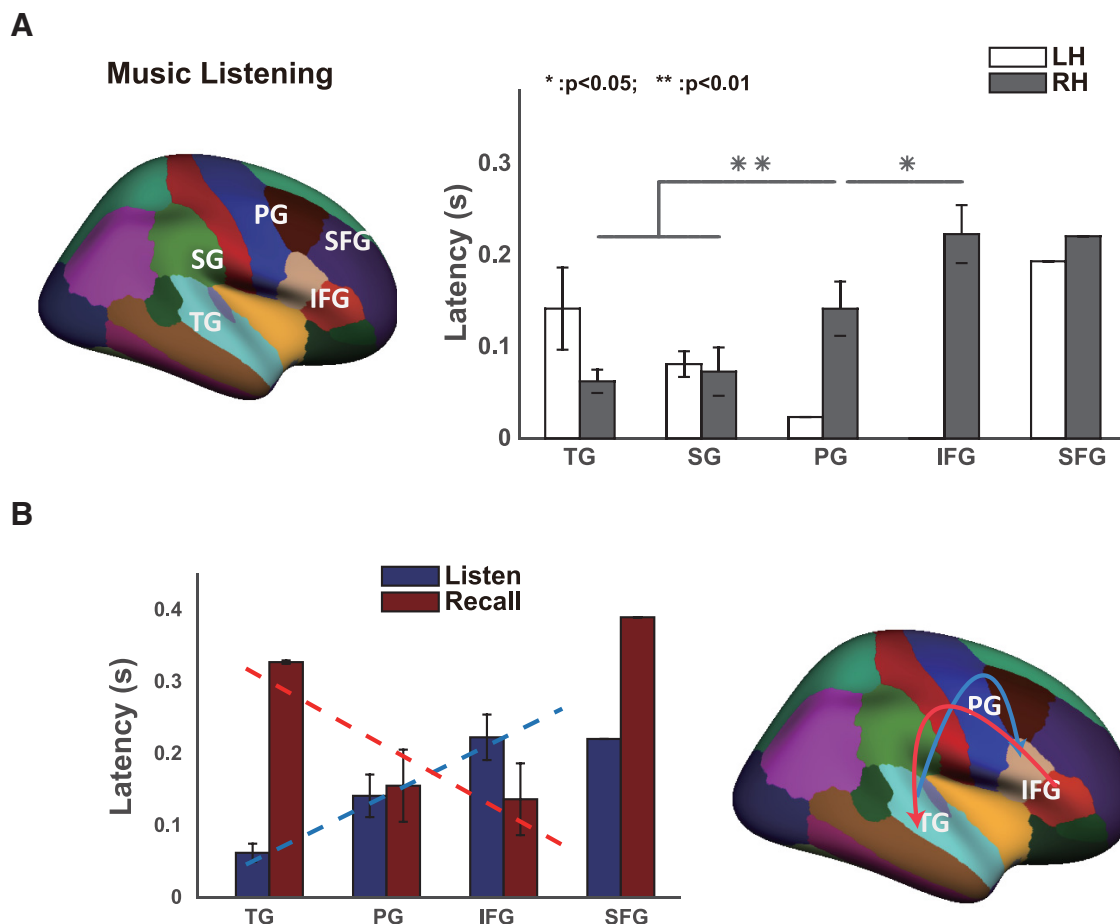
A

B

Figure 4. Dynamics of cortical responses during the onset phase of music listening and recall. **A**, Latency distribution (right) at different brain regions (left) in music listening condition. Open bars represent left hemisphere (LH). Filled bars represent right hemisphere (RH). There was no high-gamma response in the IFG of the left hemisphere. Error bars indicate SEM across electrodes. Asterisks indicate statistical significance of the comparison between different regions. *$p < 0.05$; **$p < 0.01$. **B**, Latency analysis of the onset phases of ECoG signals during music listening and recall. Left, Latency distributions of music listening (blue bars) and recall (red bars) on the right hemisphere. Error bars indicate SEM across electrodes. Dashed lines indicate the linear regression of the latencies across the brain regions for music listening (blue) and recall (red). Right, Schematics represent the flow of high-gamma response across cortical regions in the onset phase of listening (blue arrow) and recall (red arrow), respectively.

SG, and IFG. Figure 5D shows the electrodes from all subjects with significant correlation coefficients using power-ITC features, located on the bilateral pSTG. The electrode from the right pSTG (Fig. 5D, left) showed the highest correlation coefficient ($r = 0.5947$ at -10 ms lag time). The negative time lag indicated a possible involvement of some imagery during perception. The electrode on the left pSTG had a correlation coefficient of 0.4875 at the lag time of 90 ms (Fig. 5D, right). We grouped the electrodes with significant peak correlation coefficients into two groups (TG/SG and FG) according to the anatomical labels. Figure 5E summarizes the lag time of peak correlation coefficients of these electrodes. Among them, the FG electrodes had the longest lag time ($\Delta t = 270$ ms), whereas the TG/SG electrodes had an average lag time of 108 ms. This relationship is consistent with the observations based on the onset phase of music listening (i.e., the sensory areas were activated before the higher cortical areas during music listening).

The correlation analysis was also performed on the sustained phase of the neural responses during music recall. Figure 6A showed an example of the neural response (red curve) recorded from one representative electrode overlaying with the intensity profile of a recalled music segment (black curve). Figure 6B plots the correlation coefficients at different lag times for the same electrode shown in Figure 6A. The correlation coefficient reaches peak at a negative lag time, indicating that the neural

activity was preceding the music intensity profile. The peak correlation coefficients of the electrodes showing significant correlation during music recall were plotted on the average brain in Figure 6C, D. Three phase-ITC electrodes from 2 subjects located on the right SG, anterior TG, and posterior TG showed significant correlation coefficients with the intensity profile of the recalled music (Fig. 6C). Among them, the anterior TG electrode showed the longest negative lag time ($r = 0.1573$ at -190 ms lag time), followed by the posterior TG electrode ($r = 0.2362$ at -120 ms lag time). The SG electrode showed a positive lag time ($r = 0.1794$ at 70 ms lag time). Significant correlations between the neural response and the intensity profile of the recalled music were found in 7 power-ITC electrodes from 4 subjects on the left anterior temporal lobe, right posterior temporal lobe, and supramarginal gyrus and left and right frontal lobe (Fig. 6D). The frontal electrode had the largest negative lag time ($\Delta t = -170$ ms), indicating that the neural activity there was ahead of the recalled music. In contrast, the neural activity on the right SG lagged behind the recalled music ($r = 0.4802$ at 270 ms lag time). Figure 6E summarizes the lag time of peak correlation coefficients of all electrodes, separating into TG/SG and FG groups as in music listening condition. The peak correlation coefficients of FG electrodes occurred preceding the recalled music (average time lag $= -97$ ms), whereas that of TG/SG electrodes oc-
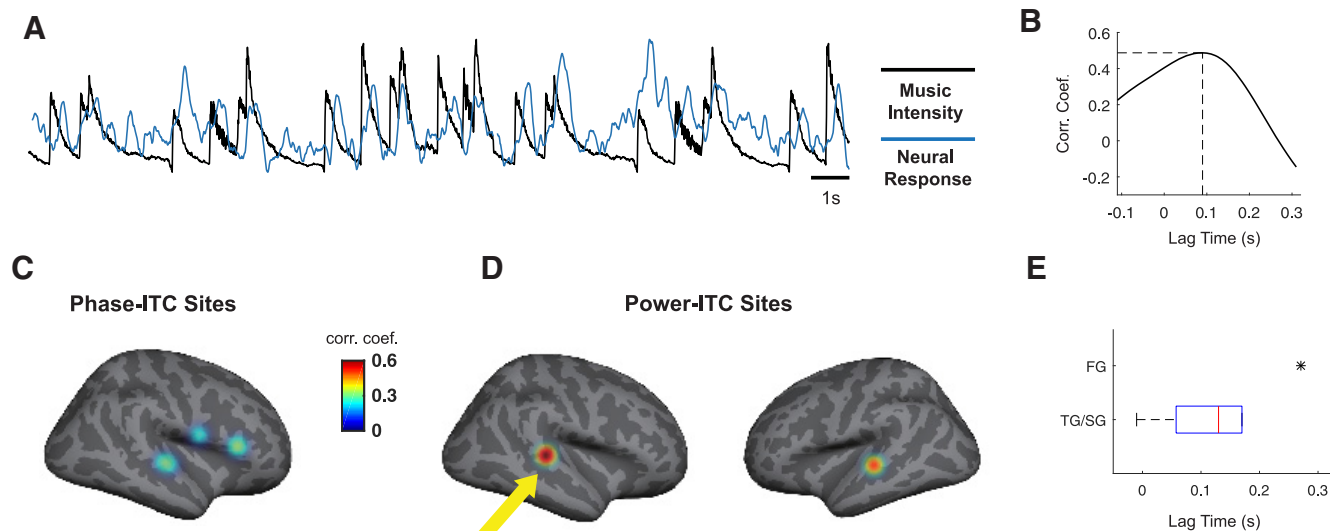
Music Listening

Figure 5. Dynamic tracking of music intensity profile during the sustained phase of music listening. **A**, An example of music intensity profile (black curve) and corresponding neural response (high-gamma response, blue curve) from one electrode on the left pSTG. The neural response is shifted to illustrate the peak correlation at a lag time of 90 ms. **B**, The correlation coefficient of neural response and music intensity is plotted versus the lag time for the example shown in **A**. **C, D**, Locations of phase-ITC (**C**) and power-ITC (**D**) electrode sites with significant correlation with music intensity profile on the group-averaged brain. Color represents the peak correlation coefficient at each site. Yellow arrow indicates the origin of the example waveform in **A**. **E**, Comparison of the lag time at peak correlation coefficients between the FG sites and TG and SG sites. Boxplot represents the median (red line) and range of the lag time, with edges of the blue box indicating 25th and 75th percentiles and the whiskers extending to the most extreme datapoints that the algorithm considered to be not outliers (dashed line). *Lag time of the peak correlation coefficient in the frontal site.
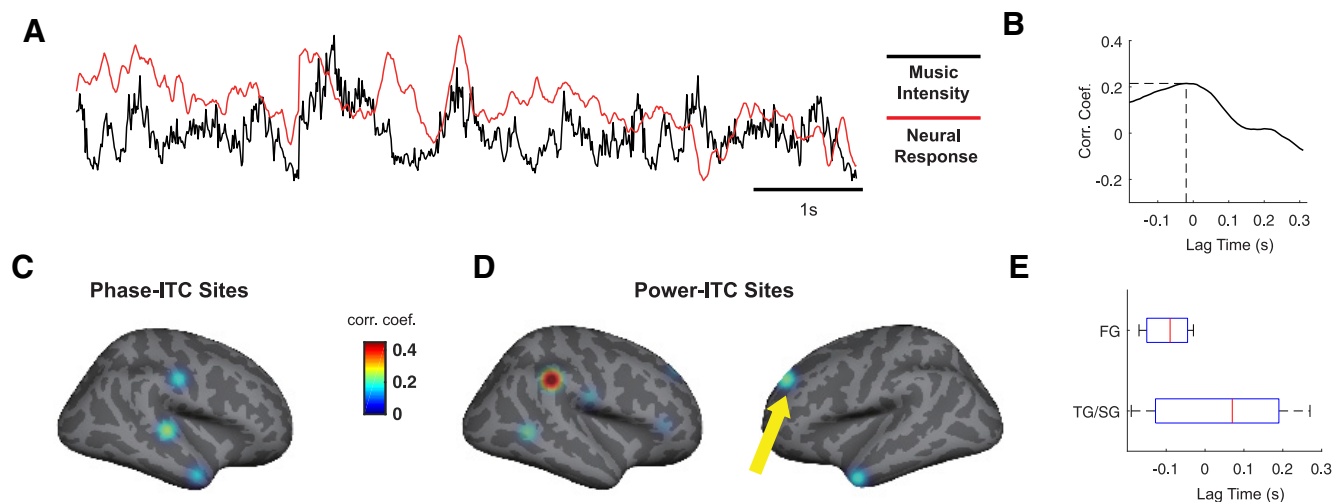
Music Recall

Figure 6. Dynamic tracking of music intensity profile during the sustained phase of music recall. **A**, An example of music intensity profile (black curve) and corresponding neural response (high-gamma response, red curve) from one electrode on the left frontal lobe. The neural response is shifted to illustrate the peak correlation at a lag time of −10 ms. **B**, The correlation coefficient of neural response and music intensity is plotted versus the lag time for the example shown in **A**. **C, D**, Locations of phase-ITC (**C**) and power-ITC (**D**) electrode sites with significant correlation in music recall on the group-averaged brain, with the color representing the peak correlation coefficient at each site. Yellow arrow indicates the origin of the example waveform in **A**. **E**, Comparison of the lag time at peak correlation coefficients between the frontal (FG) sites and temporal (TG) and supramarginal (SG) sites. Boxplot represents the median (red line) and range of the lag time, with edges of the blue box indicating 25th and 75th percentiles and the whiskers extending to the most extreme data points that the algorithm considered to be not outliers (dashed line).

curred behind the recalled music (average time lag = 37 ms). This relationship is consistent with the observations from the onset phase of music recall (i.e., the higher cortical areas were activated before sensory areas during music recall).

Discussion
Music listening and recall: bottom-up and top-down processes
Consistent with previous imaging studies (Zatorre et al., 1996, 2007; Halpern and Zatorre, 1999; Janata, 2001; Schürmann et al.,

2002; Halpern et al., 2004; Kaiser and Lutzenberger, 2005; Herholz et al., 2008, 2012; Leaver et al., 2009; Hubbard, 2010), our results demonstrated that music listening and recall activated overlapping cortical regions, including the temporal lobe, supramarginal gyrus, precentral gyrus, and the frontal lobe, providing evidence of a shared neural substrate for the two processes. EEG studies of music perception and imagery have shown EEG signals containing perception-weighted, imagery-weighted, and shared components (Schaefer et al., 2013, Stober et al., 2015). Schaefer et

al. (2013) showed a temporal component that appeared to be related to the start of a perceived stimulus and parietal and frontocentral components that showed initial activation in imagery. Similarly, our results based on ECoG signals showed different temporal sequential high-gamma responses during music listening and recall. The initialization of the response in music listening is by external auditory stimuli, which reach sensory cortex first and arrive at frontal cortex at last. The information flow from sensory cortex to frontal cortex represents a bottom-up process. In contrast, music recall initializes responses starting from frontal cortex and ending in sensory cortex, reflecting a top-down process. Reversed sequential orders of the activated cortical regions were observed during the onset phase of music listening and recall.

Imagery-related brain activation has been suggested to result from top-down induction mechanisms, including memory retrieval (Kosslyn et al., 2001; Kosslyn, 2005) and motor simulation (Price et al., 2011; Tian and Poeppel, 2012). In memory retrieval, perceptual experience retrieves objects stored in long-term memory to reactivate the sensory cortices. Through top-down reconstruction of the neural representation, the perceptual experience can be reelicited without the presence of any physical stimuli during mental imagery (Kosslyn, 2005). In motor simulation, an efference copy of the motor cortex activity is forwarded to lower sensory cortices, enabling a comparison of actual with desired movement and permitting online behavioral adjustments (Tian and Poeppel, 2012). In the present study, we also observed a top-down process from higher-level associative areas to perceptual-sensory representation areas during music imagery.

Neural correlates of music intensity profile during listening and recall

Previous studies showed that ECoG signals could track musical acoustic feature (music intensity) during music listening (Potes et al., 2012). In the present study, we showed significant correlations between ECoG signals and the music intensity profile in both music listening and recall conditions. During the sustained phase of music listening, the neural response of bilateral pSTG and right IFG tracked the intensity profile of music stimuli (Fig. 5C,D). The time lags corresponding to the peak correlation coefficients of pSTG were shorter than that of IFG, indicating that pSTG followed the music intensity profile more closely than IFG during music listening (Fig. 5E). Previous studies (Potes et al., 2012; Sturm et al., 2014) have shown that the high-gamma power envelope of pSTG tracks displayed music intensity. More importantly, during the sustained phase of music recall, the neural response tracked the intensity profile of recalled music (Fig. 6). In the absence of a physical input, music recall is a self-paced activity driven by an internal generator. During the sustained phase of music recall, the neural response of bilateral frontal lobe preceded music, whereas that of bilateral temporal lobe lagged behind (Fig. 6E). These results indicate that it may be possible to also extract musical features from ECoG signals during recall. In addition, we observed higher correlation coefficients in music listening than in music recall condition, which is in accordance with previous research showing that the predictive power for covert speech was weaker than for overt speech (Martin et al., 2014). This was likely due to the fact that the imagery condition was more variable compared with the listening condition as it was not feasible to measure the exact timing of the subject's mental

activity. Therefore, a weaker correlation in the imagery condition is expected.

Lateralization of music listening and recall

The important role of the right hemisphere in aspects of musical perception, particularly those involving tonal pitch processing, has been reported previously (Zatorre and Samson, 1991; Zatorre et al., 1992, 1994; Penhune et al., 1998; Tervaniemi and Hugdahl, 2003; Trujillo-Pozo et al., 2013). Our results showed slightly right lateralization during music listening, and a more evident right lateralization during music recall. More specifically, during the music listening, both left and right hemispheres showed activation during the onset phase, with the high-gamma band power of the right hemisphere significantly stronger than those of similar regions on the left hemisphere (Fig. 2A). As for the sustained phase of the neural responses, the right hemisphere had more consistent activation in broader regions (Fig. 3C). These findings are largely in line with the evidence from previous studies of the behavioral effects of lesions, which suggests an asymmetry favoring the right temporal cortex in tasks that involve perception of pitch patterns or spectral processing (Zatorre and Samson, 1991; Zatorre et al., 2002). More broadly, relative specialization within right auditory regions for tonal processing is supported by functional imaging data from perception tasks (Zatorre et al., 1992, 1994; Johnsrude et al., 2000; Tervaniemi and Hugdahl, 2003). Our results also showed that the high-gamma envelope of the bilateral pSTG tracked the music intensity, but that of the right hemisphere showed shorter time lag and larger correlation coefficient than the left hemisphere (Fig. 5E) during music listening.

As for music recall, only the right hemisphere was activated during the onset phase (Fig. 2B), and the sustained phase of music recall activated much broader regions on the right hemisphere than the left hemisphere (Fig. 3C). In addition, only the right hemisphere showed the phase-ITC response for either music listening or recall (Fig. 3C). The specialization within right auditory regions has also been observed in memory-related tasks (Penhune et al., 1998; Halpern and Zatorre, 1999; Perry et al., 1999; Trujillo-Pozo et al., 2013). Furthermore, the interactions with frontal-lobe regions are also frequently observed, especially on the right hemisphere (Zatorre and Samson, 1991; Zatorre et al., 1992, 1994). In the present study, we also found that the high-gamma power envelope of the right hemisphere in the frontal lobe preceded the stimuli further than that of the left hemisphere; whereas in the temporal lobe, the high-gamma power envelope of the right hemisphere had larger correlation coefficient with imagined music intensity than the left hemisphere (Fig. 6E).

To conclude, our results further confirm the right hemisphere lateralization of music-related cognitive functions and, in particular, reveal a stronger lateralization during music recall. Moreover, the present study showed a more precise tracking of music intensity profile by the neutral response of the right hemisphere compared with that of the left hemisphere.

Limitations and future research questions

The present work adopted and modified an imagery paradigm to study the neural representations of music listening and recall (Halpern and Zatorre, 1999). While the results provided evidence of neural tracking of recalled music intensity profile, a challenge in analyzing the correlation between music imagery-induced neural activity and the music being recalled is the lack of direct measurement of the imagery process. Music recall is a self-paced behavior, which may last shorter or longer than the music piece being imagined. Some studies controlled the pace

of recall by recording finger tapping (Janata and Grafton, 2003) or the output of musical instrument in case of a participant being a musician (Martin et al., 2018), presenting visual/tactile cues (Zatorre et al., 1996; Yoo et al., 2001; Brodsky et al., 2003) or inducing voluntary imagery (Kraemer et al., 2005). The present study did not use such methods so as not to introduce additional cortical activation. Instead, we hypothesized that subjects can recall the entire unpresented portion of the music pieces but at slightly different paces for each trial. Therefore, the recall ECoG responses were linearly compressed or expanded in the imagery experiments to align them to the duration of music stimuli before the correlation analysis was performed. Our results showed that the neural activity during the music recall appeared to track the intensity profile of the music being imagined (Fig. 6), albeit with a poorer precision than the music listening condition (Fig. 5). There could be other scenarios that were not controlled in the present study, for example, subjects recalling the missing portion at the original pace but lost some details in between (shorter recall duration than actual unpresented duration), or subjects thinking about other things apart from the music per se during the recall (longer recall duration than actual unpresented duration), or subjects recalling the missing portion at a variable pace during the recall. These scenarios were not examined in the present study because of the lack of adequate experimental controls in the paradigm used. Future work is needed to further explore such scenarios by developing better controlled paradigms. However, these limitations, if true, would cause us to underestimate, rather than overestimate, the neural tracking of the intensity profile of the music during recalling in the present study.

Moreover, previous studies have showed evidence for motor processing in the imagination of musical changes in loudness (Bailes et al., 2012), and mental representations of pitch and melody have been shown to involve auditory (Deutsch, 1970; Keller et al., 1995) and motor processing (Mikumo, 1994; Finney and Palmer, 2003). The present study investigated a fundamental feature of music, the intensity profile, which contains the information of rhythm, pitch, and loudness. Further studies on the mental representations of other musical features during music recall are needed. Due to the limitation imposed by patients' physical states and experiment duration, only one or a few music pieces were tested for 1 subject, which prevented us from applying classification analysis to study the predictive power of neural activity to distinguish different music pieces. It also limited the generalization of the time lag we observed. Schaefer et al. (2011) found that the correlation between EEG time course and the intensity of various music pieces during listening is maximal at a latency of 100 ms, which is close to the time lag of sensory areas in the present study, but may not stand for all music pieces. Finally, the results reported here were based on 308 electrodes on the right hemisphere from 6 subjects and 197 electrodes on the left hemisphere from 4 subjects. The unbalanced electrode numbers of two hemispheres as well as the lack of consistent coverage across all subjects could restrict the generalization of some conclusions. Future studies with more diversified stimuli and consistent electrode coverages across hemispheres and subjects can help clarify the above issues.

References

Aziz-Zadeh L, Cattaneo L, Rochat M, Rizzolatti G (2005) Covert speech arrest induced by rTMS over both motor and nonmotor left hemisphere frontal sites. J Cogn Neurosci 17:928–938.

Bailes F, Bishop L, Stevens CJ, Dean RT (2012) Mental imagery for musical changes in loudness. Front Psychol 3:525.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. The annals of statistics 29:1165–1188.

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–436.

Brodsky W, Henik A, Rubinstein BS, Zorman M (2003) Auditory imagery from musical notation in expert musicians. Percept Psychophys 65:602–612.

Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis: I. Segmentation and surface reconstruction. Neuroimage 9:179–194.

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31:968–980.

Deutsch D (1970) Tones and numbers: specificity of interference in immediate memory. Science 168:1604–1605.

Doelling KB, Poeppel D (2015) Cortical entrainment to music and its modulation by expertise. Proc Natl Acad Sci U S A 112:E6233–E6242.

Edwards E, Nagarajan SS, Dalal SS, Canolty RT, Kirsch HE, Barbaro NM, Knight RT (2010) Spatiotemporal imaging of cortical activation during verb generation and picture naming. Neuroimage 50:291–301.

Finney SA, Palmer C (2003) Auditory feedback and memory for music performance: sound evidence for an encoding effect. Mem Cognit 31:51–64.

Fischl B, Sereno MI, Tootell RB, Dale AM (1999) High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum Brain Mapp 8:272–284.

Fujioka T, Ross B, Trainor LJ (2015) Beta-band oscillations represent auditory beat and its metrical hierarchy in perception and imagery. J Neurosci 35:15187–15198.

Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ (2013) Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party." Neuron 77:980–991.

Griffiths TD, Johnsrude I, Dean JL, Green GG (1999) A common neural substrate for the analysis of pitch and duration pattern in segmented sound? Neuroreport 10:3825–3830.

Hall DA, Johnsrude IS, Haggard MP, Palmer AR, Akeroyd MA, Summerfield AQ (2002) Spectral and temporal processing in human auditory cortex. Cereb Cortex 12:140–149.

Halpern AR, Zatorre RJ (1999) When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. Cereb Cortex 9:697–704.

Halpern AR, Zatorre RJ, Bouffard M, Johnson JA (2004) Behavioral and neural correlates of perceived and imagined musical timbre. Neuropsychologia 42:1281–1292.

Hart HC, Palmer AR, Hall DA (2003) Amplitude and frequency-modulated stimuli activate common regions of human auditory cortex. Cereb Cortex 13:773–781.

Herholz SC, Lappe C, Knief A, Pantev C (2008) Neural basis of music imagery and the effect of musical expertise. Eur J Neurosci 28:2352–2360.

Herholz SC, Halpern AR, Zatorre RJ (2012) Neuronal correlates of perception, imagery, and memory for familiar tunes. J Cogn Neurosci 24:1382–1397.

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402.

Hubbard TL (2010) Auditory imagery: empirical findings. Psychol Bull 136:302–309.

Janata P (2001) Brain electrical activity evoked by mental formation of auditory expectations and images. Brain Topogr 13:169–193.

Janata P, Grafton ST (2003) Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. Nat Neurosci 6:682–687.

Johnsrude IS, Penhune VB, Zatorre RJ (2000) Functional specificity in the right human auditory cortex for perceiving pitch direction. Brain 123:155–163.

Kaiser J, Lutzenberger W (2005) Cortical oscillatory activity and the dynamics of auditory memory processing. Rev Neurosci 16:239–254.

Keller TA, Cowan N, Saults JS (1995) Can auditory memory for tone pitch be rehearsed? J Exp Psychol Learn Mem Cogn 21:635–645.

Koelsch S, Gunter TC, v Cramon DY, Zysset S, Lohmann G, Friederici AD (2002) Bach speaks: a cortical "language-network" serves the processing of music. Neuroimage 17:956–966.

Kosslyn SM (2005) Mental images and the brain. Cogn Neuropsychol 22:333–347.

Kosslyn SM, Ganis G, Thompson WL (2001) Neural foundations of imagery. Nat Rev Neurosci 2:635–642.

Kraemer DJ, Macrae CN, Green AE, Kelley WM (2005) Musical imagery: sound of silence activates auditory cortex. Nature 434:158.

Leaver AM, Van Lare J, Zielinski B, Halpern AR, Rauschecker JP (2009) Brain activation during anticipation of sound sequences. J Neurosci 29:2477–2485.

Leonard MK, Chang EF (2014) Dynamic speech representations in the human temporal lobe. Trends Cogn Sci 18:472–479.

Martin S, Brunner P, Holdgraf C, Heinze HJ, Crone NE, Rieger J, Schalk G, Knight RT, Pasley BN (2014) Decoding spectrotemporal features of overt and covert speech from the human cortex. Front Neuroeng 7:14.

Martin S, Mikutta C, Leonard MK, Hungate D, Koelsch S, Shamma S, Chang EF, Millán JD, Knight RT, Pasley BN (2018) Neural encoding of auditory features during music perception and imagery. Cereb Cortex 28:4222–4233.

Meng W, Yu C, Kim W (1992) A theory of translation from relational queries to hierarchical queries. IEEE Trans Knowl Data Eng 7:228–245.

Mikumo M (1994) Motor encoding strategy for pitches of melodies. Music Percept Ann Interdiscip J 12:175–197.

Nourski KV, Steinschneider M, McMurray B, Kovach CK, Oya H, Kawasaki H, Howard MA 3rd (2014) Functional organization of human auditory cortex: investigation of response latencies through direct recordings. Neuroimage 101:598–609.

Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011:156869.

Palmer ED, Rosen HJ, Ojemann JG, Buckner RL, Kelley WM, Petersen SE (2001) An event-related fMRI study of overt and covert word stem completion. Neuroimage 14:182–193.

Pasley BN, David SV, Mesgarani N, Flinker A, Shamma SA, Crone NE, Chang EF (2012) Reconstructing speech from human auditory cortex PLoS Biol 10:e1001251.

Penhune VB, Zattore RJ, Evans AC (1998) Cerebellar contributions to motor timing: a PET study of auditory and visual rhythm reproduction. J Cogn Neurosci 10:752–765.

Perry DW, Zatorre RJ, Petrides M, Alivisatos B, Meyer E, Evans AC (1999) Localization of cerebral activity during simple singing. Neuroreport 10:3453–3458.

Potes C, Gunduz A, Brunner P, Schalk G (2012) Dynamics of electrocorticographic (ECoG) activity in human temporal and frontal cortical areas during music listening. Neuroimage 61:841–848.

Price CJ, Crinion JT, Macsweeney M (2011) A generative model of speech production in Broca's and Wernicke's areas. Front Psychol 2:237.

Reuter M, Schmansky NJ, Rosas HD, Fischl B (2012) Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 61:1402–1418.

Rosen HJ, Ojemann JG, Ollinger JM, Petersen SE (2000) Comparison of brain activation during word retrieval done silently and aloud using fMRI. Brain Cogn 42:201–217.

Schaefer RS, Farquhar J, Blokland Y, Sadakata M, Desain P (2011) Name that tune: decoding music from the listening brain. Neuroimage 56:843–849.

Schaefer RS, Desain P, Farquhar J (2013) Shared processing of perception and imagery of music in decomposed EEG. Neuroimage 70:317–326.

Schürmann M, Raij T, Fujiki N, Hari R (2002) Mind's ear in a musician: where and when in the brain. Neuroimage 16:434–440.

Slepian D (1978) Prolate spheroidal wave functions, Fourier analysis, and uncertainty: V. The discrete case. Bell Labs Tech J 57:43–63.

Stober S, Sternin A, Owen AM, Grahn JA (2015) Towards music imagery information retrieval: introducing the OpenMIIR dataset of EEG recordings from music perception and imagination. Proceedings of the 16th ISMIR Conference, Malaga, Spain, pp 763–769.

Sturm I, Blankertz B, Potes C, Schalk G, Curio G (2014) ECoG high gamma activity reveals distinct cortical representations of lyrics passages, harmonic and timbre-related changes in a rock song. Front Hum Neurosci 8:798.

Tervaniemi M, Hugdahl K (2003) Lateralization of auditory-cortex functions. Brain Res Rev 43:231–246.

Tian X, Poeppel D (2012) Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. Front Hum Neurosci 6:314.

Tillmann B, Janata P, Bharucha JJ (2003) Activation of the inferior frontal cortex in musical priming. Brain Res Cogn Brain Res 16:145–161.

Trujillo-Pozo I, Martín-Monzón I, Rodríguez-Romero R (2013) Brain lateralization and neural plasticity for musical and cognitive abilities in an epileptic musician. Front Hum Neurosci 7:829.

Wells WM, Grimson WL, Kikinis R, Jolesz FA (1996) Adaptive segmentation of MRI data. IEEE Trans Med Imaging 15:429–442.

Yoo SS, Lee CU, Choi BG (2001) Human brain mapping of auditory imagery: event-related functional MRI study. Neuroreport 12:3045–3049.

Zatorre RJ, Belin P (2001) Spectral and temporal processing in human auditory cortex. Cereb Cortex 11:946–953.

Zatorre RJ, Samson S (1991) Role of the right temporal neocortex in retention of pitch in auditory short-term memory. Brain 114:2403–2417.

Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. Science 256:846–849.

Zatorre RJ, Evans AC, Meyer E (1994) Neural mechanisms underlying melodic perception and memory for pitch. J Neurosci 14:1908–1919.

Zatorre RJ, Halpern AR, Perry DW, Meyer E, Evans A (1996) Hearing in the mind's ear: a PET investigation of musical imagery and perception. Cogn Neurosci J 8:29–46.

Zatorre RJ, Belin P, Penhune VB (2002) Structure and function of auditory cortex: music and speech. Trends Cogn Sci 6:37–46.

Zatorre RJ, Chen JL, Penhune VB (2007) When the brain plays music: auditory–motor interactions in music perception and production. Nat Rev Neurosci 8:547–558.