Auditory imagery modulates frequency-specific areas

in the human auditory cortex

Ji Hoon Oh[1†], Jae Hyung Kwon[2†], Po Song Yang[1,3], and Jaeseung Jeong[2,*]

[1] College of Medicine,

The Catholic University of Korea

Seoul, 137-701, Republic of Korea

[2] Department of Bio and Brain Engineering,

Korean Advanced Institute of Science and Technology (KAIST)

Daejeon, 305-701, Republic of Korea

[3] Department of Diagnostic Radiology, Daejeon St. Mary's Hospital

Daejeon, 520-2, Republic of Korea

† These authors contributed equally to this study.

*Address correspondence to

Jaeseung Jeong, Ph.D.

Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), 373-1 Kuseong-dong, Yuseong-gu, Daejeon, South Korea 305-701; +82-42-350-4319 (office); +82-42-864-5318 (fax); jsjeong@kaist.ac.kr (e-mail)

Abstract

Neural responses in early sensory areas influenced by top-down processing. In the visual system, early visual areas have been shown to actively participate in top-down processing based on their topographical properties. Although it has been suggested that the auditory cortex is involved in top-down control, functional evidence of topographic modulation is still lacking. Here, we show that mental auditory imagery for familiar melodies induces significant activation in the frequency-responsive areas of the primary auditory cortex (PAC). This activation is related to the characteristics of the imagery: when subjects were asked to imagine high-frequency melodies, we observed increased activation in the high- versus low-frequency response area; when the subjects were asked to imagine low-frequency melodies, the opposite was observed. Furthermore, we found that A1 is more closely related to the observed frequency-related modulation than R in tonotopic subfields of the PAC. Our findings suggest that top-down processing in the auditory cortex relies on a mechanism similar to that used in the perception of external auditory stimuli, which is comparable to early visual systems.

Keywords: auditory, imagery, primary auditory cortex, top-down, tonotopy

INTRODUCTION

Top-down processing in the brain is characterized by its cognitive control over sensory processing in the framework of a cortical hierarchy, as when one constructs perceptions by drawing on one's experience and expectations. Top-down processing plays a particularly important role in perceptual interpretations because our past experiences and expectations may influence the ways we interpret information. This processing is thought to be mediated by higher cortical areas that ultimately convey neural signals to sensory areas (Fritz, David, Radtke-Schuller, Yin, & Shamma, 2010; Mechelli, 2004; Sack et al., 2008).

Although top-down processing does not necessarily require the presence of external stimuli, several studies have shown that similar mechanisms involved in bottom-up processing are recruited (Chelazzi, Miller, Duncan, & Desimone, 1993; Kosslyn et al., 1993; Kosslyn, Ganis, & Thompson, 2001; Mechelli, 2004; Slotnick, Thompson, & Kosslyn, 2005; Stokes, Thompson, Cusack, & Duncan, 2009; Thirion et al., 2006; Zatorre, Halpern, Perry, Meyer, & Evans, 1996). This observation has predominantly been made in early sensory areas, particularly in the visual system. The primary visual cortex (V1) is known to actively participate in top-down processing, as evidenced by its topographical properties. For example, a subject's attention to one of two overlapping orthogonal orientations strongly bias V1 activity toward the selected orientation (Kamitani & Tong, 2005), and imagining previously presented grid patterns or a rotating wedge activates the relevant retinotopic areas (Slotnick et al., 2005; Thirion et al., 2006). These findings suggest that the bottom-up processing of perceived external stimuli and the top-down processing of mental visual imagery share similar properties in early visual areas. However, in the auditory

system, it is not known how top-down processing extends to early auditory areas and recruits similar physiological mechanisms as those involved in the perception of external stimuli (Kosslyn et al., 2001; Kraemer, Macrae, Green, & Kelley, 2005).

It has been demonstrated that tonotopic representation of the primary auditory cortex was modified by frequency-specific learning in both animals (Bao, Chan, & Merzenich, 2001; Bao, Chan, Zhang, & Merzenich, 2003; Bieszczad & Weinberger, 2010a; 2010b; 2012; Recanzone, Schreiner, & Merzenich, 1993; Rutkowski & Weinberger, 2005) and humans (Menning, Roberts, & Pantev, 2000; Morris, Friston, & Dolan, 1998). For example, during associative learning, frequency tuning shifted to the frequency of a tone that signals reinforcement; neural responses to the frequency of signal tones increased, whereas responses to the frequency of non-signal tones decreased (Bao et al., 2001; Bieszczad & Weinberger, 2010a; 2010b; 2012; Rutkowski & Weinberger, 2005; for reviews, see Weinberger, 2004; 2007). Moreover, this learning-induced representational expansion of the primary auditory cortex reflected the behavioral importance of the sound stimulus. The performance level was positively correlated with the area of representation only at the conditioned frequency (Rutkowski & Weinberger, 2005), and the amount of signal-specific area was correlated with the memory strength for the signal tone (Bieszczad & Weinberger, 2010b). Similar to the above findings, it has been demonstrated that the neural response of human auditory cortex was increased in the loci of frequency representation of conditioning-stimulus (Menning et al., 2000; Morris et al., 1998).

This associative representational plasticity in auditory areas provides an opportunity to examine whether top-down control in the auditory cortex is specific to auditory imagery. Thus, we trained subjects to memorize simple melodies consisted of high- and low-frequency tone blocks and mentally rehearse them. After certain

periods, we examined whether the auditory imagery of learned melodies elicited frequency-specific activation in the auditory cortex of human subjects.

Previous studies of mental auditory imagery have showed the involvement of the auditory cortex in top-down processing (Halpern & Zatorre, 1999; Zatorre & Halpern, 2005). Matching the appropriate sound with a silenced scene (Bunzeck, Wuestenberg, Lutz, Heinze, & Jancke, 2005) and the detection of a sound emerging from silence (Voisin, 2006) can both activate the auditory cortex. Furthermore, a previous study (Kraemer et al., 2005) based on a spontaneous auditory imagery paradigm showed that a subjective, imagined experience (e.g., instrumental tones) induces neural activity in an anatomically defined area of the primary auditory cortex (PAC). However, whether mental auditory imagery activates the PAC in its entirety or only modulates content-relevant areas remains largely unknown.

Similar to other early sensory areas, the PAC exhibits a fine-grained topographical map, which represents mirror-symmetric frequency gradients residing across Heschl's gyrus (Da Costa et al., 2011; Formisano et al., 2003; Humphries, Liebenthal, & Binder, 2010; Woods et al., 2009). Within the PAC, the two largest tonotopic subfields, A1 and R, have been measured along the anterior-posterior axis of the mirror-symmetric frequency progression (Da Costa et al., 2011). This tonotopic organization allows for the investigation of the neural response of A1 based on its frequency characteristics.

As previous studies suggested that the auditory system shares some important cortical properties with the visual system (King & Nelken, 2009; Petkov et al., 2004), we hypothesized that the earliest functional area would be modulated by the frequency content of specific top-down processes. In the present study, we aimed to

examine whether auditory imagery modulates neural activity in the PAC in a frequency-related manner.

We used an auditory imagery task consisting of simple and control melodies in two different frequency blocks: a low-frequency block, which consisted of tones within the range of the D3 (164 Hz) to E4 (294 Hz) scales, and a high-frequency block, which consisted of tones within the range of the C6 (1,047 Hz) to B6 (1,980 Hz) scales (Figure 1). These simple and control melodies had the same number of tones and frequency ranges, but the control melodies consisted of dissonant chords and tempos, and thus they were difficult for the subjects to anticipate upcoming tones or phrases.

Insert Figure 1

To induce auditory imagery, excerpted melodies with silenced gaps were presented to the subjects (n=13) during a functional MRI (fMRI) scan. We extracted several frequency blocks from each melody (6 sec) and replaced the blocks with silenced gaps (Figure 2). The subjects participated in two fMRI scans within an interval of two weeks. In the first scan, the subjects were naïve to both the simple and control melodies. After the first scan, the subjects were presented with simple melodies without gaps and were required to repeatedly listen to the melodies. In the second scan, the subjects were only familiar with the simple melodies. During both scans, the order of the melodies was kept the same, and the subjects were instructed to press buttons on a keypad (Cedrus Corporation, USA, CA) according to the tempo of the melody. After the second experiment, we asked all subjects whether they could

rehearse the melody in their minds during the gaps in the simple melodies. All subjects reported that they had a strong sense of hearing the melody in their minds.

Insert Figure 2

Because the location and structure of the PAC varies widely among individuals (Rademacher, 2001), the tonotopic map of each subject was acquired independently (Figure 3). On the basis of these maps, we localized the high- and low-frequency selective voxels in the PAC for each melody and monitored their neural activities during two different silenced gaps. To prevent interruptions caused by the noise of the scanning machine, a sparse temporal acquisition paradigm (Bunzeck et al., 2005; Formisano et al., 2003; Humphries et al., 2010) was used during the tonotopy mapping and melody presentation.

METHODS

Participants

A total of 13 right-handed subjects (6 males and 7 females; mean age: 22.6 years; age range: 20-28 years) participated in this study. Tone discrimination and Montreal Battery of Evaluation of Amusia (MBEA) tests revealed that all subjects possessed a normal hearing ability and could clearly distinguish between adjacent frequency tones (Peretz, Champod, & Hyde, 2003). All subjects provided written informed consent for the study, which was approved by the Institutional Review Board (IRB) of the Korean Advanced Institute of Science and Technology (KAIST).

Procedure

Sound Stimuli for Tonotopy Mapping

Auditory stimuli were created in MATLAB 7.9 (MathWorks, USA) and were processed by SoundForge 8.0 (Sony, Japan). The stimuli were a series of tone bursts composed of three sine tones (Humphries et al., 2010). The low-frequency tone bursts were centered at 250 Hz and included the following frequencies: 175 Hz (0.7 times the center frequency) and 325 Hz (1.3 times the center frequency). The high-frequency tone bursts were centered at 1,500 Hz and included the following frequencies: 1,050 Hz (0.7 times the center frequency) and 1,950 Hz (1.3 times the center frequency). These frequency ranges were selected to include the frequency ranges of the simple and control melodies.

Each sine wave pulsed with a 10-Hz square wave (5 msec rise and fall time, 40 msec plateau time). Each tone burst was separated by a 50-msec inter-stimulus interval (ISI), and every eighteen tone bursts were followed by a 100-msec ISI (Figure 1). These two different types of ISIs were used in various timescales to avoid a potential adaptation effect in the auditory cortex (Ulanovsky, Las, Farkas, & Nelken, 2004). In each trial, the total duration of the tone bursts was 7.5 sec, which was followed by a 2.5-sec period of silence. A tonotopy session consisted of 42 low-frequency trials, 42 high-frequency trials, and 12 silent baseline trials. The order of these trials was randomized within each session.

Sound Stimuli of Simple and Control Melodies

We composed five simple melodies with standard tones using GarageBand '09 software (Apple Inc., USA). Each melody consisted of low- and high-frequency blocks that were arranged in an alternating order. The low-frequency blocks consisted

of tones within the range of the D3 (164 Hz) to E4 (294 Hz) scales, and the high-frequency blocks consisted of tones within a frequency range between the C6 (1,047 Hz) and B6 (1,980 Hz) scales. Power spectral analyses (SoundForge 8.0, Sony Inc., Japan) confirmed that the prominent frequencies of all of the composed melodies were within these ranges; the prominent frequency was 194 Hz in the low-frequency blocks and 1,573 Hz in the high-frequency blocks. In each block, tones with specific pitches were presented in a rhythmic pattern based on familiar chords (such as C major or E). Each block consisted of a 5-sec stimulus and a 1-sec silent period. Each melody contained 10 to 12 blocks and was 60 to 72 sec long. At the end of every other melody, a baseline silent condition (12 sec) was inserted that did not have tone or induce auditory imagery.

We composed the control melodies from the simple melodies by disrupting the timing of the tones while preserving other properties (the total number of tones, frequency range of the tones, and duration and intensity of each tone). Thus, every simple melody had its conjugate control melody (Supplementary Figure S1). Because we disrupted the temporal structure of the control melodies, these melodies were unpredictable and awkward, and the simple and control melodies sounded quite different.

We utilized control melodies as the conjugate of the simple melodies because the beta values were measured by contrasting the simple and control melodies. To maintain only the effect of imagery in the subtracted contrasts, we composed the control melodies to be as similar as possible to their conjugate melodies.

Experimental Procedure: Behavioral Hearing Tests

All subjects participated in a behavioral test to confirm their hearing ability and the absence of tone deafness using the MBEA and tone discrimination tests. The MBEA test was administered with an online amusia test protocol (http://www.brams.umontreal.ca/plab). The tone discrimination test was administered by presenting pure sine tones in identical or different pairs (starting at 1,500 Hz with 30-Hz steps). Using a 3-down, 1-up staircase method, we measured the discrimination threshold of each subject. The subjects who received a normal MBEA score (more than 75% correct) and possessed a normal ability for tone discrimination participated in the study.

Experimental Procedure: Tonotopy and the First Melody Presentation Session

During the first functional MRI scan, all subjects participated in two scanning sessions: a tonotopy acquisition session and a melody presentation session. In the tonotopy acquisition session, the subjects passively listened to 42 high- and 42 low-frequency tones through OPTIME1 electrostatic headphones (MR ConFon, Germany) during the scan. The total running time of this session was 16 minutes.

In the melody presentation session, the subjects listened to a series of melodies during the scan. A total of 20 melodies (10 simple and 10 control melodies) were presented in a randomized order. Unlike the original melodies, these melodies included silent gaps within them (Figure 2). Because each melody contained high- and low-frequency blocks that were presented in an alternating order, two different versions of the same melody were heard, which included high-frequency and low-frequency silenced blocks. The high-frequency silenced melody included 3 to 4 silenced blocks only during the high-frequency blocks. Similarly, the low-frequency silenced melody included silenced blocks only during the low-frequency blocks

(Figure 2B). The ratio of the silenced blocks was kept below 30% of the total number of blocks. Because we composed 10 different original melodies (5 simple original melodies and 5 control original melodies) and each melody included 2 types of partially silenced blocks, a total of 20 different melodies were presented during the melody presentation session. The order of the simple and control melodies was shuffled within a session, and the total running time of a session was 21 minutes. A T1-weighted magnetization-prepared rapid gradient-echo (MP-RAGE) image was acquired at the end of the tonotopy and melody presentation session in the first and second experiments.

For the first experiment, the subjects were instructed to listen to the melodies in the scanner. To confirm the subject's alertness during the experiment, we instructed them to regularly press the buttons on the response pad (LUMINA LU-441, Cedrus Corporation, USA) while they listened to the melodies. We did not inform the subjects that the melodies consisted of repetitive blocks or when the silenced gaps would occur. Because the excerpted melodies with gaps were somewhat confusing, and because the subjects were unfamiliar with the melodies, it was difficult for the subjects to infer or to imagine the original tones during the silenced gaps in the first listening period.

Experimental Procedure: The Second Melody Presentation Session

After the first scanning experiment, five mp3 files of training melodies were presented to the subjects (the entire structure and notes of these five melodies are provided in Supplementary Figure S1 and S2). The training melodies were exactly the same as the melodies used in the previous experiments with the exception that they lacked silenced gaps. The subjects were required to listen to these melodies for at

least 10 hours per week. During this training period, we regularly sent an SMS message to the subjects (twice per week) to remind them to listen to and learn these melodies.

Approximately two weeks after the first fMRI scan, all subjects participated in the second experiment. Similar to the first scan, the subjects listened to simple and control melodies that contained several silenced blocks. The type and ratio of the silenced blocks were similar to those of the first scan, which included either high- or low-frequency silenced blocks.

The subjects were able to contextually recall the original tones during the silenced blocks in these melodies because they had trained with the simple melodies. In contrast, the control melodies were unfamiliar to these subjects, and they were unable to anticipate the musical context during the silenced gaps. Before scanning, every subject had rehearsed the tempo to follow when pressing the buttons in time to the excerpted melodies. The subjects were instructed to press the buttons in rhythm both during the melody-presentation periods and silent-gap periods. During the melody-presentation period, the subjects were instructed to press the buttons according to the tempo that they heard. During the silent-gap periods, the subjects were instructed to press the buttons according to their own estimation of the ongoing tempo of the melody. During the silent gaps of unpredictable control melodies, the subjects were instructed to continue pressing the buttons regularly.

Experimental Design: Tonotopy and Melody Presentation

To reduce the interference caused by the noise of the scanner on the sound stimuli, we used a sparse temporal acquisition paradigm (Belin, Zatorre, Hoge, Evans, & Pike, 1999; Humphries et al., 2010) during both the tonotopy and melody presentation

sessions. In the tonotopy session, a trial lasted 10 sec, consisting of a 7.5-sec tone presentation followed by a 2.5-sec gap. The echo-planner imaging (EPI) signal was only acquired during this 2.5-sec gap, and this protocol prevented the interference of the scanner noise with the tone wave sounds (TR (repetition time) =10 sec; acquisition time =1.93 sec).

During the melody presentation sessions, we used a TR of 12 sec, and the functional images were acquired within 2 sec (TR = 12 sec, delay in TR = 10 sec, acquisition time = 1.93 sec). Each melody consisted of 2 alternating frequency blocks, and each block lasted for 6 sec. Thus, a block of the same frequency type was presented every 12 sec. To accommodate the maximum hemodynamic response of the primary auditory cortex (Formisano et al., 2003) and to avoid disturbing the subject's auditory imagery, every silenced block was presented 6 sec before the beginning of the EPI acquisition. Thus, no EPI noise was heard by the subjects during the occurrence of silenced blocks.

MRI Hardware and Imaging

The fMRI data were collected on a 3.0T SIEMENS Verio scanner (Siemens, Germany) with a 12-channel head coil. The auditory stimulus was presented through an audio control unit (MR ConFon, Germany). To suppress the gradient and the ambient noise, we applied foam-padded earmuffs around the head coil. The foam padding and padded headphones provided an overall attenuation of ~30 dB of the gradient noise. The auditory stimuli were generated using custom-made software in MATLAB. A LUMINA LP-400 controller (Cedrus Corporation, USA) was used to synchronize the EPI signal with the auditory stimulus.

Functional and Structural Scan Parameters

The parameters for the scanning (flip angle, thickness and gaps) of the tonotopy acquisition session were adjusted according to a previous tonotopy study that used a similar tonotopy block design (Humphries et al., 2010). In each tonotopy session, the functional images consisted of 30 slices using a gradient-echo EPI pulse sequence (FOV (field of view) = 192 mm, matrix size = 96×96, in-plane resolution = 3×3 mm, thickness = 2 mm, TE (echo planner delay) =27 ms, TR =10 sec, delay in TR = 8.0 sec, flip angle = 77°). A total of 96 volumes were collected per session.

The scanning parameters of the melody presentation session were similar to those of a tonotopy acquisition session, with the exception of the TR. We maintained the other parameters under the same conditions because the design of the tonotopy and melody presentation exhibited similar sparse acquisition paradigms (Figures 1 and 2). In the melody presentation sessions, the functional images consisted of 30 slices using a gradient-echo EPI pulse sequence (FOV = 192 mm, matrix size = 96×96, in-plane resolution = 3×3 mm, thickness = 2 mm, TE = 27 ms, TR = 12 sec, delay in TR = 10 sec, flip angle = 77°). A total of 112 volumes were acquired per session. At the end of each session, a T1-weighted anatomical scan was collected using an MP-RAGE pulse sequence (FOV = 250 mm, matrix size = 256×256, in-plane resolution = 1×1 mm, thickness = 1 mm, flip angle = 9°, TE = 2.52 msec, TR = 1.9 sec).

Data Analysis: PAC Localization

To identify the frequency selective voxels in the temporal cortex, we employed the MarsBAR toolbox (http://marsbar.sourceforge.net) and SPM5 (Wellcome Department of Imaging Neuroscience). The tonotopy data acquired from each subject were then

separately analyzed. The functional MRI data were realigned, and the slice timing was corrected and normalized to the Montreal Neurological Institute (MNI) template brain. The normalized functional images were then smoothed with a kernel size of 4 mm full-width half-maximum (FWHM). For each subject dataset, a general linear model (GLM) was applied to the functional voxel time series, and the T-values for the contrast of the high- versus low-frequency blocks were measured.

The regions of interest (ROIs) were defined as significant with a family-wise error (FWE) (corrected p-value < .05) cluster of activation signals in the anatomical region of the temporal cortical areas of both hemispheres. We extracted the relevant voxels using the MarsBAR toolbox, and their anatomical coordinates were defined as high- or low-frequency selective voxels in the PAC.

Data Analysis: A1 and R localization

T1-inflated images were generated using the FreeSurfer software (Martinos Center for Biomedical Imaging, Harvard Medical School) to visualize the tonotopic map of each subject. SPM-analyzed non-normalized tonotopy contrast images were co-registered with an inflated T1 image through the FreeSurfer MATLAB code. Next, we examined the presence of mirror-symmetric frequency gradients in each hemisphere. Previous studies (Da Costa et al., 2011; Formisano et al., 2003; Humphries et al., 2010) have reported that the cortical areas responding preferably to low frequencies are located at the center of the PAC, and the cortical areas responding preferably to high frequencies are located on both sides of the anterior-posterior axis. The more posterior region of this axis corresponds to human A1, and the more anterior portion of this axis corresponds to human R (Da Costa et al., 2011). If there was a mirror-symmetric frequency progression (high-low-low-high) along this axis, then we labeled the high-

frequency areas of the more posterior region as a high-frequency area of A1 (HA1) and the anterior region as a high-frequency area of R (HR) (Figure 5).

Data Analysis: Beta Estimate Analysis

We constructed the model of a musical imagery task under the four following conditions: high-frequency silenced gaps in a simple melody (HS), low-frequency silenced gaps in a simple melody (LS), high-frequency silenced gaps in a control melody (HC) and low-frequency silenced gaps in a control melody (LC). The onset and duration of each gap were provided for the first-level specification in SPM, and t-tests were used to compare differences between the simple and control silenced gaps. Thus, two contrasts were acquired for each subject (HS > HC; T-contrast vector [HS LS HC LC] = [1 0 -1 0], LS > LC; T-contrast vector [HS LS HC LC] = [0 1 0 -1]).

We employed MarsBAR to extract the beta values of the ROI for each subject. Two contrast images (i.e., HS > HC and LS > LC) were used in the ROI analysis. ROI-related estimates were extracted from these contrast images and scaled to the raw data. The extracted values were set to 'scale grand mean to zero' and saved in the MATLAB workspace. We used the FreeSurfer software for the subfield (A1 and R) analysis and overlaid two non-normalized contrast images (i.e., HS > HC and LS > LC) on the inflated surface. Next, labels of the two subfields (i.e., the high-frequency region of A1 and the high-frequency region of R) were added to the surface. Each label datum contained the related beta values within its area. We used this information in additional statistical analyses.

RESULTS

We found frequency selective voxels in the auditory cortex that were significant for each subject (Figure 3). Consistent with previous studies (Da Costa et al., 2011; Formisano et al., 2003; Humphries et al., 2010), high- and low-frequency selective voxels residing near Heschl's gyrus and mirror-symmetric gradients (high-low-low-high) were observed along the anterior-posterior axis of the inflated cortex (Figure 3). These tonotopy results guided the localization of the primary auditory cortical region in each subject, and this ROI information was traced in successive experiments.

Insert Figure 3

Next, we examined how the extent of PAC activation was associated with the type of silenced block. To interpret these results, two types of averaged beta estimates were obtained from each subject. T-contrasts of the simple high-frequency silenced blocks were compared to the control high-frequency silenced blocks, and the simple low-frequency silenced blocks were compared to the control low-frequency silenced blocks.

No significant differences were found between the high- and low-frequency ROIs in the subjects prior to the their familiarization with the simple melodies (Figure 4A). In the high-frequency silenced blocks, beta estimates of high- and low-frequency ROIs did not significantly differ among the subjects ($P = .54$, paired Wilcoxon signed-rank test, two-tailed, alpha level $= .05$). A similar result was obtained for the low-frequency silenced blocks ($P = .53$, paired Wilcoxon signed-rank test, two-tailed, alpha level $= .05$). This result indicates that the neural activity associated with high-

and low-frequency ROIs did not differ during high- and low-frequency silenced gaps prior to the subject's familiarization with (and learning of) the simple melodies.

Insert Figure 4

However, when the subjects were trained with the simple melodies, the activation patterns of the auditory cortex were altered by the frequency characteristics of the melodies (Figure 4B). After the subjects learned the simple melodies, the beta estimates of the high-frequency ROIs were significantly higher than those of the low-frequency ROIs during the high-frequency silenced blocks (P = .04, paired Wilcoxon signed-rank test, one-tailed, alpha level = .05). This frequency-related modulation was also observed in the low-frequency silenced blocks: the beta estimates of the low-frequency ROIs were higher than those of the high-frequency ROIs during the low-frequency silenced blocks (P = .02, paired Wilcoxon signed-rank test, one-tailed, alpha level = .05). Because the subjects were familiar with and had learned these simple melodies, they could imagine the original tones even when the simple melody was unexpectedly silenced; however, this did not occur with the control melodies.

Thus, this result indicates that the subjective experience of musical imagery in simple melodies is closely associated with the functionally relevant areas of the PAC: when imagining high-frequency melodies, the area that processes high-frequency stimuli was more active than the low-frequency processing area, and similarly, the low-frequency processing area was more active than the high-frequency processing area during the imagery of low-frequency melodies.

These results cannot be accounted for by the characteristics of the selected melodies or presented tones near the silenced blocks. Although the subjects were

presented with the same melodies before and after learning, the average beta value of each ROI was significantly higher only when the subjects had learned the simple melodies. This increase in beta values was observed for both the high-frequency ($P < .001$, paired Wilcoxon signed-rank test, two-tailed, alpha level $= .05$) and low-frequency ($P < .001$, paired Wilcoxon signed-rank test, two-tailed, alpha level $= .05$) ROIs. Thus, we confirmed that the familiarity of the simple melodies induced a significant activation of the PAC during the silenced gaps in the second experiment.

To assess whether the frequency-related modulation of the PAC originated from earlier cortical structures, we divided the PAC into two subdivisions, A1 and R, on the basis of recent tonotopic results (Da Costa et al., 2011; Humphries et al., 2010). It has been shown that along the anterior-to-posterior axis, there are frequency gradients of preferred frequencies from high to low (A1) and reverse frequency gradients from low to high (R), which follows A1 (Da Costa et al., 2011). In our tonotopy results, 10 of the 13 subjects showed mirror-symmetric progression in each hemisphere (Figure 5), which was traced in the subsequent analyses.

Insert Figure 5

Because the low-frequency regions are located at the center of the PAC, the A1 and R areas of low frequency are adjacent to each other (Da Costa et al., 2011; Formisano et al., 2003). Thus, it was difficult to divide the low-frequency regions into two subfields using our tonotopy design. In contrast, the high-frequency regions were located at both ends of the PAC region and were discernible if there was mirror-symmetric frequency progression. Within the PAC, the more posterior of the high-frequency regions corresponded to A1, and the anterior region corresponded to R.

Thus, we referred to these regions as the high-frequency area of A1 (HA1) and the high-frequency area of R (HR).

Insert Figure 6

Before the subjects learned the simple melodies, the beta values for these two regions were not significantly different in the familiar high- and low-frequency silenced blocks (HA1: P = .76, paired Wilcoxon signed-rank test, two-tailed, alpha level = .05; HR: P = .72, paired Wilcoxon signed-rank test, two-tailed, alpha level = .05) (Figure 6A). However, when the subjects were trained with simple melodies, only the HA1 field showed a content-related modulation of auditory imagery (Figure 6B). HA1 was more active during the imagery of high-frequency melodies than during the low-frequency melodies (P = .02, paired Wilcoxon signed-rank test, two-tailed, alpha level = .05), but this relationship was not observed in the HR field (P = .88, paired Wilcoxon signed-rank test, two-tailed, alpha level = .05). These results suggest that the frequency-related modulation observed in high-frequency ROIs might originate from a specific subfield of the PAC, i.e., the high-frequency area of A1, but not from the relevant area of R.

DISCUSSION

Previous studies demonstrated that anatomically defined PAC areas were modulated by top-down processing during auditory imagery (Kraemer et al., 2005) and attention (Petkov et al., 2004). However, because the topographical location and size of the PAC varies greatly across individuals (Rademacher, 2001), the exact location and

border of the PAC cannot be reliably inferred from an anatomical image alone. Here, we show the modulation of functionally defined areas of the PAC based on the subjects' tonotopic maps during auditory imagery.

First, we found that increased activity of the PAC during auditory imagery was strongly related to the frequency property of the learned melodies. The imagery of the high-frequency melodies led to an increased activation of high-frequency processing areas in the PAC compared to low-frequency processing areas. Similarly, the imagery of the low-frequency melodies led to an increased activation of low-frequency selective areas. This result indicates that the extension of top-down processing to the PAC is frequency specific.

Furthermore, we found that this frequency-related modulation of the PAC is related to subfield A1 but not to subfield R. Although we only examined the high-frequency region of these subfields, this result suggests that the increased activity of high-frequency ROIs during the imagery of high-frequency melodies might have originated from subfield A1.

Studies of auditory cortical hierarchy have suggested that simple auditory stimuli (i.e., pure tones) are mainly processed in the core region near Heschl's gyrus, whereas more complex stimuli (i.e., speech) are likely to drive the activity of the surrounding auditory cortical areas (Okada et al., 2010; Rauschecker & Scott, 2009). This relationship is similar during auditory imagery. The imagery of semantic sounds or lyrics to a melody only recruits secondary auditory cortical regions (Bunzeck et al., 2005; Kraemer et al., 2005; Yao, Belin, & Scheepers, 2011), whereas the imagery of pure and instrumental tones extend to regions of the primary auditory cortex (Kraemer et al., 2005; Yoo, Lee, & Choi, 2001).

Because we utilized simple melodies with no lyrics or speech, our results demonstrating the increased activation of primary auditory regions during auditory imagery is consistent with previous findings; the imagery of simple auditory stimuli extended to the earliest auditory area, the PAC, and to its subfield A1. Our findings on the frequency-related modulation of these areas suggest that top-down processing might selectively recruit early auditory regions rather than induce a broad cortical activation.

Behavioral evidence of imagery during the second experiment

Previous studies used rigorous methods to validate the occurrence of sensory imagery. In a visual imagery task, subjects were required to press buttons to indicate whether a spatial probe was inside or outside the imagery contour of a wedge (Slotnick et al., 2005). In an auditory imagery task, subjects were trained to press buttons at the end of the melody in both imagery and perceptive conditions and the latency between these data was measured to confirm the occurrence of imagery (Halpern & Zatorre, 1999). Although this type of quantitative validation of auditory imagery was not attempted in this study, the following factors support the inference that adequate imagery occurred during the second experiment.

First, we measured the number of total button presses made during both the presentation of melodies and the silent gaps, which required imagery. In the high- and low-frequency blocks of each melody (i.e., a high-frequency block in simple melody #3), the total number of button presses made during the presentation of the tones and silent gaps was not significantly different, with the exception of melody #1 (Figure S4). These results suggested that the subjects might have correctly imagined the original melodies during the silent gaps in the second experiment because the total

number of tones was different between each melody. Second, the raster plot of the button responses performed by a subject showed that he/she had pressed the buttons in a similar manner during the presentation of the simple melody tones with silent gaps (Figure S5), although the exact timing and patterns varied among the subjects. These results are consistent with the anecdotal reports that subjects were imagining learned melodies during the silent gap periods.

Because the subjects were also required to press buttons during the presentation of the control melodies, we measured the ratio of button presses during the control melodies to button presses during the simple melodies, which was $23.3 \pm 12.4\%$ (mean $\pm$ s.d.) across all subjects. This result demonstrated that the subjects also pressed buttons during the presentation of the control melodies.

Top-Down Modulation and Plasticity in the Auditory Cortex

It has been previously found that musical imagery and perception draw on similar neural mechanisms in the auditory cortex (Halpern & Zatorre, 1999; Herholz, Lappe, Knief, & Pantev, 2008; 2009b; Zatorre & Halpern, 2005). Imagery of familiar melodies elicited prominent mismatch negativity (MMN) (Herholz et al., 2008; Herholz, Lappe, Knief, & Pantev, 2009b), which is similar to the MMN response in violations of predominant musical patterns (Herholz, Lappe, & Pantev, 2009a). These results suggested that the early response of auditory cortical neurons in imagery and perception might share similar neural principles (Herholz, Lappe, & Pantev, 2009a). Our results are consistent with these previous findings; the auditory area that processes perceived stimuli also responded to imagery of familiar melodies.

It is also well established that associative learning induces a highly learning-specific plasticity in the auditory cortex (Bakin & Weinberger, 1990; Bakin, Lepan, &

Weinberger, 1992; Bao et al., 2001; Bieszczad & Weinberger, 2010a; 2010b; 2012; Recanzone et al., 1993; Rutkowski & Weinberger, 2005). In animals, frequency receptive fields shifted to the conditioned frequency during classical fear conditioning (Bakin & Weinberger, 1990) and the area of tonotopic representation increased in the behaviorally trained frequency bands (Bao et al., 2001; Bieszczad & Weinberger, 2010a; 2010b; 2012; Recanzone et al., 1993; Rutkowski & Weinberger, 2005). Moreover, the level of behavioral importance of the conditioned stimulus (Rutkowski & Weinberger, 2005), and the enhancement of memory strength (Bieszczad & Weinberger, 2010b) were related to the representational area in the primary auditory cortex.

This kind of associative auditory cortical plasticity could be induced by a brief training of five trials (Edeline & Weinberger, 1993) and could be retained up to 8 weeks following training (Weinberger, Javid, & Lepan, 1993), whereas modification of tonotopic maps required more training trials (~750 trials per daily session) and longer training periods (several weeks) to develop an increased response in the represented area for the learned frequency (Recanzone et al., 1993). In our study, subjects were required to train with simple melodies for approximately 10 hours in 2 weeks, although we did not quantitatively measure the amount of training time in each subject. Previous human study have shown that modulation of neural responses to the learned frequency was obtained by several training trials before PET scanning (Morris et al., 1998). Thus, the increased beta values we measured after learning simple melodies may have originated from learning-induced plastic changes in the primary auditory cortex. This observation suggests that neural responses to learned frequency could be increased by a top-down signal that contains learning-related frequency information, as well as by a bottom-up perceptual stimulus.

The BOLD Response to Each Melody Type

To confirm whether increased beta values in our results were caused by learning, we measured the BOLD response to the silenced blocks in each melody type (simple and control melodies) and compared their values before and after the learning of simple melodies (Figure S3).

The BOLD response to the simple melodies was significantly increased after simple melodies were learned compared to before ($P < .001$, paired Wilcoxon signed-rank test), both in high- and low-frequency silenced conditions (high-frequency silenced condition, $P < .01$; low-frequency silenced condition, $P < .05$). However, the BOLD response to the control melodies was not significantly different before and after learning ($P = .47$), both in high- and low-frequency silenced conditions (high-frequency silenced condition, $P = .65$; low-frequency silenced condition, $P = .55$). These results suggest that training significantly enhanced the auditory activity associated with the simple but not control melodies.

After the simple melodies were learned, the BOLD responses to the simple melodies were significantly higher in both the high- and low-frequency silenced blocks (Figure S3B: high-frequency silenced blocks, $P < .001$; low-frequency silenced blocks, $P < .01$). The BOLD differences between the simple and control melodies were significantly larger in high-frequency silenced blocks compared to low-frequency silenced blocks ($P < .001$). This result is consistent with the findings of higher beta values in high-frequency silenced blocks observed in Figure 4B, because the beta values were acquired by subtracting control melodies from simple ones. This relationship of higher beta values was also observed in Figure 6B. In HA1, the beta values of high-frequency silenced blocks were significantly larger than low-frequency

silenced blocks. Because the beta values of Figure 6 were derived from high frequency ROIs (HA1 and HR), this result is also consistent with the result of Figure 4B, which showed larger beta values in high-frequency silenced blocks than low-frequency silenced ones. Taken together, these results suggest that increased neural responses of high frequency ROI in high-frequency silenced blocks than low-frequency silenced ones (Figure 4B) might be originated from the region of HA1, rather than HR.

In contrast, before the simple melodies were learned, these relationships were not observed (Figure S3A). The neural activity of the tonotopic regions was not significantly different in the high-frequency silenced blocks, and the BOLD response to the control melodies was higher than to the simple melodies in the low-frequency silenced blocks ($P < .05$). These results suggest that frequency-related learning increased the neural responses in the primary auditory cortex with a frequency selectivity; enhanced neural activities were associated with simple melodies rather than control melodies. This finding is also consistent with a previous human study that showed specific cortical plasticity to the conditioned frequency in the auditory cortex (Morris et al., 1998).

The Potential Role of Subjective Loudness on Beta Values

In our beta analysis results, it should be noted that the high-frequency silenced blocks exhibited larger beta values than the low-frequency silenced blocks (Figure 4). This result might not have originated from the absolute volume of the melodies because we adjusted the intensity of the tones (dB) when composing each melody. Furthermore, because the tone intensity did not seem to be a critical factor affecting the activity of the auditory cortex, particularly in the PAC (Brechmann, Baumgart, & Scheich, 2002;

Woods et al., 2009), the sound volume difference could not explain the observed difference in the beta values of the high- and low-frequency silenced blocks.

Although the absolute sound volumes were equally adjusted, the subjective loudness of the tones, which is measured by 'phons,' may have been different between the high- and low-frequency tones. According to the equal-loudness contours (ISO 226: 2003), the same dB of high-frequency tones has a larger phon value than that of low-frequency tones. Thus, the high frequency tones may have been perceived as louder than the low-frequency tones in our experiment. If the subjective sense of loudness in learning melodies affects the vividness of the imagery, then the observed higher beta values in the high-frequency silenced blocks might have originated from the different phon values of our melodies.

Similarities between Auditory and Visual Areas during Top-Down Processing

Our findings of frequency-selective modulation in the early auditory cortex are consistent with previous findings in the visual system. On the basis of the retinotopic maps of the primary visual areas (Kamitani & Tong, 2005), several fMRI studies have shown that V1 actively participates in top-down processing during spatially selective visual attention (Somers, Dale, Seiffert, & Tootell, 1999). Furthermore, top-down modulation is closely associated with the basic properties of V1, including orientation selectivity (Kamitani & Tong, 2005) and retinotopic characteristics (Slotnick et al., 2005; Thirion et al., 2006). Thus, in the visual system, top-down effects are thought to extend to the functionally relevant areas of V1 in the first cortical stage. Our results showing that the frequency-selective voxels in the PAC associated with the auditory imagery context are more active than the context-irrelevant voxels suggest that

auditory top-down processing also extends to the functionally related areas of the PAC on the basis of tonotopic properties.

Although the PAC has a similar hierarchical position to that of other early sensory areas, such as V1, it is likely that the PAC is functionally analogous to higher areas of other sensory systems, such as the inferior temporal cortex (IT) in the visual system (King & Nelken, 2009). The IT is known to be involved in the regulation of visual attention (Chelazzi et al., 1993), but the effects of its retinotopic properties are still unclear. In contrast, other early visual areas, such as V1, V2, V3, V4 and MT, have well-characterized retinotopic properties that actively participate in top-down processing with topographic influences (Slotnick et al., 2005). Based on the findings of the present study, we speculate that the auditory system shares similar neural principles with the visual system in terms of top-down processing; in particular, A1, which is a subfield of the PAC, is responsible for auditory imagery as V1 is for visual imagery, namely, in a top-down manner.

Author contributions

J.H. Oh and J. Jeong conceived of this study and wrote the manuscript. J.H. Oh, J.H. Kwon, and P.S. Yang designed and conducted the experiment. J.H. Oh and J.H. Kwon analyzed the behavioral and fMRI data.

REFERENCES

Bakin, J. S., & Weinberger, N. M. (1990). Classical conditioning induces CS-specific receptive field plasticity in the auditory cortex of the guinea pig. *Brain research*, *536*(1-2), 271–286.

Bakin, J. S., Lepan, B., & Weinberger, N. M. (1992). Sensitization induced receptive field plasticity in the auditory cortex is independent of CS-modality. *Brain research*, *577*(2), 226–235.

Bao, S., Chan, V. T., & Merzenich, M. M. (2001). Cortical remodelling induced by activity of ventral tegmental dopamine neurons. *Nature*, *412*(6842), 79–83.

Bao, S., Chan, V. T., Zhang, L. I., & Merzenich, M. M. (2003). Suppression of cortical representation through backward conditioning. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(3), 1405–1408.

Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., & Pike, B. (1999). Event-related fMRI of the auditory cortex. *NeuroImage*, *10*(4), 417–429.

Bieszczad, K. M., & Weinberger, N. M. (2010a). Learning strategy trumps motivational level in determining learning-induced auditory cortical plasticity. *Neurobiology of Learning and Memory*, *93*(2), 229–239.

Bieszczad, K. M., & Weinberger, N. M. (2010b). Representational gain in cortical area underlies increase of memory strength. *Proceedings of the National Academy of Sciences*, *107*(8), 3793–3798.

Bieszczad, K. M., & Weinberger, N. M. (2012). Extinction reveals that primary sensory cortex predicts reinforcement outcome. *European Journal of Neuroscience*, Sensory cortical plasticity induced by extinction, *35*(4), 598–613.

Brechmann, A., Baumgart, F., & Scheich, H. (2002). Sound-level-dependent

representation of frequency modulations in human auditory cortex: a low-noise fMRI study. *Journal of Neurophysiology*, *87*(1), 423–433.

Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H., & Jancke, L. (2005). Scanning silence: Mental imagery of complex sounds. *NeuroImage*, *26*(4), 1119–1127.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*(6427), 345–347.

Da Costa, S., Van Der Zwaag, W., Marques, J. P., Frackowiak, R. S. J., Clarke, S., & Saenz, M. (2011). Human Primary Auditory Cortex Follows the Shape of Heschl's Gyrus. *Journal of Neuroscience*, *31*(40), 14067–14075.

Edeline, J. M., & Weinberger, N. M. (1993). Receptive field plasticity in the auditory cortex during frequency discrimination training: selective retuning independent of task difficulty. *Behavioral neuroscience*, *107*(1), 82–103.

Formisano, E., Kim, D., Di Salle, F., van de Moortele, P., Ugurbil, K., & Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, *40*(4), 859–869.

Fritz, J. B., David, S. V., Radtke-Schuller, S., Yin, P., & Shamma, S. A. (2010). Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *Nature Publishing Group*, *13*(8), 1011–1019.

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cerebral cortex (New York, NY : 1991)*, *9*(7), 697–704.

Herholz, S. C., Lappe, C., & Pantev, C. (2009a). Looking for a pattern: An MEG study on the abstract mismatch negativity in musicians and nonmusicians. *BMC Neuroscience*, *10*(1), 42.

Herholz, S. C., Lappe, C., Knief, A., & Pantev, C. (2008). Neural basis of music imagery and the effect of musical expertise. *European Journal of Neuroscience*, *28*(11), 2352–2360.

Herholz, S. C., Lappe, C., Knief, A., & Pantev, C. (2009b). Imagery Mismatch Negativity in Musicians. *Annals of the New York Academy of Sciences*, *1169*(1), 173–177.

Humphries, C., Liebenthal, E., & Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *NeuroImage*, *50*(3), 1202–1211.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*(5), 679–685.

King, A. J., & Nelken, I. (2009). Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nature Neuroscience*, *12*(6), 698–701.

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Maljkovic, V., Weise, S. B., Chabris, C. F., Hamilton, S. E., et al. (1993). Visual mental imagery activates topographically organized visual cortex: PET investigations. *Journal of Cognitive Neuroscience*, *5*(3), 263–287.

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, *2*(9), 635–642. doi:10.1038/35090055

Kraemer, D. J. M., Macrae, C. N., Green, A. E., & Kelley, W. M. (2005). Musical imagery: sound of silence activates auditory cortex. *Nature*, *434*(7030), 158.

Mechelli, A. (2004). Where Bottom-up Meets Top-down: Neuronal Interactions during Perception and Imagery. *Cerebral Cortex*, *14*(11), 1256–1265.

Menning, H., Roberts, L. E., & Pantev, C. (2000). Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *Neuroreport*, *11*(4), 817–822.

Morris, J. S., Friston, K. J., & Dolan, R. J. (1998). Experience-dependent modulation of tonotopic neural responses in human auditory cortex. *Proceedings. Biological sciences / The Royal Society*, *265*(1397), 649–657.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., Serences, J. T., et al. (2010). Hierarchical Organization of Human Auditory Cortex: Evidence from Acoustic Invariance in the Response to Intelligible Speech. *Cerebral Cortex*, *20*(10), 2486–2495.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*, 58–75.

Petkov, C. I., Kang, X., Alho, K., Bertrand, O., Yund, E. W., & Woods, D. L. (2004). Attentional modulation of human auditory cortex. *Nature Neuroscience*, *7*(6), 658–663.

Rademacher, J. (2001). Probabilistic Mapping and Volume Measurement of Human Primary Auditory Cortex. *NeuroImage*, *13*(4), 669–683.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724.

Recanzone, G. H., Schreiner, C. E., & Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *13*(1), 87–103.

Rutkowski, R. G., & Weinberger, N. M. (2005). Encoding of learned importance of sound by magnitude of representational area in primary auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(38), 13664–13669.

Sack, A. T., Jacobs, C., De Martino, F., Staeren, N., Goebel, R., & Formisano, E. (2008). Dynamic Premotor-to-Parietal Interactions during Spatial Imagery. *Journal of Neuroscience*, *28*(34), 8417–8429.

Slotnick, S., Thompson, W., & Kosslyn, S. (2005). Visual Mental Imagery Induces Retinotopically Organized Activation of Early Visual Areas. *Cerebral Cortex*.

Somers, D. C., Dale, A. M., Seiffert, A. E., & Tootell, R. B. (1999). Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(4), 1663–1668.

Stokes, M., Thompson, R., Cusack, R., & Duncan, J. (2009). Top-Down Activation of Shape-Specific Population Codes in Visual Cortex during Mental Imagery. *Journal of Neuroscience*, *29*(5), 1565–1572.

Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J.-B., Lebihan, D., & Dehaene, S. (2006). Inverse retinotopy: inferring the visual content of images from brain activation patterns. *NeuroImage*, *33*(4), 1104–1116.

Ulanovsky, N., Las, L., Farkas, D., & Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *Journal of Neuroscience*, *24*(46), 10440–10453.

Voisin, J. (2006). Listening in Silence Activates Auditory Areas: A Functional Magnetic Resonance Imaging Study. *Journal of Neuroscience*, *26*(1), 273–278.

Weinberger, N. M. (2004). Specific long-term memory traces in primary auditory cortex. *Nature Reviews Neuroscience*, *5*(4), 279–290.

Weinberger, N. M. (2007). Associative representational plasticity in the auditory cortex: a synthesis of two disciplines. *Learning & memory (Cold Spring Harbor, N.Y.)*, *14*(1-2), 1–16.

Weinberger, N. M., Javid, R., & Lepan, B. (1993). Long-term retention of learning-

induced receptive-field plasticity in the auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(6), 2394–2398.

Woods, D. L., Stecker, G. C., Rinne, T., Herron, T. J., Cate, A. D., Yund, E. W., Liao, I., et al. (2009). Functional Maps of Human Auditory Cortex: Effects of Acoustic Features and Attention. *PLoS ONE*, *4*(4), e5183.

Yao, B., Belin, P., & Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience*, *23*(10), 3146–3152.

Yoo, S. S., Lee, C. U., & Choi, B. G. (2001). Human brain mapping of auditory imagery: event-related functional MRI study. *Neuroreport*, *12*(14), 3045–3049.

Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the Mind's Ear: A PET Investigation of Musical Imagery and Perception. *Journal of Cognitive Neuroscience*, *8*(1), 29–46.

Zatorre, R., & Halpern, A. (2005). Mental Concerts: Musical Imagery and Auditory Cortex. *Neuron*, *47*(1), 9–12.

Figure Legends

Figure 1. The detailed design of tonotopy sound stimuli. (A) Waveform of a tone (B) Time structure of a tone block that consisted of tone bursts. Each frequency tone block lasted 7.5 sec and was followed by 2.0 sec of EPI scanning. (C) Schematic time-frequency relationship of tone sequence. The order of the high- and low-frequency stimuli and silenced baseline were randomized within a session.

Figure 2. A schematic representation of simple and control melodies. (A) Two frequency types of tone blocks were presented in an alternating order. A simple melody was composed of a familiar chord and regular tempo. In the control melody, the tones were jittered to preserve similar physical properties (the number of tones, frequency range and volume) to those of the simple melody. The subjects listened to the melodies for the occasionally silenced blocks (30% of total blocks) during scanning. (B) A BOLD signal was acquired every 12 sec to avoid interrupting the subjective experience of the auditory imagery.

Figure 3. (A) An individual subject's tonotopy results are rendered onto a volumetric brain image of the right hemisphere. (B) The resulting tonotopic maps are projected onto each subject's inflated cortical surface (anterior view of the right hemisphere). (C) The false-discovery rate (FDR) color-coded tonotopic results are presented on the flattened surface. The region outlined by a white dashed line represents high-frequency (1,050, 1,500 and 1,950 Hz) responsive voxels; the region outlined by a black dashed line represents low-frequency (175, 250 and 325 Hz) responsive voxels. (D, E, F) Tonotopic representations of the volumetric (D), the anterior view of the

inflated cortical surface (E) and the flattened surface (F) of the left hemisphere. Data presented in D, E, and F are from different subjects.

Figure 4. The beta estimates acquired from individual tonotopic coordinates. The beta values are shown for the simple silenced blocks versus the control silenced blocks in each frequency condition. (A) Before becoming familiar with the simple melodies. (B) After learning the simple melodies. After the first experiment, the subjects were required to repeatedly listen to the simple melodies without gaps for one week. The error bars denote the s.e.m. across subjects.

Figure 5. Localizations of two subfields of the PAC, A1 and R. Each panel represents the high-frequency responsive areas in different subjects. Based on the frequency-gradient progression, we delineated the outer border of the high-frequency area of A1 (HA1) and the high-frequency area of R (HR). The more posterior regions of the frequency maps correspond to HA1, and the anterior regions correspond to HR. The left column represents the result from the right hemisphere, and the right column represents the left hemisphere. Each figure represents the results obtained from a different subject.

Figure 6. The beta estimates acquired from two subfields, HA1 and HR. The beta values were measured for the simple versus control melodies in the high- and low-frequency silenced blocks. (A) Before learning the simple melodies. In both HA1 and HR, the beta values were not significantly different between the high- and low-frequency silenced blocks. (B) After the learning simple melodies. Only the HA1 showed significantly different beta values between the high- and low-frequency

silenced blocks. The beta value of HA1 was higher during the occurrence of the high-frequency silenced blocks (imagery of high-frequency melodies) than the low-frequency silenced blocks (imagery of low-frequency melodies) ($P < .05$, paired Wilcoxon signed-rank test). This significant relationship was not observed in the HR region.