Psychophysiology, 48 (2011), 415–419. Wiley Periodicals, Inc. Printed in the USA.

DOI: 10.1111/j.1469-8986.2010.01070.x

Pitch and loudness information encoded in auditory imagery as revealed by event-related potentials

JIANHUI WU,[a] ZULIN YU,[b] XIAOQIN MAI,[c] JINGHAN WEI,[a] AND YUEJIA LUO[d]

[a]Institute of Psychology, Chinese Academy of Sciences, Beijing, China
[b]State Key Laboratory of Brain and Cognitive Science, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China
[c]Center for Human Growth and Development, University of Michigan, Ann Arbor, Michigan
[d]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China

Abstract

Two experiments using the ERP method and a task that involved comparing an imagined-S1 (the first stimulus) with a perceived-S2 (the second stimulus) were conducted to investigate whether imagined auditory representations encode pitch and loudness information. It was found that the amplitude of the imagery-related late positive complex (LPC) decreased with pitch but increased with loudness of the imagined sound, which was consistent with amplitude modulations of the auditory perception-related N1 component, thereby providing the first neural evidence that auditory imagery encodes perceptual attributes of auditory experiences.

Descriptors: Auditory imagery, Pitch, Loudness, Event-related potentials, Late positive complex

The most influential theory of mental imagery is Kosslyn's depictive (picture-like) representation of imagery processing, emphasizing the close similarities between imagery and perception (Kosslyn, Thompson, & Ganis, 2006). Neuroimaging and neuropsychological data indicate that the modality-specific visual and auditory cortex is involved in visual and auditory imagery, respectively (Aleman, Formisano, Koppenhagen, Hagoort, de Haan, & Kahn, 2005; Kraemer, Macrae, Green, & Kelley, 2005; Slotnick, Thompson, & Kosslyn, 2005; for reviews, see Kosslyn, Ganis, & Thompson, 2001; and Kosslyn & Thompson, 2003). In the visual modality, neuroimaging data further indicate that, during visual imagery, spatial information is mapped in the primary visual cortex (Klein, Dubois, Mangin, Kherif, Flandin, et al., 2004; Kosslyn, Thompson, Kim, & Alpert, 1995), providing strong evidence for the depictive theory. In the auditory modality, behavioral studies also indicate that auditory imagery represents perceptual attributes of sound such as loudness and pitch (Farah & Smith, 1983; Intons-Peterson, 1980; for reviews, see Intons-Peterson, 1992; and Hubbard, 2010). In Intons-Peterson's (1980) study, the time required to mentally adjust the loudness of one imagined sound to that of another imagined sound increased with the distance between the loudness of the imagined sounds, suggesting auditory imagery encodes loudness. Farah and Smith (1983) found that one could detect the target sound at a lower loudness threshold when the pitch of the imagined sound is the same as the following target sound, suggesting auditory imagery also encodes pitch. These behavioral results may be confounded by the subject's knowledge of the task; however, for example, subjects may use their prior knowledge to simulate what would happen in a real-world situation (Pylyshyn, 2002). Evidence from studies of brain activity, which cannot be simulated by a subject's tacit knowledge, however, is still scarce.

The aim of the present study was to provide neural evidence that auditory imagery encodes pitch and loudness information by using the event-related potential (ERP) method. Auditory stimuli typically elicit an N1 component in the ERP with a peak latency of about 100 ms and a scalp distribution over frontocentral areas. The amplitude of the N1 is related to perception and increases with sound intensity but decreases with sound frequency (for a review, see Näätänen & Picton, 1987). Previous ERP studies have shown that a late positive complex (LPC) is related to the generation of both visual and auditory imagery (Farah, Peronnet, Weisberg, & Monheit, 1989; Meyer, Elmer, Baumann, & Jancke, 2007; Wu, Mai, Chan, Zheng, & Luo, 2006). In the present study, we investigated whether the amplitude modulation of this imagery-related LPC will follow the same pattern as that of the auditory perception-related N1. If imagery is similar to perception, as the depictive theory suggests, the amplitude of the imagery-related LPC should also increase with loudness (the subjective dimension of intensity), but decrease with pitch (the subjective dimension of frequency) of the imagined sound.

We developed a paradigm called "imagined-S1/perceived-S2 comparison" (Wu, Mai, Yu, Qin, & Luo, 2010). Before starting the EEG (electroencephalography) recordings, participants learned to associate each of three visual cues with one of three pure tones varying in pitch (Experiment 1) or loudness (Exper-

This work was supported by the NSFC (30930031, 30900442), and the Ministry of Education, China (PCSIRT, IRT0710), and the Project for Young Scientists Fund, IP, CAS (O9CX042004), and the GSCAS (2006). We thank Professor Sun Junxian for the use of the sound-level meter and Justin Ryder for his editorial assistance.

Address correspondence to: Yue-jia Luo, State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, 19 Xin Jie Kou Wai Street, Beijing, 100875, China. E-mail: luoyj@bnu.edu.cn

iment 2). During EEG recording, participants were presented with one of these three visual cues and asked to imagine hearing the corresponding tone (imagined-S1). After a short delay, one of the three tones was actually presented (perceived-S2), and participants were required to make a same/different comparison between the real tone and the previously imagined one. The effective use of imagery was ensured by imagery training prior to EEG recordings and evaluated by means of imagery vividness and subvocalization questionnaires after the recording. The content of the questionnaires has been reported in our previously published paper that focused on both behavioral and ERP results associated with the perceived-S2 stimulus (Wu et al., 2010). In the present paper, we focus on the ERP elicited by the imagined-S1 stimulus in the same experiment in addition to interpreting the results of the perceived-S2 period from the perspective of mental imagery.

Methods

Participants

As reported previously (Wu et al., 2010), data were obtained from 22 right-handed participants (mean age 21.6 $\pm$ 0.9 years, 10 males) in Experiment 1 and 23 right-handed participants (mean age 21.3 $\pm$ 1.3 years, 11 males) in Experiment 2. Two participants were excluded in Experiment 1 and one participant in Experiment 2 due to excessive movement artifacts. All reported normal hearing and normal or corrected-to-normal vision. None of the participants reported a history of neurological or psychiatric disorders. Subjects were paid for their participation and gave their informed consent.

Stimuli

Three shapes (circle, square, and triangle) were chosen as visual cues to induce auditory imagery. Cues were presented on a computer screen placed 75 cm away from the participants' eyes and subtended at an angle of approximately 2° both horizontally and vertically. In Experiment 1, three pure tone bursts of different pitch (low-pitch: 400 Hz; medium-pitch: 1000 Hz; and high-pitch: 2500 Hz) at a constant loudness (75 dB SPL [sound pressure level]) were chosen as the imagined sounds. In Experiment 2, tones were at a constant pitch (400 Hz) while their loudness varied (soft: 50 dB SPL; medium: 75 dB SPL; and loud: 85 dB SPL). Tone bursts (250 ms duration, including the 25 ms rise/fall times) were broadcast from a loudspeaker (Fostex FE107E, Tokyo, Japan) placed beside the computer screen. SPLs were measured with a condenser microphone (Brüel & Kjaer 4135, Nærum, Denmark) and a sound-level meter (Brüel & Kjaer 2610), accurate to $\pm$ 1 dB over 0.1–10 kHz. The association between visual cues and tones was counterbalanced across participants.

Procedure

The procedure in both Experiment 1 and Experiment 2 was as follows. Participants were seated in a relaxed position on a comfortable chair in a dimly lit, sound-attenuated, and electrically isolated room. There were three training sessions before the EEG recordings. The first session was a familiarization session in which the visual cues and corresponding pure tones were presented simultaneously at least 50 times for each pair, until participants reported that they had learned these new associations. The second session was an imagery training session in which only the visual cues were presented and the participants were encouraged to vividly imagine hearing the corresponding sounds; the real sound was then presented, and the participants were required to adjust their previously imagined sound to the real sound. The third session was an imagery-comparison training session in which one of the visual cues was presented and the participants were asked to imagine hearing the corresponding sound; after a short delay (600 ms) one of the three sounds was then presented and participants were asked to make a same/different comparison between the real sound and the previously imagined one by pressing a button as accurately and quickly as possible (see Figure 1). Following the button press, "Correct" or "Incorrect" feedback was presented to encourage both response accuracy and speed.

After these three training sessions, the participants completed ten blocks of the EEG recording experiment with short breaks between blocks. Each block started with another short familiarization and imagery training session, and then the imagery-comparison task was performed (see Figure 1). Unlike the training session, however, no response feedback was provided during the experiment. Sequential effects of trial-to-trial transitions were counterbalanced within each block. The number of trials requiring "same" and "different" responses was equal. Each sound was both imagined and heard 240 times, leading to a total of 720 trials completed by each participant.

After the ten experimental blocks, the participants completed a brief questionnaire by rating the vividness of their auditory imagery on a 7-point scale (1 = no imagery at all, 7 = very vivid imagery), and reporting whether they had experienced subvocalization (i.e., silent movements of the lips, tongue, or larynx) while imagining hearing sounds.

EEG Recording and Analysis

During each experimental block, EEG data were continuously recorded from 64 cap-mounted Ag/AgCl electrodes arranged according to the 10–20 international placement system (Compumedics Neuroscan, Charlotte, NC) with an on-line reference to the left mastoid and off-line algebraic re-reference to the average of the left and right mastoids. The EEG data were amplified with a bandpass filter of 0.05–100 Hz and digitized at 500 Hz. The vertical and horizontal electrooculogram (VEOG and HEOG) were recorded from two pairs of electrodes: one pair placed 1 cm above and below the left eye, and another pair placed 1 cm lateral from the outer canthi of both eyes. Interelectrode impedances were maintained below 5 kΩ.

The EEG data were processed offline using the Neuroscan 4.3 software. Ocular artifacts were removed using a regression procedure implemented in the Neuroscan software (Semlitsch, Anderer, Schuster, & Presslich, 1986). Data were digitally filtered at 30 Hz lowpass and were epoched into periods of 1000 ms (including 200 ms of pre-stimulus time as a baseline) time-locked to the onset of the visual cue (S1-period) and the real sound during the familiarization session before each block of imagery-comparison task. Epochs containing artifacts exceeding $\pm$ 70 μV were rejected from the analysis. The data were then averaged for each condition.

The present analysis focused on the LPC elicited by the imagined-S1. The LPC does not have a clear peak and was therefore measured as the mean amplitude in the time window of 330–500 ms at the following 18 sites: Fz, FCz, Cz, CPz, Pz, POz, F3, FC3, C3, CP3, P3, PO3, F4, FC4, C4, CP4, P4, and PO4.

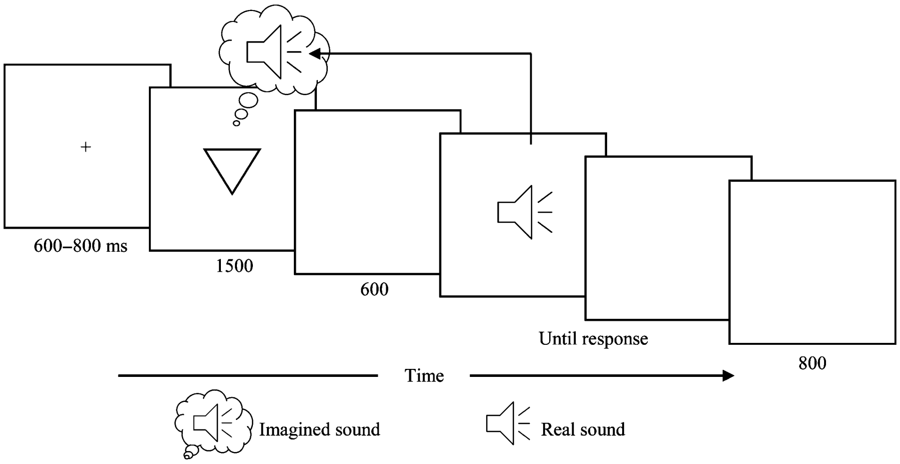

Figure 1. Schematic description of the experimental paradigm. After the fixation, one of the visual cues was presented for 1500 ms and the subjects were required to imagine the corresponding sound; after a 600-ms delay, one of the real sounds was presented for 250 ms (or less if the RT was less than 250 ms) and the subjects' task was to compare the real sound with the preceding imagined sound by pressing a button. The screen then remained blank until a response. The next fixation appeared 800 ms after the response.

These amplitudes were subjected to repeated measures analysis of variance (ANOVA) with factors of pitch (imagining low and high-pitch sound) or loudness (imagining soft and loud sound) $\times$ anterior-posterior scalp location (F, FC, C, CP, P, and PO) $\times$ laterality (left, midline, and right). The data from medium-pitch/loudness sounds were not analyzed due to the interference of both high-pitch/loud and low-pitch/soft sounds on the imagery of medium sound. Behaviorally incorrect trials were also not analyzed. For the direct comparison between the imagery-related LPC and the auditory perception-related N1 within the same experiments, we also analyzed the effect of pitch and loudness of the perceived sound on the N1 peak amplitude at these same 18 sites when the sound was actually presented during the familiarization session before each block of imagery-comparison task. The Greenhouse-Geisser correction was used to compensate for sphericity violations.

Results

Experiment 1

For the amplitude of the LPC elicited by the imagined-S1, the main effect of pitch was not significant, but there was a significant main effect of anterior-posterior electrodes, $F(5,105) = 12.70$, $p < .01$, $\varepsilon = .27$, and there was a marginally significant pitch $\times$ anterior-posterior electrodes interaction, $F(5,105) = 3.85$, $p < .1$, $\varepsilon = .24$. Further analysis revealed that the imagined low-pitch sound elicited a more positive LPC than did the imagined high-

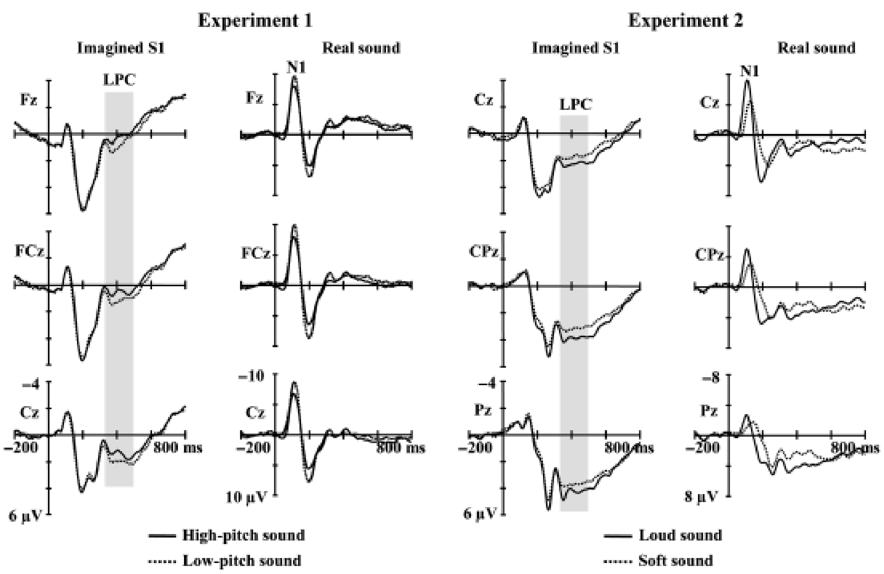

Figure 2. Grand averaged ERPs illustrating pitch effect (Experiment 1, left panel) and loudness effect (Experiment 2, right panel) of imagined sound (S1) and real sound presented in the familiarization session before each block of the main imagery-comparison task. The gray areas highlighted the time windows of LPC (330–500 ms) used for statistical analysis.

pitch sound at fronto-central areas (see Figure 2), F: $F(1,21) = 3.22$, $p < .1$, FC: $F(1,21) = 6.80$, $p < .05$ and C: $F(1,21) = 7.52$, $p < .05$, but not at more posterior areas. Imagining both low-pitch and high-pitch sounds have maximum amplitudes over centro-parietal areas.

For the amplitude of the N1 elicited by the perceived sound, the low-pitch sound elicited a significantly greater N1 than did the high-pitch sound (see Figure 2), $F(1,21) = 12.27$, $p < .01$, and there was a significant main effect of anterior-posterior electrodes, $F(5,105) = 69.09$, $p < .001$, $\varepsilon = .27$. Further analysis revealed that hearing both low-pitch and high-pitch sounds have maximum amplitudes over fronto-central areas.

Experiment 2

For the effect of loudness of the imagined sound on the amplitude of LPC, ANOVA revealed a significant main effect, $F(1,22) = 6.24$, $p < .05$. There was also a significant main effect of anterior-posterior electrodes, $F(5,110) = 13.01$, $p < .001$, $\varepsilon = .28$, and a marginally significant loudness $\times$ anterior-posterior electrodes interaction effect, $F(5,110) = 3.90$, $p < .1$, $\varepsilon = .30$. Further analysis revealed that imagining the loud sound elicited a more positive LPC than did imagining the soft sound at centro-parietal areas (see Figure 2), C: $F(1,22) = 4.96$, $p < .05$, CP: $F(1,22) = 9.98$, $p < .01$, P: $F(1,22) = 10.53$, $p < .01$, and PO: $F(1,22) = 11.21$, $p < .01$, but not at more anterior areas. Imagining both soft and loud sounds have maximum amplitudes over centro-parietal areas.

For the amplitude of the N1 elicited by the perceived sound, the loud sound elicited a significantly greater N1 than did the soft sound (see Figure 2), $F(1,22) = 15.04$, $p < .01$, and there was a significant main effect of anterior-posterior electrodes, $F(5,110) = 35.72$, $p < .001$, $\varepsilon = .26$. Further analysis revealed that hearing both soft and loud sounds have maximum amplitudes over fronto-central electrodes.

Discussion

The present study used the ERP method to investigate whether perceptual attributes of auditory experiences are represented in auditory imagery. The main findings of the present study can be summarized as follows: imagining a low-pitch sound elicited a higher amplitude LPC than did imagining a high-pitch sound, and imagining a loud sound elicited a higher amplitude LPC than did imagining a soft sound. This response pattern parallels that of the effects of sound frequency and intensity on the auditory perception-related N1, which has been well documented in the literature (for a review, see Näätänen & Picton, 1987) and also replicated in the present study.

The similar pattern of amplitude modulation between the imagery-related LPC and the auditory perception-related N1 suggests that auditory imagery may encode pitch and loudness information in a similar manner as auditory perception, thereby implying a functional similarity between auditory imagery and auditory perception. Behavioral studies also have suggested that auditory imagery reflects perceptual attributes of sound such as pitch and loudness (Farah & Smith, 1983; Intons-Peterson, 1980). The results of the present study provide to our knowledge the first neural evidence that auditory imagery may represent perceptual attributes of auditory experiences and provide evidence supporting the depictive theory of mental imagery in the auditory modality.

Previous studies provide insight into the possible cortical mechanisms underlying the amplitude modulation of the auditory perception-related N1 and imagery-related LPC. Studies using various neuroimaging techniques have shown both tonotopic and amplitopic organization of the human auditory cortex (e.g., Bilecen, Seifritz, Scheffler, Henning, & Schulte, 2002; Romani, Williamson, & Kaufman, 1982). Particularly worth mentioning among these is the magnetic equivalent of the electric N1, the N1m. The source of N1m, which is located in the auditory cortex (for a review, see Jacobson, 1994), varies with the frequency and intensity of the experienced sound: lower-frequency or higher-intensity sounds activate more superficial cortical regions (Pantev, Bertrand, Eulitz, Verkindt, Hampson, et al., 1995; Pantev, Hoke, Lehnertz, & Lütkenhöner, 1989; Pantev, Roberts, Elbert, Ross, & Wienbruch, 1996). Therefore, the larger amplitude of the electrical N1 elicited by lower-frequency or higher-intensity sounds might be attributed to the shorter distance between the brain source and the recording electrodes (Jacobson, Lombardi, Gibbens, Ahmad, & Newman, 1992; Wunderlich & Cone-Wesson, 2001). The same logic might apply to the imagery-related LPC in the present study: the amplitude modulation of the scalp recorded imagery-related LPC might be attributed to the tonotopicity and amplitopicity in the auditory cortex during auditory imagery. Previous brain imaging studies have revealed that the auditory cortex, especially the primary auditory cortex, is activated by auditory imagery (Aleman et al., 2005; Kraemer et al., 2005). To confirm this assumption of tonotopicity and amplitopicity of the auditory cortex when imagining auditory stimuli, further studies with high spatial resolution neuroimaging will be needed.

The ERPs elicited by the perceived-S2, in particular the effects of the discrepancy between the imagined and perceived sounds on the N2 component and the behavioral performance of the same/different comparison, have been reported in our previous paper (Wu et al., 2010). The different comparison pairings between presented sounds (S2) and imagined sounds (S1) led to three levels of discrepancy (no, small, and large discrepancy) defined by the degree of physical separation. The results can be summarized as follows: an N2 component with latency of approximately 220 ms was reliably elicited when the heard S2 was different from the imagined S1, suggesting that the N2 ERP, typically elicited in an S1–S2 matching paradigm and considered to reflect mismatch processing, can still be elicited when the S1 was imagined instead of perceived. Furthermore, the amplitude of the N2 increased with the degree of discrepancy, and accuracies were lower and reaction times were longer for the small discrepancy than large discrepancy conditions. These behavioral results replicate the fourth task by Intons-Peterson (1980) in which the discrimination times increased with decreasing distances between loudness of two imagined sounds. The effect of the degree of discrepancy on both the N2 amplitude and the behavioral performance provide further evidence that auditory imagery includes perceptual information of auditory experiences (i.e., pitch in Experiment 1 and loudness in Experiment 2), supporting the assumption of functional similarity between auditory imagery and perception.

In summary, auditory imagery encodes pitch and loudness information as evidenced by the similar pattern of amplitude modulation between the imagery-related LPC and the auditory perception-related N1. The previously reported effects of discrepancy between the real sound and the previously imagined sound on both behavioral performance and N2 component amplitude (Wu et al., 2010) provide further evidence supporting this assumption.

REFERENCES

Aleman, A., Formisano, E., Koppenhagen, H., Hagoort, P., de Haan, E. H. F., & Kahn, R. S. (2005). The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words. *Cerebral Cortex*, *15*, 221–228.

Bilecen, D., Seifritz, E., Scheffler, K., Henning, J., & Schulte, A. C. (2002). Amplitopicity of the human auditory cortex: An fMRI study. *NeuroImage*, *17*, 710–718.

Farah, M. J., Peronnet, F., Weisberg, L. L., & Monheit, M. (1989). Brain activity underlying mental imagery: Event-related potentials during mental image generation. *Journal of Cognitive Neuroscience*, *1*, 302–316.

Farah, M. J., & Smith, A. F. (1983). Perceptual interference and facilitation with auditory imagery. *Perception and Psychophysics*, *33*, 475–478.

Hubbard, T. L. (2010). Auditory imagery: Empirical findings. *Psychological Bulletin*, *136*, 302–329.

Intons-Peterson, M. J. (1980). The role of loudness in auditory imagery. *Memory and Cognition*, *8*, 385–393.

Intons-Peterson, M. J. (1992). Components of auditory imagery. In D. Reisberg (Ed.), *Auditory imagery* (pp. 45–71). Hillsdale, NJ: Erlbaum.

Jacobson, G. P. (1994). Magnetoencephalographic studies of auditory system function. *Journal of Clinical Neurophysiology*, *11*, 343–364.

Jacobson, G. P., Lombardi, D. M., Gibbens, N. D., Ahmad, B. K., & Newman, C. W. (1992). The effects of stimulus frequency and recording site on the amplitude and latency of multichannel cortical auditory evoked potential (CAEP) component N1. *Ear and Hearing*, *13*, 300–306.

Klein, I., Dubois, J., Mangin, J. F., Kherif, F., Flandin, G., Poline, J. B., et al. (2004). Retinotopic organization of visual mental images as revealed by functional magnetic resonance imaging. *Cognitive Brain Research*, *22*, 26–31.

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

Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin*, *129*, 723–746.

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. New York: Oxford University Press.

Kosslyn, S. M., Thompson, W. L., Kim, I. J., & Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature*, *378*, 496–498.

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

Meyer, M., Elmer, S., Baumann, S., & Jancke, L. (2007). Short-term plasticity in the auditory system: Differential neural responses to perception and imagery of speech and music. *Restorative Neurology and Neuroscience*, *25*, 411–431.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*, 375–425.

Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., & Elbert, T. (1995). Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalography and Clinical Neurophysiology*, *94*, 26–40.

Pantev, C., Hoke, M., Lehnertz, K., & Lütkenhöner, B. (1989). Neuromagnetic evidence of an amplitopic organization of the human auditory cortex. *Electroencephalography and Clinical Neurophysiology*, *72*, 225–231.

Pantev, C., Roberts, L. E., Elbert, T., Ross, B., & Wienbruch, C. (1996). Tonotopic organization of the sources of human auditory steady-state responses. *Hearing Research*, *101*, 62–74.

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, *25*, 157–182.

Romani, G. L., Williamson, S. J., & Kaufman, L. (1982). Tonotopic organization of the human auditory cortex. *Science*, *216*, 1339–1340.

Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology*, *23*, 695–703.

Slotnick, S. D., Thompson, W. L., & Kosslyn, S. M. (2005). Visual mental imagery induces retinotopically organized activation of early visual areas. *Cerebral Cortex*, *15*, 1570–1583.

Wu, J., Mai, X., Chan, C. C., Zheng, Y., & Luo, Y. (2006). Event-related potentials during mental imagery of animal sounds. *Psychophysiology*, *43*, 592–597.

Wu, J., Mai, X., Yu, Z., Qin, S., & Luo, Y. (2010). Effects of discrepancy between imagined and perceived sounds on the N2 component of the event-related potential. *Psychophysiology*, *47*, 289–298.

Wunderlich, J. L., & Cone-Wesson, B. K. (2001). Effects of stimulus frequency and complexity on the mismatch negativity and other components of the cortical auditory-evoked potential. *Journal of the Acoustical Society of America*, *109*, 1526–1537.

(Received November 6, 2009; Accepted April 22, 2010)