

BRIEF REPORT

Event-related potentials during mental imagery of animal sounds

JIANHUI WU,^{a,b} XIAOQIN MAI,^a CHETWYN C.H. CHAN,^c YAQIN ZHENG,^d AND YUEJIA LUO^{a,d}

^aKey Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

^bGraduate School of the Chinese Academy of Sciences, Beijing, China

^cApplied Cognitive Neuroscience Laboratory, Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hong Kong, China

^dState Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China

Abstract

To investigate the neural correlates of imagined animal sounds, event-related potentials (ERPs) were recorded while subjects were presented with (1) animal pictures without any imagery instruction (control) or (2) animal pictures with instructions to imagine the corresponding sounds (imagery). The results revealed imagery effects starting with an enhancement of the P2, possibly indexing the top-down allocation of attention to the imagery task, and continuing into a more positive-going deflection in the time window of 350–600 ms poststimulus, probably reflecting the formation of auditory imagery. A centro-parietally distributed late positive complex (LPC) was identified in the difference waveform (imagery minus control) and might reflect two subprocesses of imagery formation: sound retrieval from stored information and representation in working memory.

Descriptors: Auditory imagery, Event-related potentials, Late positive complex, Memory, Subvocalization

“Mental imagery occurs when perceptual information is accessed from memory, giving rising to the experience of seeing with the mind’s eye, hearing with the mind’s ear and so on” (Kosslyn, Ganis, & Thompson, 2001). Over the past 30 years, there has been great interest in the study of mental imagery. One of the main findings from neuroimaging and neuropsychological data is that imagery shares some brain structures with perception in the same modality (for a review, see Kosslyn et al., 2001). Another finding is that imagery tasks activate brain regions related to memory (for a review, see Mellet, Petit, Mazoyer, Denis, & Tzourio, 1998). Most of the research into the neural substrates of mental imagery has focused on the visual modality. In recent years, however, there has been an increasing interest in the study of auditory imagery (e.g., Aleman et al., 2005; Bunzeck, Wuestenberg, Lutz, Heinze, & Jancke, 2005; Halpern & Zatorre, 1999; Halpern, Zatorre, Bouffard, & Johnson, 2004; Jancke & Shah, 2004; Kraemer, Macrae, Green, & Kelley, 2005; Schur-

mann, Raij, Fujiki, & Hari, 2002). One notable aspect of the previous auditory imagery studies is that most of them have concentrated on music or language. Nevertheless, auditory imagery includes not only music and language but also natural sounds in the environment, for example, animal sounds. Furthermore, these studies have generally focused on the anatomical brain areas but provided limited temporal information related to imagery processing. The event-related potential (ERP), in spite of its higher temporal resolution, has received considerably less attention in the imagery field. Several ERP studies of visual imagery have shown that the imagery effect starts about 300 ms after stimulus onset (Farah, Peronnet, Weisberg, & Monheit, 1989; Ferlazzo, Conte, & Gentilomo, 1993; Nittono, Suehiro, & Hori, 2002). However, little ERP research has been conducted on auditory imagery. To our knowledge, there is only one exception, a study by Fallgatter, Mueller, and Strik (1997). This previous study, focusing on the different brain topographies underlying mental imagery in three sensory modalities (visual, auditory, and tactile), conducted spatiotemporal segmentation analysis only within a limited time range chosen by maximal global field power differences between the three modalities. Thus the time course of imagery processing within auditory modality has not yet been addressed.

The aim of our study was to investigate the time course of auditory imagery processing of animal sounds using ERPs. Almost everyone is familiar with these animal sounds, which makes it relatively easy for subjects to imagine them. During the experiment, subjects were asked in the visual control condition to simply look at the animal picture and in the auditory imagery

This work was supported by the NSF China (30325026, 30670698) and the Ministry of Education (106025). We thank Chao Liu for constructive advice on experimental design and Zulin Yu for assistance with sound processing. We also thank Nina K. Simms and Ellen E. Hamilton for their editorial assistance and we especially thank Prof. Jinghan Wei and Mr. Shaozheng Qin for their valuable discussions throughout the project.

Address reprint requests 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.

condition to look at the picture and imagine the corresponding sound. The difference waveform between the two conditions should represent the neural correlates of auditory imagery. A similar paradigm has been used in previous studies of mental imagery in both auditory and visual modalities (e.g., Aleman et al., 2005; Bunzeck et al., 2005; Farah et al., 1989; Jancke & Shah, 2004; Rajj, 1999; Schurmann et al., 2002; Yamamoto & Mukai, 1998). To ensure that subjects performed the required task in all conditions, we interspersed a monitoring task requiring a same-different comparison over two consecutive trials. To evaluate comparatively the behavioral performance related to imagined sounds and thus to better decide whether subjects had imagined sounds as required, we also designed an auditory perception condition during which animal pictures and sounds were presented simultaneously.

Methods

Participants

Twenty-two undergraduate students from China Agriculture University (mean age 20.8 ± 1.8 years, 11 men, all right-handed) participated in this study as paid volunteers. None had a history of neurological or psychiatric disorders. All reported normal hearing and normal or corrected-to-normal vision. The data from an additional two participants were discarded due to excessive movement artifacts.

Stimuli

Nine kinds of animals (cow, duck, dog, sheep, cat, pig, horse, rooster, and tiger) were chosen with the constraint of their high familiarity and high auditory imageability, which were evaluated by 20 subjects who did not participate in the present ERP study. Each animal consisted of three black and white pictures and one characteristic vocalization, as selected by another 10 subjects. Pictures were presented on a computer monitor placed 75 cm in front of the participants' eyes. These images subtended a visual angle of approximately 8.5° horizontally and 6.5° vertically and each was presented for a duration of 2000 ms. With audio editing software (Adobe Audition 1.0), the duration of each sound was modified to 1000 ms and the intensity was adjusted to a comfortable listening level of ~ 70 dB SPL. Sounds were delivered binaurally through headphones by the Stim interface system (Neuroscan Labs, Sterling, VA).

Procedure

During EEG recordings, subjects were seated in a relaxed position on a comfortable chair in a dimly lit and electrically isolated room. Three conditions were used in this experiment: control, imagery, and perception. Each condition was presented in a separate session, with each session taking approximately 6.5 min. Figure 1 shows the control and imagery sessions. For every subject, the nine animals were divided pseudorandomly into the three conditions to ensure that each animal had the same chance to be selected in each condition. Each picture was presented 12 times besides those presented in the response trials (i.e., these trials during which a same-different comparison task was required as described below), which resulted in 108 nonresponse trials per condition. Interstimulus intervals were randomized between 800 and 1200 ms, during which a central fixation was presented.

The experiment began with the session for the visual control condition (control) to avoid the influence of any sound infor-

mation given before this session. In this session, animal pictures were presented without sounds and the subjects were instructed to follow the pictures attentively. A question mark would appear randomly on 20% of the trials. Closely after the question mark, an animal picture would be presented and the participant's task was to determine as quickly and accurately as possible whether the picture was the same as or different from the last picture preceding the question mark and to indicate their answer by pressing a button. The other two conditions, auditory perception (perception) and auditory imagery (imagery), were counterbalanced in the following two sessions. In the perception condition, the animal pictures and corresponding sounds were simultaneously presented and subjects were required to both look at the picture and listen to the sound. The question mark would appear as in the control session and after the question mark, a picture or sound would be presented. Subjects were required to compare the picture or sound with its preceding counterpart. During the imagery session, pictures were presented without sounds and subjects were required to look at the picture and vividly imagine the corresponding sound. The same-different comparison task was identical to that in the perception session, but in the imagery condition, subjects compared the heard sound with the previously imagined one. We supposed that subjects would perform well on the same-different comparison task only if they had imagined the sound in the prior trial, due to the consideration that the previous vivid imagining should lead to a direct sound comparison during the response trials. Specifically, the accuracy rate and reaction time (RT) of sound comparison in the imagery condition would be similar to those in the perception condition if subjects had imagined the sound well.

Before each session, subjects were given the opportunity to practice the required tasks. To facilitate imagery generation of animal sound during the imagery session, the subjects were first presented with these animal pictures and corresponding sounds repeatedly. This prior familiarization period was also conducted before the other two sessions for the sake of consistency between conditions.

After the ERP recordings, subjects were given a questionnaire to report whether they had experienced auditory imagery in the control and imagery conditions, to rate the imagery vividness on a 9-point scale (1 = *no imagery at all*, 9 = *very vivid imagery*), and to report whether they had experienced subvocalization (i.e., silent movements of their lips, tongue, or larynx) while imagining animal sounds.

EEG Recording and Analysis

During all three conditions, electroencephalogram (EEG) was recorded from 64 scalp sites using Ag/AgCl electrodes mounted in an elastic cap (Neuroscan Inc.), with reference to linked mastoids. The vertical electrooculogram (VEOG) and horizontal electrooculogram (HEOG) were recorded with two pairs of electrodes, one placed above and below the left eye, and another 10 mm from the outer canthi of both eyes. All interelectrode impedance was maintained < 5 k Ω . Signals were amplified with 0.05–100 Hz bandpass filter and digitized at 500 Hz.

The EEG data from the imagery and control conditions were digitally filtered with 30 Hz low-pass and were epoched into periods of 1200 ms (including a 200-ms prestimulus baseline). Ocular artifacts were removed from the EEG signal using a regression procedure implemented in the Neuroscan software (Semlitsch, Anderer, Schuster, & Presslich, 1986). Trials with various artifacts were rejected, with a criterion of ± 75 μ V. The

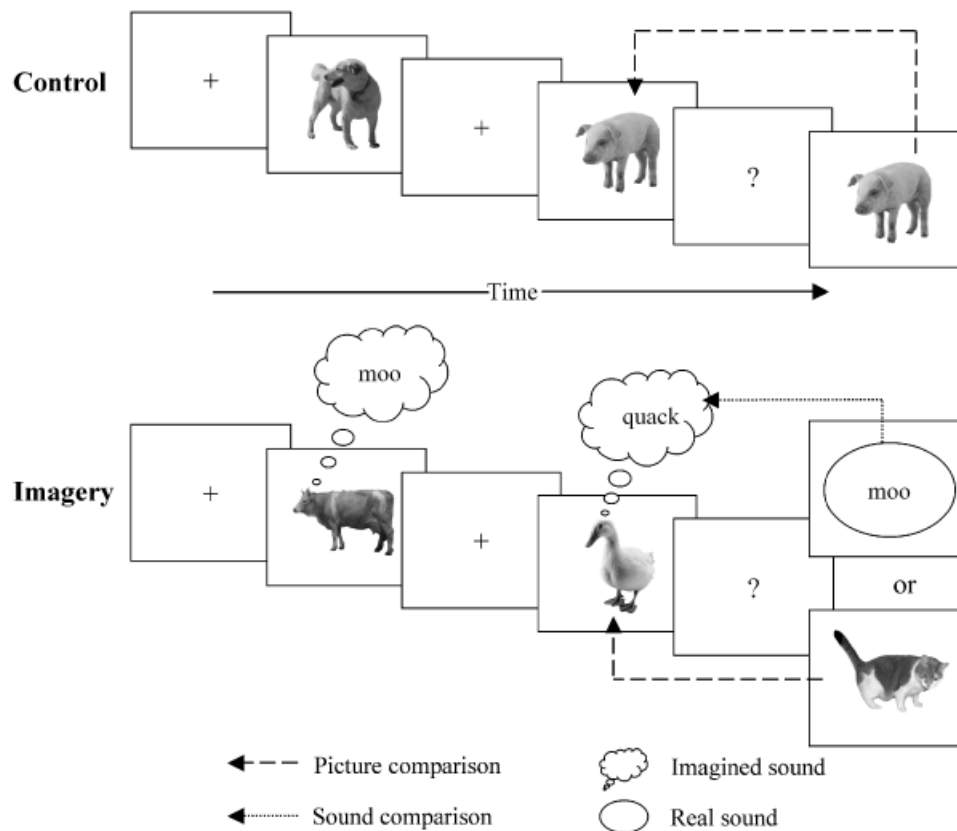


Figure 1. Schematic description of the experimental paradigm. The question mark and the following response trial in the control and imagery sessions were presented randomly on 20% of the trials and the subjects' task was to compare the picture or sound presented in the response trials with its preceding counterpart.

ERPs were then averaged separately for both experimental conditions. Only nonresponse trials were included in the averaging procedure. In addition, based on the logic that the same-different comparison task in response trials might interfere with the imagery performance of next trial, we excluded the next trial after each response trial.

For the early components (P1, N1, and P2), the peak amplitudes and latencies were measured; for the late component, mean amplitudes were measured in the time window of 350–600 ms after stimulus onset. The P1 component was mainly distributed over posterior electrodes and was analyzed accordingly at the following nine sites: Pz, POz, Oz, P3, PO3, O3, P4, PO4, and O4, and the repeated measures ANOVA factors were Condition (imagery vs. control) and Electrode Site. The other components, which are more broadly distributed, were analyzed at the following 18 sites: Fz, FCz, Cz, CPz, Pz, POz, F3, FC3, C3, CP3, P3, PO3, F4, FC4, C4, CP4, P4, and PO4, and the repeated measures ANOVA factors included Condition (imagery and control) \times Laterality (left, midline, and right) \times Anterior-Posterior (F, FC, C, CP, P, and PO). The Greenhouse–Geisser correction was used to compensate for sphericity violations. Post hoc analyses were conducted to explore the interaction effects.

Results

Behavior and postexperimental questionnaire

Mean response error was 3% and mean RT was 862 ms for all response trials in all conditions. Both error rate and RT did not reach significant difference between sound comparison in the

imagery and perception condition (3% vs. 1.5%, $t[21] = 1.82$, $p > .05$ and 938 ms vs. 953 ms, $t[21] = 0.54$, $p > .05$, respectively). Postexperimental questionnaire revealed that no subjects had experienced imagery of animal sounds while viewing pictures in the control condition. The vividness rating showed that all subjects had experienced vivid auditory imagery in the imagery condition (7.26 ± 0.92). Nineteen out of the 22 subjects reported that they had experienced subvocalization in the imagery condition.

ERPs

As shown in Figure 2, the P1, N1, and P2 were elicited by both imagery and control conditions. The amplitude was significantly greater in the imagery condition than in the control for the P2, $F(1,21) = 19.48$, $p < .001$, but not for the P1 and N1. The Condition \times Anterior-Posterior interaction for P2 amplitude reached marginal significance, $F(5,105) = 2.78$, $p < .10$, $\epsilon = .454$, but the Condition \times Laterality interaction did not. Post hoc contrasts revealed that the effect of Condition for P2 amplitude was maximal at front-central locations. The difference in latencies between two conditions did not reach the level of significance for all three components.

Imagery also elicited a more positive deflection than control in the ERP waveform between 350 and 600 ms approximately, $F(1,21) = 7.20$, $p < .05$. The Condition \times Anterior-Posterior interaction reached significance, $F(5,105) = 7.27$, $p < .01$, $\epsilon = .365$, but the Condition \times Laterality interaction did not. Post hoc tests revealed that the effect of condition was significant at central, centro-parietal, parietal, and parietal-occipital loca-

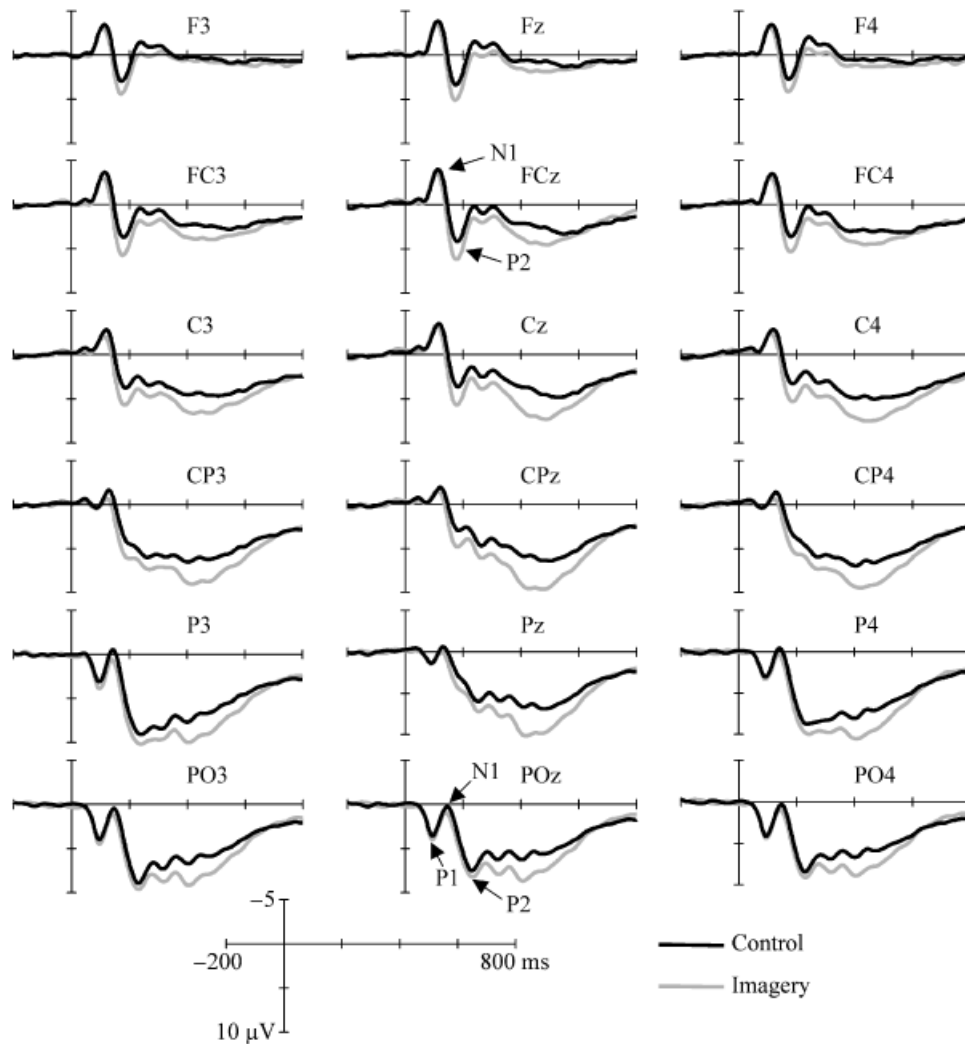


Figure 2. Grand averaged ERPs elicited by imagery and control at 18 electrode sites.

tions, but not significant at frontal and front-central locations. This distribution is also illustrated in Figure 3, where the difference waveform was obtained by subtracting the ERPs elicited by the control from that elicited by the imagery, and a centro-parietally distributed late positive complex (LPC) was identified with a peak latency of about 450 ms at CPz.

Discussion

The present study examined the ERP correlates of auditory imagery of animal sounds. The behavioral and questionnaire results suggest that the subjects focused their attention on the presented stimuli and successfully executed the required auditory imagery task during the imagery session. Because there was no significant difference in the P1 and N1 between imagery and control, it might indicate that the visual processing during these stages is involved equally in both conditions.

The first ERP difference was that the imagery condition showed increased P2 amplitude compared to the control. Although generally considered to be an exogenous component, the P2 component is also related to attention allocation, depending on, for example, subjects' state (e.g., Bernal et al., 2000), stimulus

salience (e.g., Huang & Luo, 2006), or task information (e.g., Luck & Hillyard, 1994). In the present study, the task demands were different between the two conditions. In the control condition, subjects only needed to attend to the picture, whereas in the imagery condition, subjects had to both look at the picture and imagine the sound. We thus believe that the additional mental imagery task might lead to the top-down allocation of attention, which should contribute to the larger amplitude of the P2 component.

The second imagery effect was the greater positive deflection in the time window of about 350–600 ms after stimulus presen-

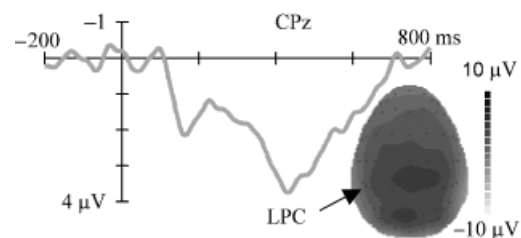


Figure 3. The difference waveform (imagery minus control) at CPz and the scalp topography at the peak latency of the LPC.

tation. The centro-parietally distributed LPC was identified in the difference waveform between the two conditions. The ERP component during this late stage is generally considered to be an endogenous one that varies as a function of internal mental process. As Johnson (1986) suggested, one factor influencing LPC (i.e., P3) amplitude is the amount of information extracted from the stimulus by the subject. In this study, the increased amount of information during the imagery compared to the control (i.e., the imagined sound) should contribute to the LPC. The starting time of the imagery effect (i.e., 350 ms) in the present study is similar to that in visual imagery studies with both ERP (Farah et al., 1989; Ferlazzo et al., 1993; Nittono et al., 2002) and MEG (Raij, 1999) methods. We propose that in the present study, 350 ms might be the time point when auditory imagery began to form after the stimulus evoking imagery had been visually perceived and top-down attention had been allocated, and the LPC might reflect the auditory imagery formation.

Before further analyses on the cognitive processes underling the LPC, we consider another related phenomenon, which is that most of the subjects reported that they had experienced subvocalization while imagining animal sounds. Smith, Reisberg, and Wilson (1992) and Smith, Wilson, and Reisberg (1995) reported the functional involvement of subvocalization during auditory imagery of both linguistic and musical materials. The current self-report results suggest that subvocalization is also concurrent with the auditory imagery of animal sounds. According to Smith et al. (1992, 1995) assumption, subvocalization in auditory imagery should be similar to the subvocal rehearsal of Baddeley's (1986) working memory model.

Baddeley and Logie (1992) have proposed that the phonological loop subsystem of working memory is the basis for auditory imagery. Brain mapping research has demonstrated that brain areas involved in the phonological loop are also activated by auditory imagery (Aleman et al., 2005). Baddeley and Andrade (2000) further suggest the involvement of both working memory and long-term memory in mental imagery. Halpern and Zatorre (1999) have even successfully separated memory retrieval processes from auditory imagery processes in their PET research. According to Kosslyn's (1994) theory of mental imagery, visual imagery, being based on stored information, is

activated and maintained in the visual buffer. Based on these analyses, we assume that at least two main subprocesses are involved in the formation of auditory imagery in the present study: animal sound retrieval from stored information and representation in the phonological loop component of working memory. Along with the converging experimental evidence that the LPC (i.e., P3) is related to both working memory and cued recall (Allan & Rugg, 1997; Donchin & Coles, 1988; Johnson, Kreiter, Zhu, & Russo, 1998; Kok, 2001; McCarthy, Luby, Gore, & Goldman-Rakic, 1997), we believe that the proposed two subprocesses might be involved in the elicitation of the LPC in the present study.

We propose above that it is the LPC that reflects auditory imagery formation; however, it could be argued that imagery formation should begin at the P2 and extend into the LPC range. The self-report results demonstrated here that no subject had experienced imagery in the control condition, and all subjects had experienced vivid imagery as required in the imagery condition, showing that the imagery formation in this study is a controlled, not automatic, process while viewing pictures. It is more likely that the controlled processing (i.e., imagery formation in the present study) should occur at later components such as the P300 (e.g., Rosler, Clausen, & Sojka, 1986).

In summary, our findings indicate that auditory imagery probably begins with attention allocation to the imagery task, which is denoted by an increase in P2 amplitude, and continues into imagery formation, which is associated with the greater positive-going deflection within the 350–600-ms time window. Further study would be necessary to investigate the potentially different cognitive and neural dynamics between automatic and controlled imagery. In addition, a LPC was identified in the difference waveform between imagery and control and might reflect two subprocesses of imagery formation: retrieval of sound images from stored information and their representations in the working memory. Nevertheless, the design of this study did not enable us to answer the following questions: Can the neural correlates of these subprocesses be temporally separated? Do these subprocesses occur sequentially, simultaneously or interactively? Future research will need to address these issues.

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.
- Allan, K., & Rugg, M. D. (1997). An event-related potential study of explicit memory on tests of cued recall and recognition. *Neuropsychologia*, *35*, 387–397.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.
- Baddeley, A. D., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General*, *129*, 126–145.
- Baddeley, A. D., & Logie, R. H. (1992). Auditory imagery and working memory. In D. Reisberg (Ed.), *Auditory imagery* (pp. 179–197). Hillsdale, NJ: Erlbaum.
- Bernal, J., Harmony, T., Rodriguez, M., Reyes, A., Yanez, G., Fernandez, T., et al. (2000). Auditory event-related potentials in poor readers. *International Journal of Psychophysiology*, *36*, 11–23.
- Bunzeck, N., Wuestenber, T., Lutz, K., Heinze, H. J., & Jancke, L. (2005). Scanning silence: Mental imagery of complex sounds. *NeuroImage*, *26*, 1119–1127.
- Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral Brain Science*, *11*, 357–374.
- Fallgatter, A. J., Mueller, T. J., & Strik, W. K. (1997). Neurophysiological correlates of mental imagery in different sensory modalities. *International Journal of Psychophysiology*, *25*, 145–153.
- 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.
- Ferlazzo, F., Conte, S., & Gentilomo, A. (1993). Event-related potentials and recognition memory: The effects of word imagery value. *International Journal of Psychophysiology*, *15*, 115–122.
- 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*, *9*, 697–704.
- Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia*, *42*, 1281–1292.
- Huang, Y. X., & Luo, Y. J. (2006). Temporal course of emotional negativity bias: An ERP study. *Neuroscience Letters*, *398*, 91–96.
- Jancke, L., & Shah, N. J. (2004). Hearing syllables by seeing visual stimuli. *European Journal of Neuroscience*, *19*, 2603–2608.
- Johnson, R., Jr. (1986). A triarchic model of P300 amplitude. *Psychophysiology*, *23*, 367–384.

- Johnson, R., Jr., Kreiter, K., Zhu, J., & Russo, B. (1998). A spatio-temporal comparison of semantic and episodic cued recall and recognition using event-related brain potentials. *Brain Research. Cognitive Brain Research*, 7, 119–136.
- Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, 38, 557–577.
- Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.
- Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews. Neuroscience*, 2, 635–642.
- Kraemer, D. J., Macrae, C. N., Green, A. E., & Kelley, W. M. (2005). Musical imagery: Sound of silence activates auditory cortex. *Nature*, 434, 158.
- Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, 31, 291–308.
- McCarthy, G., Luby, M., Gore, J., & Goldman-Rakic, P. (1997). Infrequent events transiently activate human prefrontal and parietal cortex as measured by functional MRI. *Journal of Neurophysiology*, 77, 1630–1634.
- Mellet, E., Petit, L., Mazoyer, B., Denis, M., & Tzourio, N. (1998). Reopening the mental imagery debate: Lessons from functional anatomy. *NeuroImage*, 8, 129–139.
- Nittono, H., Suehiro, M., & Hori, T. (2002). Word imageability and N400 in an incidental memory paradigm. *International Journal of Psychophysiology*, 44, 219–229.
- Raij, T. (1999). Patterns of brain activity during visual imagery of letters. *Journal of Cognitive Neuroscience*, 11, 282–299.
- Rosler, F., Clausen, G., & Sojka, B. (1986). The double-priming paradigm: A tool for analyzing the functional significance of endogenous event-related brain potentials. *Biological Psychology*, 22, 239–268.
- Schurmann, M., Raij, T., Fujiki, N., & Hari, R. (2002). Mind's ear in a musician: Where and when in the brain. *NeuroImage*, 16, 434–440.
- 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.
- Smith, J. D., Reisberg, D., & Wilson, M. (1992). Subvocalization and auditory imagery: Interactions between the inner ear and inner voice. In D. Reisberg (Ed.), *Auditory imagery* (pp. 95–119). Hillsdale, NJ: Erlbaum.
- Smith, J. D., Wilson, M., & Reisberg, D. (1995). The role of subvocalization in auditory imagery. *Neuropsychologia*, 33, 1433–1454.
- Yamamoto, S., & Mukai, H. (1998). Event-related potentials during mental imagery. *NeuroReport*, 9, 3359–3362.

(RECEIVED May 10, 2006; ACCEPTED August 12, 2006)