Mental imagery of face enhances face-sensitive event-related potentials to ambiguous visual stimuli

Lingxi Lu[a, b], Changxin Zhang[c], Liang Li[a, b, *]

[a] School of Psychological and Cognitive Sciences, Beijing Key Laboratory of Behavior and Mental Health, Key Laboratory on Machine Perception (Ministry of Education), Peking University, Beijing 100080, China
[b] Beijing Institute for Brain Disorders, Beijing 100069, China
[c] Faculty of Education, East China Normal University, Shanghai 200062, China

ARTICLE INFO

ABSTRACT

Visual mental imagery forms mental representations of visual objects when correspondent stimuli are absent and shares some characters with visual perception. Both the vertex-positive-potential (VPP) and N170 components of event-related potentials (ERPs) to visual stimuli have a remarkable preference to faces. This study investigated whether visual mental imagery modulates the face-sensitive VPP and/or N170 components. The results showed that with significantly larger amplitudes under the face-imagery condition than the house-imagery condition, the VPP and P2 responses, but not the N170 component, were elicited by phase-randomized ambiguous stimuli. Thus, the brain substrates underlying VPP are not completely identical to those underlying N170, and the VPP/ P2 manifestation of the category selectivity in imagery probably reflects an integration of top-down mental imagery signals (from the prefrontal cortex) and bottom-up perception signals (from the early visual cortex) in the occipito-temporal cortex where VPP and P2 originate.

1. Introduction

The top-down visual processing plays an important role in perceiving visual objects, particularly when optical inputs are weak, ambiguous, or absent, leading to that an observer can "see" an object in an illusory way even when the visibility of the object is low or absent (e.g. see a face in noise). Visual mental imagery is a typical type of top-down processing that forms mental representations of objects when the corresponding visual stimulation is absent (Gosselin & Schyns, 2003; Hansen et al., 2010; Nestor, Vettel, & Tarr, 2013; Smith, Gosselin, & Schyns, 2012; for a review see Ganis & Schendan, 2011). Currently, the mechanism underlying visual mental imagery is largely unknown.

It has been suggested that imagining visual objects in the mind may activate corresponding brain areas for perception of visual objects (O'Craven & Kanwisher, 2000; Klein et al., 2004). For example, some functional magnetic resonance imaging (fMRI) studies have shown that imagery of illusory faces is associated with activation of multiple cortical regions of the visual system, including the fusiform face area (FFA) and the occipital face area (OFA) (Li et al., 2010; Nestor et al., 2013; Righart et al., 2009), which are also the brain areas specific for face perception (Kanwisher, McDermott, & Chun, 1997; Haxby, Hoffman, & Gobbini, 2000). Extended brain regions for face imagery may also involve the orbitofrontal cortex (Li et al., 2010), inferior frontal gyri (Zhang et al., 2008), and a distributed network between frontal and occipito-temporal areas (Liu et al., 2014).

Among these brain structures, activation of the FFA during mental imagery of faces appears to be the most consistent across different studies using different experimental

---

* Corresponding author at: School of Psychological and Cognitive Sciences, Peking University Beijing 100080, China.
   Email address: liangli@pku.edu.cn, liangli2@hotmail.com (L. Li).

paradigms, supporting the view that visual mental imagery and perception share some common neural substrates. For example, even when the visual sensory input of a face image is absent, imagery can induce marked activation of the FFA (e.g., observers are instructed to imagine a face with their eyes closed), though the imagery-induced activation is weaker than that induced by a face photograph (O'Craven & Kanwisher, 2000). Also, in both healthy individuals (Li et al., 2010; Nestor et al., 2013) and patients with prosopagnosia (Righart et al., 2010), when the visual input is random noise, the FFA becomes activated as long as an illusory face is reported. The Hadjikhani, Kveraga, Naik and Ahlfors magnetoencephalographic (MEG) study (2009) has further shown that when the visual stimulus containing some schematic clues of a face (e.g., the pattern of one mouth and two eyes) are incidentally perceived as a face, the M165 response occurs both with a similar amplitude to that elicited by a real face stimulus and with an activation source in the ventral FFA. All these reports suggest that the FFA is an interface that is involved in functional integration between the top-down imagery process and the bottom-up sensory process. It is important to establish a neurophysiological model of the top-down modulating effect of visual mental imagery of faces.

Numerous electro-encephalographic (EEG) studies have shown that both the vertex positive potential (VPP) component and the N170 component of event-related potentials (ERPs) are markedly face sensitive: both VPP and N170 are enhanced when eliciting stimuli are face-like objects compared to non-face objects (e.g. cars, animals, leaves, flowers, mushrooms, tools, shoes, road signs or words) (for VPP, see Bötzel, Schulze, & Stodieck, 1995; Itier & Taylor, 2004; Jeffreys, 1996; Jeffreys & Tukmachi, 1992; Proverbio & Galli, 2016; Rossion & Jacques, 2008; Rossion, Joyce, Cottrell, & Tarr, 2003; for N170, see Bentin, Allison, Puce, Perez, & McCarthy, 1996; De Haan, Pascalis, & Johnson, 2002; Eimer, 2000; Itier & Taylor, 2004; Proverbio & Galli, 2016; Rossion & Jacques, 2008; Rossion et al., 2000 Taylor, McCarthy, Saliba, & Degiovanni, 1999 ; for a recent review see Rossion, 2014). It is of interest to know whether these two face-sensitive ERP components (N170, VPP) are sensitive to face mental imagery when any face features in eliciting stimuli are completely eliminated.

The face-responsive N170 component mainly occurs at occipito-temporal recording sites on the scalp, with the amplitude peak near 170 ms following the onset of a face stimulus and manifests the early stage of processing face structural information (Bentin et al., 1996; Rossion et al., 2000). Interestingly, N170 can be elicited by non-face noise images either during the maintenance of the working memory of a real face image (Sreenivasan, Katz, & Jha, 2007) or when a noise image is treated as a human face with a gender (Wild & Busey, 2004). Moreover, N170 can even be elicited by simple schematic or line drawings interpreted as faces or eyes (Bentin & Golland, 2002; Bentin et al., 2002). However, it is not clear whether top-down imagery can still affect N170 when any bottom-up face-signal inputs are substantially limited or even completely eliminated.

The VPP emerges at fronto-central recording sites on the scalp with the peak amplitude between 160 and 200 ms following the onset of a face stimulus and may also represent the stage of structural encoding of a face, because it becomes delayed or sometimes attenuated if the face stimulus is disrupted by an inversion, scramble, or masking treatment (George, Evans, Fiori, Davidoff, & Renault, 1996; Jeffreys, 1989; Joyce & Rossion, 2005; Jemel et al., 2003). It is not clear whether VPP can be used for studying whether a preference to face also occurs in mental imagery.

Some previous studies have suggested that VPP and N170 reflect identical brain processes in face perception (Itier & Taylor, 2002; Joyce & Rossion, 2005; Jemel et al., 2003; Rossion & Jacques, 2008; Rossion, Campanella et al., 1999; Rossion, Delvenne et al., 1999; Rossion et al., 2003). Particularly, the changing pattern of amplitude of VPP and that of N170 are highly correlated across different reference sites (Joyce & Rossion, 2005), and their latencies also change synchronously either with addition of noise to face pictures (Jemel et al., 2003) or with inversion or contrast reversal of face pictures (Itier & Taylor, 2002). Some studies have also suggested that both VPP and N170 would be originated from both fusiform gyrus regions and lateral inferior occipital cortex (Rossion, Campanella et al., 1999; Rossion et al., 2003). However, several lines of studies have also suggested that VPP and N170 are based on different brain mechanisms (Itier & Taylor, 2004 Bötzel et al., 1995; George et al., 1996; Taylor et al., 1999; Saavedra et al., 2012): (1) the VPP generator involves a larger network including both inferio-temporal cortex and the superior temporal sulcus, but the N170 generator is confined to the parahippocampal place area (PPA) and FFA (George et al., 1996); (2) the hippocampus may be another origin of VPP, indicating that VPP is also associated with memory-related process (Bötzel et al., 1995); (3) a steady age-related change of N170 is observed and VPP is absent in young children, indicating a developmental difference between N170 and VPP (Taylor et al., 1999); (4) cognitive decline affects the VPP and N170 in the opposite direction (Saavedra et al., 2012); (5) the face-orientation inversion effect is reliable on the amplitude of N170, but not on the amplitude of VPP (Itier & Taylor, 2004).

This study aimed to investigate whether instruction-induced mental imagery of faces can top-down modulate the two face-sensitive ERP components, VPP and N170, when the ERP-eliciting stimuli are completely ambiguous pictures that have no physical features of faces. The controlling condition for face-imagery was mental imagery of houses.

2. Methods

2.1. Participants

Sixteen university students (8 males and 8 females, aged between 20 and 28 years with the mean of 22.9 ± 2.0 years) participated in this study. They were right handed and had either normal or correct-to-normal vision. These participants gave their informed consent before the experiment and were paid a modest stipend for their participation. The experimental procedures were approved by the Committee for Protecting Human and Animal Subjects of the Department of Psychology at Peking University.

2.2. Stimuli

Ten grayscale pictures of faces (5 females and 5 males) and 10 grayscale pictures of houses (5 two-floor ones and 5 three-floor ones) with balanced physical attributes of brightness and contrast were used as the original pictures. The hair and ears of face pictures were removed. The

houses were placed into a face-shape contour to reduce the physical interstimulus variance. These original pictures were transferred to completely ambiguous figures using the phase randomization methods of MATLAB 7.0, leading to that all the pixels of a picture were selected and their phases were randomly redistributed to form a new figure. To avoid potential effects of content of original pictures, these ambiguous figures were classified into two categories according to their original content: the ambiguous figures from face pictures and those from house pictures. Each original picture was processed for 15 times. Thus, in total 150 completely ambiguous figures were made from face pictures and 150 completely ambiguous figures were made from house pictures. After the phase-randomization manipulation, these ambiguous figures had neither recognizable face objects nor recognizable house objects. The third category of stimuli were 150 Gaussian noise stimuli, whose brightness was matched to that of completely ambiguous figures.

In addition to the completely ambiguous figures, pictures containing either a real face or a real house were also used. More specifically, 50 face pictures and 50 house pictures (10 original face pictures and 10 original house pictures each submitted to the processing described below 5 times) were processed into half-ambiguous pictures, in which either a face or a house could be easily observed. The half-ambiguous pictures were produced by randomly selecting 50% of the pixels of the original pictures and randomizing the phases of the pixels. In this way, 50 half-ambiguous face pictures and 50 half-ambiguous house pictures were obtained. In each condition with either the face-picture origin or the house-picture origin, 50 original (clear) pictures and 50 half-ambiguous pictures were used (Fig. 1).

2.3. Procedures

During the experiment, participants sat in a dimly-lit and sound-attenuated room (EMI Shielded Audiometric Examination Acoustic Suite), with a distance of 75 cm away from a 15-in. computer screen (resolution, 1024 × 768 pixels; refresh rate, 60 Hz) that presented the visual stimuli (visual angle, 10.5° × 7.9°). There were two blocks (face-imagery block, house-imagery block) of 550 randomly ordered trials, including in each block 150 presentations of ambiguous figures from face pictures, 150 presentations of ambiguous figures from house pictures, 150 presentations of Gaussian noise images, and either 100 presentations of real-face pictures (50 clear and 50 half-ambiguous ones, only in the face-imagery block) or 100 presentations of real-house pictures (50 clear and 50 half-ambiguous ones, only in the house-imagery block). The order of the 2 blocks was counterbalanced across participants. Each block was divided into 5 sessions and participants had a short rest between sessions.

In a trial, a fixation cross was presented for a duration randomly between 800 and 1000 ms, followed by an 800-ms presentation of a stimulus, and then participants responded to this stimulus by pressing of the two buttons of a joystick (Fig. 1). Before each block, participants were instructed to imagine either a face (in the face-imagery block) or a house (in house-imagery block). The instruction for the face-imagery block was that "We are going to show you a number of face pictures. Some of them are clear, but some are ambiguous. Your task is to judge the gender and press the corresponding buttons. The task in some trials may be challenging and you might find some figures too ambiguous to tell the gender. You just imagine a face from an ambiguous figure and try your best to determine the gender." The instruction for the house-imagery block was similar, except that the participants' task was to judge whether the imaged
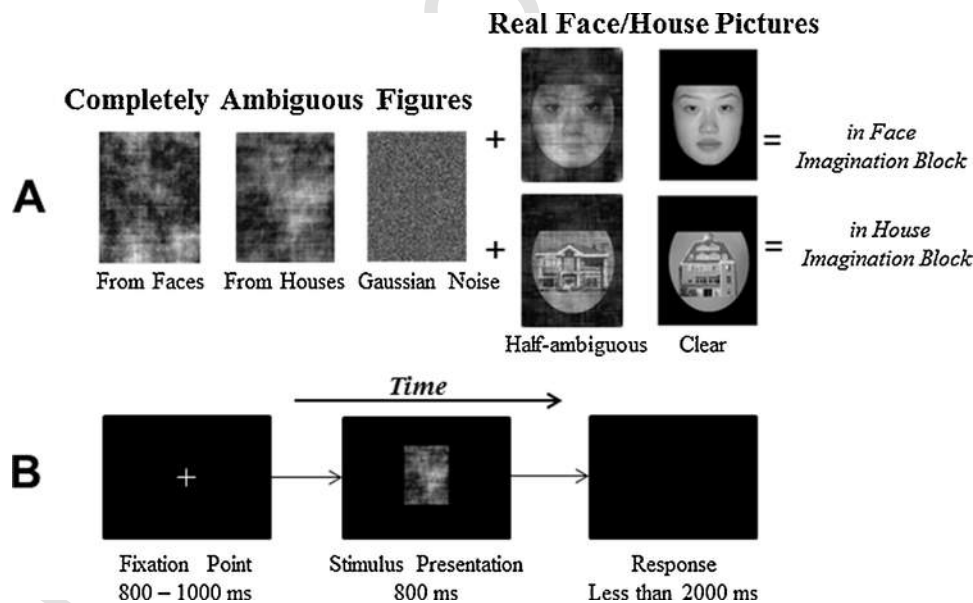

Fig. 1. Stimuli and Procedures. The upper panel (A): Examples of the experimental stimuli. Completely ambiguous figure stimuli were used in both the face-imagery and house-imagery blocks. Corresponding real face/house pictures were used in the face-imagery block and the house-imagery blocks. The lower panel (B): Experimental procedures. During stimulus presentation, participants were instructed to imagine either a face or a house, and then make 2-alternative-forced-choice responses (male/female, two/three floors).

houses had either 2 or 3 floors. In this way, participants performed the 2-alternative-forced-choice (2AFC) gender-judgment task (for the face-imagery block) or floor-judgment task (for the house-imagery block) not only for completely ambiguous figures, but also for real faces in the face-imagery block and for real houses in the house-imagery block. To avoid interferences from participants' actions, participants were instructed to press the button only after the visual stimulus terminated.

Before the formal testing, participants received a training to ensure that they had fully understood the procedure and performed the task correctly. The purpose of this study was fully explained to individual participants only at the end of the experiment.

2.4. EEG recordings

Scalp electro-encephalographic activity was recorded using a 64-channel NeuroScan SynAmps system (Compumedics Limited, Victoria, Australia) with the reference electrode placed on the nose tip. Two vertical and horizontal EOG sites were used to monitor eye movements and eye blinks, placed superior and inferior to the left eye and also at the outer canthi of the two eyes. Electrode impedance was kept below 5 kΩ. EEG signals were processed at on-line band-pass from 0.05 to 100 Hz and digitized with a sample rate of 1000 Hz. Data average was performed after sorting by stimulus type (ambiguous figures from faces, ambiguous figures from houses, Gaussian noise, real face/house pictures) and imagery category (face block, house block). EEG and EOG were epoched from 200 ms before (served as baseline for the baseline correction) and 800 ms after the stimulus onset. Epochs with ocular artifacts exceeding ± 100 μV were rejected, and then the waveform was off-line low-pass filtered at the cut-off frequency of 20 Hz (12 dB/octave). Linear deflections were removed before averaging.

The peak amplitude and latency of the N170 component were processed at the occipito-temporal sites in the right hemisphere (P6, P8, PO6, PO8) and left hemisphere (P5, P7, PO5, PO7) during the time window of 140–220 ms after the stimulus onset. The VPP was processed at the fronto-central sites of FCz and Cz 140–220 ms after the stimulus onset; the occipito-temporal P2 component was processed at the sites in the right hemisphere (P6, P8, PO6, PO8) and left hemisphere (P5, P7, PO5, PO7) during the time window of 200–260 ms after the stimulus onset. In the peak-to-peak measurement of VPP, frontal N1 (the negative peak around 100–120 ms after the stimulus onset) was processed at the electrode sites of FCz and Cz during the time window of 70–150 ms after the stimulus onset, and the N1-VPP peak-to-peak amplitude was obtained by subtracting the baseline-to-peak amplitude of frontal N1 from the baseline-to-peak amplitude of VPP.

3. Results

3.1. Behavioral testing

The results of the behavioral testing showed that the mean accuracy rate of the 2AFC task for recognizing real face pictures and real house pictures were above 0.85 (face, 0.94 ± 0.01; house, 0.88 ± 0.02), indicating that participants were able to concentrate their attention on the stimulus,

perform the task, and discriminate features of the real objects (faces or houses).

For the completely ambiguous figures, in the face-imagery task and in the in the house-imagery task participants judged the figures as either male or female or as two or three floors only at random. Statistically, in the face-imagery task, no significant differences were found between the probability of judging the ambiguous pictures as either male faces (0.56 ± 0.03) or female faces (0.44 ± 0.03). In the house-imagery task, no significant difference was found between the probability of judging the ambiguous pictures as either two-floor houses (0.46 ± 0.03) or three-floor houses (0.54 ± 0.03).

3.2. ERP recordings

3.2.1. Amplitudes and latencies of ERPs to completely ambiguous figures

Fig. 2 shows the group grand average ERP waveforms induced by each of the 3 types of ambiguous stimuli in either the face- imagery condition or the house- imagery condition with recordings from the electrode sites of Cz and P8. Clearly, the ambiguous stimuli reliably elicited both the VPP component (at site Cz) and the P2 component (at site P8), but not the N170 component (also see below).

The results of recordings at site Cz showed that the amplitude of VPP was larger under the face-imagery condition than under the house-imagery condition. The amplitude and latency of VPP were statistically analyzed using a repeated-measures analyses of variance (ANOVA) with the following 3 variables: (1) imagery category (face imagery, house imagery), (2) stimulus type (ambiguous figures from face pictures, ambiguous figures from house pictures, Gaussian noise), and (3) electrode site (FCz, Cz). A three-variable within-subject ANOVA showed a significant main effect of imagery category [$F(1,15) = 18.27$, $p = 0.001$, $\eta p^2 = 0.55$], significant main effect of stimulus type [$F(2,30) = 6.58$, $p = 0.004$, $\eta p^2 = 0.31$], and significant main effect of electrode site [$F(1,15) = 4.87$, $p = 0.043$, $\eta p^2 = 0.25$]. No significant interactions were observed (all $p > 0.05$). LSD Post hoc analyses showed that the amplitude of VPP was significantly larger under the face-imagery condition (2.95 ± 0.63 μV) than under the house-imagery condition (1.80 ± 0.52 μV) [$t(15) = 4.28$, $p < 0.001$, Cohen's $d = 0.50$]. Thus, although identical ambiguous stimuli were used, introducing mental imagery of face led to stronger VPP than introducing mental imagery of house.

Moreover, the mean VPP amplitude to the ambiguous stimuli from face pictures (2.87 ± 0.64 μV) was larger than mean VPP amplitude to Gaussian noise (1.74 ± 0.50 μV) [$t(15) = 3.42$, $p = 0.004$, Cohen's $d = 0.49$], and the mean VPP amplitude to the ambiguous stimuli from house pictures (2.51 ± 0.62 μV) was larger than mean VPP amplitude to Gaussian noise [$t(15) = 2.49$, $p = 0.025$, Cohen's $d = 0.34$]. Also, no significant difference was found between VPP elicited by ambiguous stimuli from face pictures and that elicited by ambiguous stimuli from house pictures ($p = 0.27$). Thus, the effect of original picture content before phase randomization was not significant. The results also showed that VPP at the site Cz (2.68 ± 0.65 μV) was larger than that at the site FCz (2.07 ± 0.49 μV)
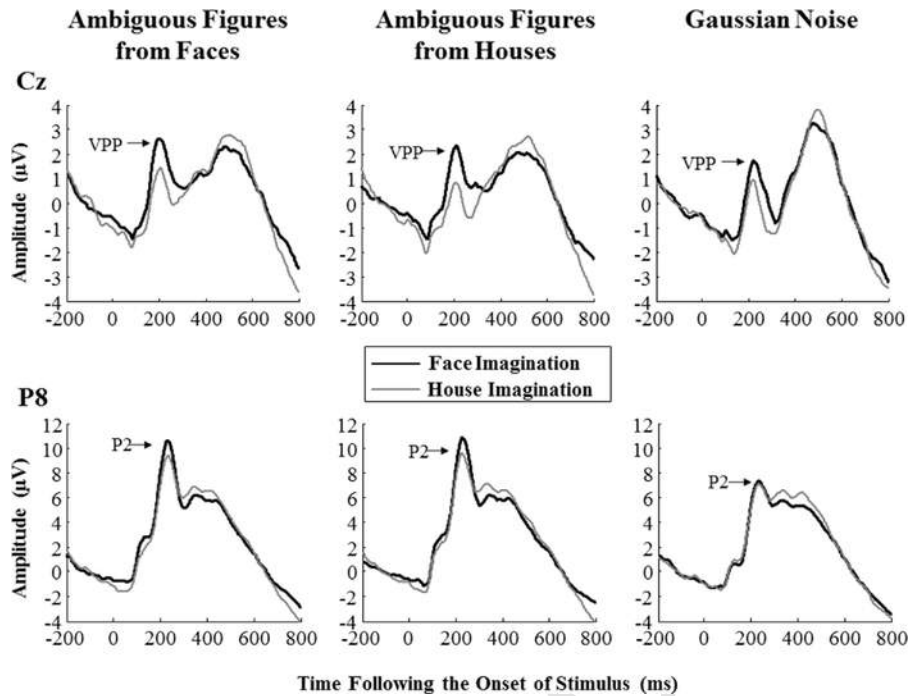
Fig. 2. Waveforms of VPP and P2. Group grand average waveforms at electrode site Cz (the upper panel) and P8 (the lower panel) under either the face-imagery condition or the house-imagery condition. VPP (time window: 140–220 ms) and P2 (time window: 200–260 ms) were enhanced under the face-imagery condition (dark line) compared to the house-imagery condition (grey line).

[t(15) = 2.27, p = 0.043, Cohen's d = 0.27]. The comparisons in VPP amplitudes are shown in the upper panel of Fig. 3.

For the latency of VPP, neither the main effect for each of the 3 variables nor the interactions between the variables were significant. The mean latency of VPP was 191.4 ± 8.5 ms and 191.4 ± 7.3 ms under the face-imagery condition and the house-imagery condition, respectively.

The amplitude and latency of P2 were analyzed also using a repeated-measures ANOVA with the following 3 variables (1) imagery category (face imagery, house imagery), (2) stimulus type (ambiguous stimuli from face pictures, ambiguous stimuli from house pictures, Gaussian noise), and (3) hemisphere (left, right). Greenhouse-Geisser Adjustment was used for violation of Sphericity Assumption.

The lower panel of Fig. 2 shows the comparisons of grand average ERP waveforms under the face-imagery condition and those under the house-imagery condition with the recordings at the electrode site of P8. For the P2 amplitude, a three-variable within-subject ANOVA revealed a significant main effect of imagery category [F(1,15) = 4.58, p = 0.049, ηp^2 = 0.23], significant main effect of stimulus type [F(2.30) = 35.19, p < 0.001, ηp^2 = 0.70], and significant main effect of hemisphere [F(1,15) = 5.58, p = 0.032, ηp^2 = 0.27]. For interactions among the 3 variables, only the interaction between stimulus type and hemisphere was significant [F(1.32, 19.79) = 10.55, p = 0.002, ηp^2 = 0.41]. LSD Post hoc comparison showed that the amplitude of P2 was larger under the face-imagery condition (10.38 ± 0.85 μV) than under the house-imagery condition (9.44 ± 0.98 μV) [t(15) = 2.14, p = 0.049, Cohen's d = 0.27].

Moreover, the simple effect of the two-way (stimulus type by hemisphere) interaction was analyzed. The amplitude of P2 in the right hemisphere was larger than that in the left hemisphere when the stimuli were either ambiguous ones from face pictures [t(15) = 2.36, p = 0.032, Cohen's d = 0.42] or ambiguous ones from house pictures [t(15) = 2.94, p = 0.010, Cohen's d = 0.46], but not when the stimuli were Gaussian-noise ones (p = 0.179). The comparisons of P2 amplitudes are shown in the lower panel of Fig. 3.

For the latency of P2, a three-variable within-subject ANOVA only revealed a significant main effect of stimulus type [F(1.28, 19.25) = 4.25, p = 0.034, ηp^2 = 0.22]. LSD Post Hoc comparison showed that P2 induced by ambiguous stimuli from face pictures (229.0 ± 3.5 ms) occurred earlier than P2 induced by Gaussian noise (235.2 ± 3.4 ms) [t(15) = 2.22, p = 0.043, Cohen's d = 0.45]. The difference between P2 latency to ambiguous stimuli from house pictures (229.0 ± 3.3 ms) and that to Gaussian noise was marginal [t(15) = 2.10, p = 0.053, Cohen's d = 0.47]. The mean latency of P2 was 229.1 ± 3.3 ms and 223.1 ± 3.3 ms under the face-imagery condition and the house-imagery condition, respectively.

ERP waveforms in 5 participants exhibited an observable N170 component under either the face-imagery condition or the house-imagery condition. No N170 component was observable from the grand average ERP waveform across 16 participants.

The grand average ERP waveforms of all the analyzed electrode sites (FCz, Cz, P5, P6, P7, P8, PO5, PO6, PO7, PO8) under face-imagery and house-imagery conditions for the 4 stimulus types (completely ambiguous figures from
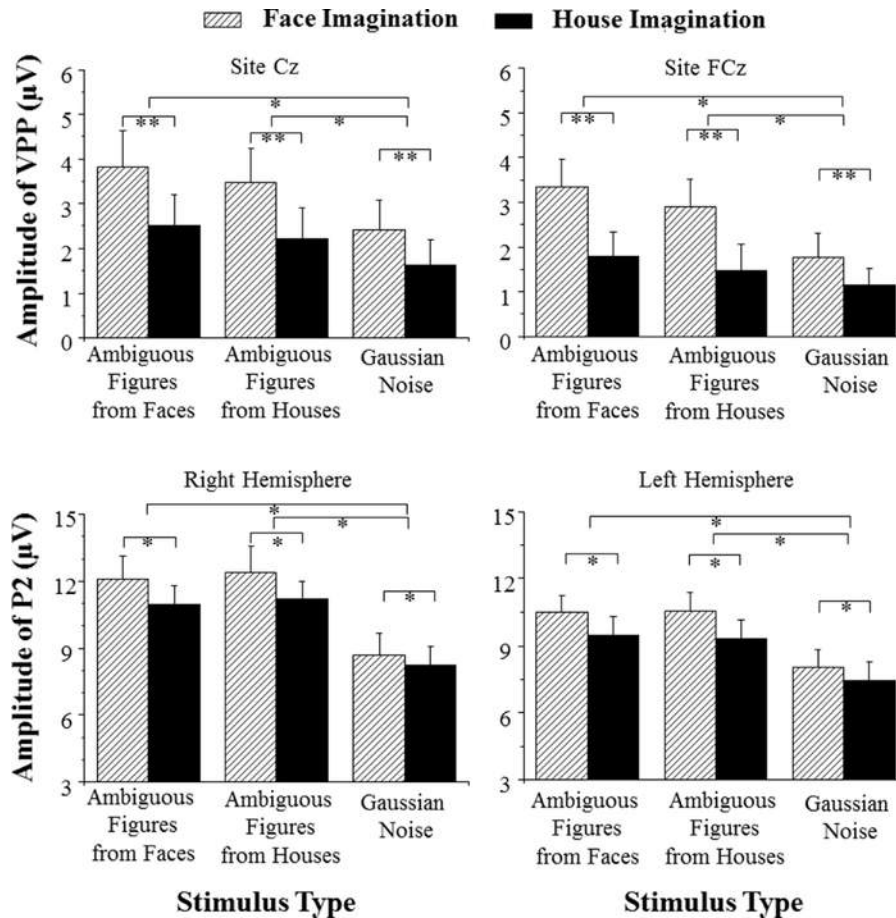
Fig. 3. Statistical comparisons of VPP and P2 amplitudes. The upper panel: Post Hoc comparisons of peak amplitudes of VPP at electrode sites Cz and FCz. The amplitude of VPP under the face-imagery condition was significantly larger than that under the house-imagery condition, and VPP elicited by ambiguous pictures was significantly larger than that elicited by Gaussian noise. The lower panel: Post Hoc comparisons of peak amplitudes of P2 in the left hemisphere (P5, P7, PO5, PO7) and the right hemisphere (P6, P8, PO6, PO8). The amplitude of P2 under the face-imagery condition was significantly larger than that under the house-imagery condition, and P2 elicited by ambiguous figures was significantly larger than that elicited by Gaussian noise. Error Bars represent the standard error of the mean (SEM). * $p < 0.05$, ** $p < 0.01$.

faces, completely ambiguous figures form houses, Gaussian noise and real pictures) are shown in Fig. 4.

3.2.2. Comparisons of ERPs under the face-imagery condition and ERPs to real face-picture stimulation

Based on the results described above, the influence of original picture content can be excluded. To compare ERPs to ambiguous stimuli under the face-imagery condition and ERPs to real face-picture stimulation (including both half-ambiguous face pictures and clean face pictures), we merged ambiguous figures from face pictures and those from house pictures to form a single ambiguous-figure condition, and then compared ERPs to these ambiguous figures under the face-imagery condition with ERPs to real face pictures.

As shown in Fig. 5, both ambiguous figures under the face-imagery condition and real face pictures elicited the VPP component at the site Cz and the P2 component at the site P8. Also, the face-specific N170 component was reliably elicited only by the real face pictures (not by the ambiguous figures).

For the amplitude of VPP component, a 2 (stimulus type: ambiguous pictures with face imagery, face-picture stimula-

tion) by 2 (electrode site: Cz, FCz) repeated-measures ANOVA showed a significant two-way interaction [F(1, 15) = 63.98, $p < 0.001$, $\eta p^2 = 0.81$]. The difference of VPP amplitude between the two stimulus types (face-picture stimulation minus ambiguous picture with face imagery) was larger at site FCz (1.93 ± 0.93 μV) compared to that at site Cz (0.16 ± 0.91 μV) [t(15) = 8.00, $p < 0.001$, Cohen's $d = 0.48$]. Also, the latency of VPP was analyzed and the results showed that no significant differences were found between 2 stimulus types (186.0 ± 9.2 ms for completely ambiguous stimuli under the face-imagery condition; 185.3 ± 4.6 ms for real-face-picture stimulation).

At the occipito-temporal sites, as the positive potential P2 partially overlapped with the negative potential N170, the amplitude of P2 elicited by face-picture stimuli was notably reduced, smaller than that induced by ambiguous stimuli under the face-imagery condition. There was no significant difference in peak latency of P2 between the 2 stimulus types (227.6 ± 3.5 ms for completely ambiguous stimuli under the face-imagery condition; 232.2 ± 3.1 ms for face-picture stimulation).
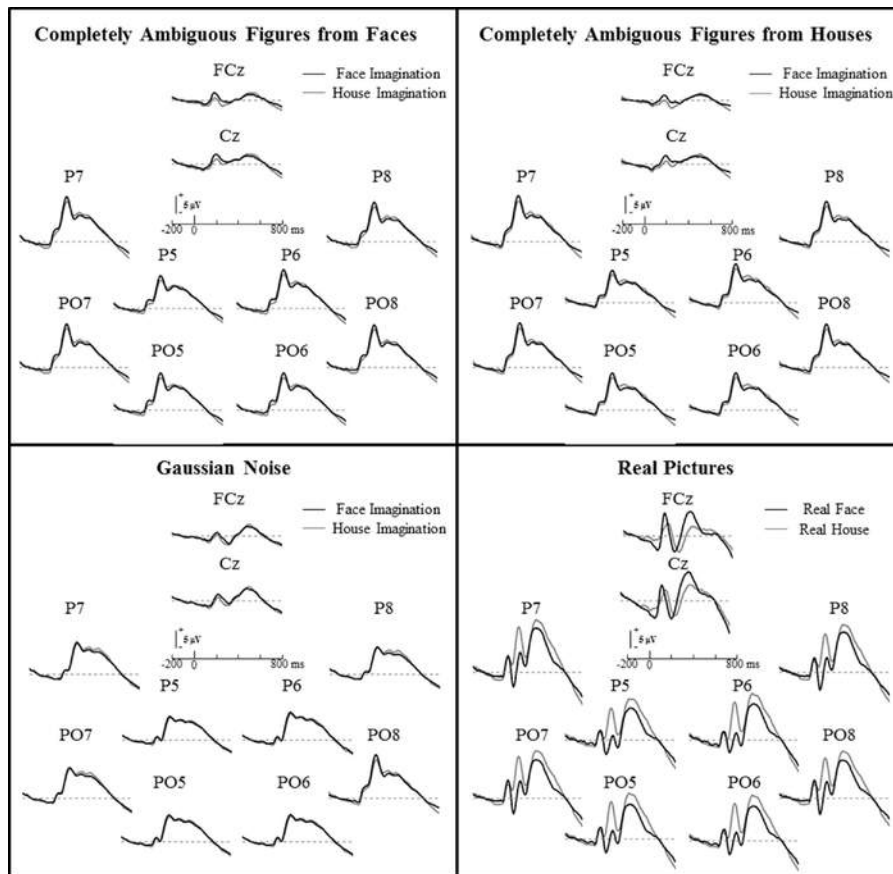
Fig. 4. Group grand average waveforms at all analyzed electrode sites. For the stimulus type of completely ambiguous figures from faces, completely ambiguous figures from houses, and Gaussian noise, the dark lines represent the face-imagery condition and the grey lines represent house-imagery condition. For the stimulus type of real pictures, the dark lines represent real-face-picture condition and the grey lines represent real-house-picture condition. The electrode sites included FCz, Cz, P5, P6, P7, P8, PO5, PO6, PO7, PO8.
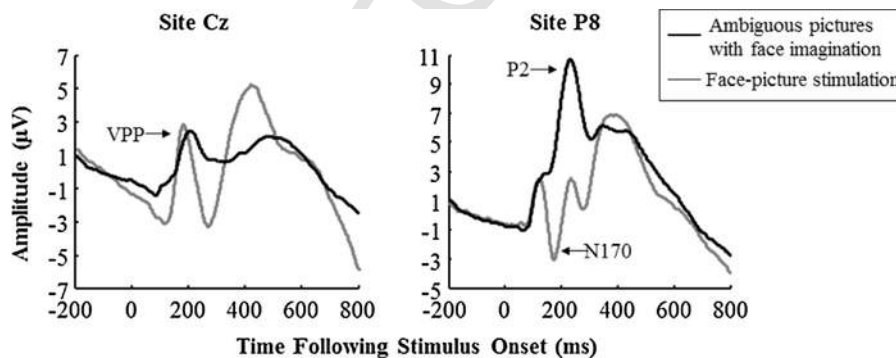

Fig. 5. Group grand average waveforms at electrode sites Cz and P8. Dark lines represent responses elicited by ambiguous figures with face imagery, and grey lines represent responses induced by real face pictures. VPP and P2 occurred in both the face-imagery condition and face-perception condition, but N170 was found only in the face-perception condition.

3.2.3. Comparisons of preference of VPP to face between the ambiguous-figure condition and the real-figure condition

Since both VPP to completely ambiguous figures and VPP to real pictures exhibited the marked preference to face, comparisons of the preference of VPP to face were conducted between the ambiguous-figure condition and the real-picture condition. As shown in Fig. 6, when the stimuli were either completely ambiguous figures or real pictures, face imagery/perception caused larger VPP amplitudes than house imagery/perception.

Fig. 6A plots the grand average ERP waveforms at fronto-central sites (averaged across the FCz and Cz sites). In the waveforms, a negative deflection (frontal N1) before VPP notably differed between the condition with completely ambiguous figures and the condition with real pictures. Therefore, besides the baseline-to-peak VPP amplitudes, peak-to-peak amplitudes of N1-VPP were also included in the following analyses.

For the VPP amplitudes shown in the left panel of Fig. 6B, a 2 (stimulus type: completely ambiguous figure, real picture) by 2 (stimulus-object condition: face, house) re-
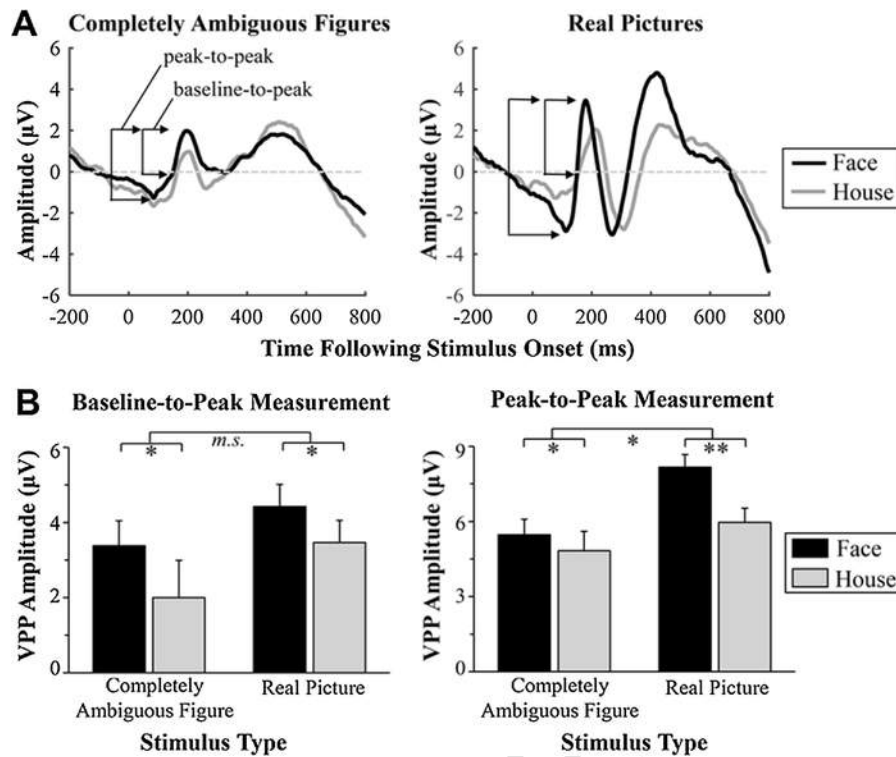
Fig. 6. The upper panels (A): Group grand average of ERP waveforms at fronto-central sites (averaged across the FCz and Cz sites) elicited either by completely ambiguous figures (left panel) with either face imagery (dark line) or house imagery (grey line), or by real-face pictures (dark line in right panel) or real-house pictures (grey line in right panel). The lower panels (B): Post hoc comparisons of baseline-to-peak amplitudes of VPP (left panel) and peak-to-peak amplitudes of N1-VPP (right panel) across stimulus types and conditions. Both baseline-to-peak and peak-to-peak measurements showed stronger VPP under face conditions (dark bar) than house conditions (grey bar). The peak-to-peak measurement also reveals a stronger face-advantage effect for real pictures than completely ambiguous figures. Error Bars represent the standard error of the mean (SEM). * $p < 0.05$, ** $p < 0.01$, *m.s.* marginally significant.

peated-measures ANOVA of the baseline-to-peak VPP amplitude revealed a marginally significant main effect of stimulus type [$F_{(1, 15)} = 4.40$, $p = 0.053$, $\eta p^2 = 0.23$] and a significant main effect of stimulus-object condition [$F_{(1, 15)} = 4.93$, $p = 0.042$, $\eta p^2 = 0.25$]. The two-way interaction was not significant ($p = 0.63$). A LSD Post hoc analysis also showed that VPP under the face condition (3.91 ± 0.71 μV) was significantly larger than that under the house condition (2.73 ± 0.51 μV) [$t(15) = 2.22$, $p = 0.042$, Cohen's $d = 0.47$]. However, the difference between VPP to real pictures (3.95 ± 0.67 μV) and that to completely ambiguous figures (2.69 ± 0.61 μV) was only marginally significant [$t(15) = 2.10$, $p = 0.053$, Cohen's $d = 0.49$]. The results showed that the preference of VPP to face imagery under the ambiguous-figure condition was similar to the preference of VPP to real-figure stimuli under the real-picture condition.

For the VPP amplitudes shown in the right panel of Fig. 6B, a 2 (stimulus type: completely ambiguous figures, real pictures) by 2 (stimulus-object condition: face, house) repeated-measures ANOVA of peak-to-peak amplitude revealed a significant main effect of stimulus type [$F_{(1, 15)} = 22.57$, $p < 0.001$, $\eta p^2 = 0.60$], a significant main effect of stimulus-object condition [$F_{(1, 15)} = 11.89$, $p = 0.004$, $\eta p^2 = 0.44$], and a significant two-way interaction between the two factors [$F_{(1, 15)} = 5.52$, $p = 0.033$, $\eta p^2 = 0.27$]. For completely ambiguous figures, the peak-to-peak amplitude of VPP under face imagery (5.47 ± 0.62 μV) was significantly

larger than that under house imagery (4.83 ± 0.50 μV) [$t(15) = 2.28$, $p = 0.037$, Cohen's $d = 0.29$]. For real pictures, the real-face picture also elicited stronger VPP (8.17 ± 0.78 μV) than real-house picture did (5.96 ± 0.56 μV) [$t(15) = 3.17$, $p = 0.006$, Cohen's $d = 0.81$]. Also, the face/house difference for real pictures (2.21 ± 0.70 μV) was significantly larger than that for completely ambiguous figures (0.64 ± 0.28 μV) [$t(14) = 2.35$, $p = 0.033$, Cohen's $d = 0.74$]. The results of peak-to-peak-amplitude data showed that for mental imagery, the face-preferential effect could also be presented by the N1-VPP amplitude.

4. Discussion

As mentioned in the Introduction, previous studies have shown that a preference to face stimulation exists in the two ERP components, N170 and VPP. The aim of the present study was to examine whether a preference of N170 and/or VPP to face also occurs in top-down mental imagery. To achieve this aim, face features of bottom-up visual stimuli were completely eliminated, and non-face (house) mental imagery was introduced as the control for face mental imagery.

In this study, presenting real-face pictures reliably elicited both the VPP component at the fronto-central sites and the N170/P2 component at the occipito-temporal sites. Nevertheless, under either the face-imagery   or   house-im-

agery condition, presenting completely ambiguous pictures only reliably elicited the VPP and P2 components, but not the N170 component (which was not detectable in the majority of the participants). Obviously, the results of this study did not fully support the notion that VPP and N170 manifest the identical brain processes (Jemel et al., 2003; Joyce & Rossion, 2005).

It has been known that N170 mainly reflects the early and automatic structural encoding stages in face processing (Bentin & Deouell, 2000; Bentin et al., 1996; Cauquil et al., 2000; Eimer, 2000). N170 can be reliably elicited by noise images only when bottom-up face signals or object signals are available, such as when working memory of a real face image is well maintained (Sreenivasan et al., 2007) or when a noise image is forcedly treated as a face, word, or house (Wild & Busey, 2004). Thus, N170 mainly reflects the bottom-up central representation of signals of structured objects (particularly faces) and is more vulnerable to destruction and noise masking.

Although VPP has similar response properties to N170 (Bötzel & Grüsser, 1989; Itier & Taylor, 2002; Jeffreys, 1996; Rossion, Delvenne et al., 1999), particularly that the changing pattern of amplitudes and peak latencies of VPP and those of N170 are highly similar across reference sites and even correlate to each other (Joyce & Rossion, 2005; Jemel et al., 2003), substantial evidence suggests that VPP and N170 are not the two "faces" of the same underlying neural generator (Bötzel et al., 1995; George et al., 1996; Itier & Taylor, 2004; Saavedra et al., 2012; Taylor et al., 1999). The results of this study make it clear that when bottom-up face signals are absent, VPP, but not N170, can be not only reliably elicited by completely ambiguous noises but also top-down enhanced by mental imagery of faces relative to mental imagery of non-face objects (houses). In other words, in addition to that both VPP and N170 are associated with face-specific bottom-up perceptual processing (e.g., Bentin & Deouell, 2000; Eimer, 2000; George et al., 1996; Itier & Taylor, 2002; Joyce & Rossion, 2005; Jemel et al., 2003; Rossion, Delvenne et al., 1999; Rossion et al., 2003), this study provides evidence showing that a preferential effect of top-down visual mental imagery of faces can be reflected by changes in VPP, even though the preferential effect of mental imagery was possibly weaker than that during perceiving real face pictures.

It has been suggested that the VPP generator involves a larger network including the fusiform gyrus (Rossion, Campanella et al., 1999; Johnston, Stojanov, Devir, & Schall, 2005), the inferio-temporal cortex and the superior temporal sulcus (George et al., 1996; Jeffreys, 1996), and probably sequential activation of occipital, lateral temporal, and mesio-temporal cortical structures (Bötzel et al., 1995). These ventral occipito-temporal regions play a role in representing forms of objects with object-category selectivity (Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999; Ishai, Ungerleider, Martin, & Haxby, 2000) based on ascending inputs from early visual areas including the inferior occipital cortex (Mechelli, Price, Noppeney, & Friston, 2003; Mechelli, Price, Friston, & Ishai, 2004). Moreover, the ventral occipito-temporal regions are also modulated by mental imagery with object-category selectivity (Ishai, Ungerleider, & Haxby, 2000; Ishai, Haxby, & Ungerleider, 2002; Kosslyn & Thompson, 2003; O'Craven & Kanwisher, 2000; Mechelli et al., 2003, 2004). The Mechelli et al. study (2004) has

shown that during visual perception, the category selective effect in the occipito-temporal regions are mediated by ascending inputs from early visual areas; during visual imagery, the category selective effects in the occipito-temporal regions are mediated by descending inputs from the prefrontal cortex. Thus, it is proposed that the preference of top-down mental imagery to face can be represented in changes of VPP, and the face-imagery-induced enhancement of VPP to completely ambiguous noises reflects the mental imagery-induced modulation of the occipito-temporal regions by descending inputs from the prefrontal cortex.

Surprisingly, in this study, the P2 component of ERPs to ambiguous stimuli was remarkable at occipito-temporal sites and exhibited a difference in amplitude between the face-imagery condition and the house-imagery condition. It has been suggested P2 functions as an index of the processing stage for face racial perception (Keyes, Brady, Reilly, & Foxe, 2010; Stahl, Wiese, & Schweinberger, 2008) and even deeper cognitive processing of categorizing ambiguous faces (Latinus & Taylor, 2005, 2006). Interestingly, adding noise to face pictures leads to an increase in the P2 amplitude (Bankó, Gál, Körtvélyes, Kovács, & Vidnyánszky, 2011; Németh, Kovács, Vakli, Kovács, & Zimmer, 2014) but a decrease in the N170 amplitude during face perception (Jemel et al., 2003; Horovitz et al., 2004; Bankó et al., 2011; Németh et al., 2014). This study, for the first time, suggests that P2 can be top-down modulated by mental imagery, showing a category selectivity. Since both P2 and VPP are top-down modulated by mental imagery simultaneously, in the future it is important to investigate the functional integration between the neural circuitry underlying P2 and that underlying VPP.

5. Conclusions

(1) The present study indicates that VPP but not N170 can be reliably elicited by completely ambiguous stimuli when object structural features are eliminated, suggesting that the brain substrates underlying VPP are not fully identical to those underlying N170.

VPP can be modulated by mental imagery with a preference to face signals, when the bottom-up inputs contains no object structural features, indicating that VPP reflects an interface of integrating bottom-up perception signals and top-down mental imagery signals with category selectivity.

(2) Ambiguous-picture-evoked P2 also responds preferentially to face mental imagery when bottom-up object structural features are eliminated.

(3) The interaction between the neural substrates underlying P2 and those underlying VPP will be an important issue in future studies.

Statement of conflicts of interest

None declared.

Acknowledgements

References

Bötzel, K., Grüsser, O.J., 1989. Electric brain potentials evoked by pictures of faces and non-faces: A search for face-specific EEG-potentials. Experimental Brain Research 77 (2), 349–360.

Bötzel, K., Schulze, S., Stodieck, S.R., 1995. Scalp topography and analysis of intracranial sources of face-evoked potentials. Experimental Brain Research 104 (1), 135–143.

Bankó, É.M., Gál, V., Körtvélyes, J., Kovács, G., Vidnyánszky, Z., 2011. Dissociating the effect of noise on sensory processing and overall decision difficulty. Journal of Neuroscience 31 (7), 2663–2674.

Bentin, S., Deouell, L.Y., 2000. Structural encoding and identification in face processing: ERP evidence for separate mechanisms. Cognitive Neuropsychology 17 (1-3), 35–55.

Bentin, S., Golland, Y., 2002. Meaningful processing of meaningless stimuli: The influence of perceptual experience on early visual processing of faces. Cognition 86 (1), B1–B14.

Bentin, S., Allison, T., Puce, A., Perez, E., McCarthy, G., 1996. Electrophysiological studies of face perception in humans. Journal of Cognitive Neuroscience 8 (6), 551–565.

Bentin, S., Sagiv, N., Mecklinger, A., Friederici, A., von Cramon, Y.D., 2002. Priming visual face-processing mechanisms: Electrophysiological evidence. Psychological Science 13 (2), 190–193.

Cauquil, A.S., Edmonds, G.E., Taylor, M.J., 2000. Is the face-sensitive N170 the only ERP not affected by selective attention?. Neuroreport 11 (10), 2167–2171.

De Haan, M., Pascalis, O., Johnson, M.H., 2002. Specialization of neural mechanisms underlying face recognition in human infants. Journal of Cognitive Neuroscience 14 (2), 199–209.

Eimer, M., 2000. Event-related brain potentials distinguish processing stages involved in face perception and recognition. Clinical Neurophysiology 111 (4), 694–705.

Ganis, G., Schendan, H.E., 2011. Visual imagery. Wiley Interdisciplinary Reviews: Cognitive Science 2 (3), 239–252.

George, N., Evans, J., Fiori, N., Davidoff, J., Renault, B., 1996. Brain events related to normal and moderately scrambled faces. Cognitive Brain Research 4 (2), 65–76.

Gosselin, F., Schyns, P.G., 2003. Superstitious perceptions reveal properties of internal representations. Psychological Science 14 (5), 505–509.

Hadjikhani, N., Kveraga, K., Naik, P., Ahlfors, S.P., 2009. Early (N170) activation of face-specific cortex by face-like objects. Neuroreport 20 (4), 403–407.

Hansen, B.C., Thompson, B., Hess, R.F., Ellemberg, D., 2010. Extracting the internal representation of faces from human brain activity: An analogue to reverse correlation. Neuroimage 51 (1), 373–390.

Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2000. The distributed human neural system for face perception. Trends in Cognitive Sciences 4 (6), 223–233.

Horovitz, S.G., Rossion, B., Skudlarski, P., Gore, J.C., 2004. Parametric design and correlational analyses help integrating fMRI and electrophysiological data during face processing. Neuroimage 22 (4), 1587–1595.

Ishai, A., Ungerleider, L.G., Martin, A., Schouten, J.L., Haxby, J.V., 1999. Distributed representation of objects in the human ventral visual pathway. Proceedings of the National Academy of Sciences 96 (16), 9379–9384.

Ishai, A., Haxby, J.V., Ungerleider, L.G., 2002. Visual imagery of famous faces: Effects of memory and attention revealed by fMRI. Neuroimage 17 (4), 1729–1741.

Ishai, A., Ungerleider, L.G., Haxby, J.V., 2000. Distributed neural systems for the generation of visual images. Neuron 28 (3), 979–990.

Ishai, A., Ungerleider, L.G., Martin, A., Haxby, J.V., 2000. The representation of objects in the human occipital and temporal cortex. Journal of Cognitive Neuroscience 12 (Suppl. 2), 35–51.

Itier, R.J., Taylor, M.J., 2002. Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: A repetition study using ERPs. Neuroimage 15 (2), 353–372.

Itier, R.J., Taylor, M.J., 2004. N170 or N1? Spatiotemporal differences between object and face processing using ERPs. Cerebral Cortex 14 (2), 132–142.

Jeffreys, D.A., Tukmachi, E., 1992. The vertex-positive scalp potential evoked by faces and by objects. Experimental Brain Research 91 (2), 340–350.

Jeffreys, D.A., 1989. A face-responsive potential recorded from the human scalp. Experimental Brain Research 78 (1), 193–202.

Jeffreys, D.A., 1996. Evoked potential studies of face and object processing. Visual Cognition 3 (1), 1–38.

Jemel, B., Schuller, A.M., Cheref-Khan, Y., Goffaux, V., Crommelinck, M., Bruyer, R., 2003. Stepwise emergence of the face-sensitive N170 event-related potential component. Neuroreport 14 (16), 2035–2039.

Johnston, P.J., Stojanov, W., Devir, H., Schall, U., 2005. Functional MRI of facial emotion recognition deficits in schizophrenia and their electrophysiological correlates. European Journal of Neuroscience 22 (5), 1221–1232.

Joyce, C., Rossion, B., 2005. The face-sensitive N170 and VPP components manifest the same brain processes: The effect of reference electrode site. Clinical Neurophysiology 116 (11), 2613–2631.

Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: A module in human extrastriate cortex specialized for face perception. The Journal of Neuroscience 17 (11), 4302–4311.

Keyes, H., Brady, N., Reilly, R.B., Foxe, J.J., 2010. My face or yours? Event-related potential correlates of self-face processing. Brain and Cognition 72 (2), 244–254.

Klein, I., Dubois, J., Mangin, J., 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 (1), 26–31.

Kosslyn, S.M., Thompson, W.L., 2003. When is early visual cortex activated during visual mental imagery?. Psychological Bulletin 129 (5), 723–746.

Latinus, M., Taylor, M.J., 2005. Holistic processing of faces: Learning effects with Mooney faces. Journal of Cognitive Neuroscience 17 (8), 1316–1327.

Latinus, M., Taylor, M.J., 2006. Face processing stages: Impact of difficulty and the separation of effects. Brain Research 1123 (1), 179–187.

Li, J., Liu, J., Liang, J., Zhang, H., Zhao, J., Rieth, C.A., et al., 2010. Effective connectivities of cortical regions for top-down face processing: A dynamic causal modeling study. Brain Research 1340, 40–51.

Liu, J., Li, J., Feng, L., Li, L., Tian, J., Lee, K., 2014. Seeing Jesus in toast: Neural and behavioral correlates of face pareidolia. Cortex 53, 60–77.

Mechelli, A., Price, C.J., Noppeney, U., Friston, K.J., 2003. A dynamic causal modeling study on category effects: Bottom-up or top-down mediation?. Journal of Cognitive Neuroscience 15 (7), 925–934.

Mechelli, A., Price, C.J., Friston, K.J., Ishai, A., 2004. Where bottom-up meets top-down: Neuronal interactions during perception and imagery. Cerebral Cortex 14 (11), 1256–1265.

Németh, K., Kovács, P., Vakli, P., Kovács, G., Zimmer, M., 2014. Phase noise reveals early category-specific modulation of the event-related potentials. Frontiers in Psychology 5 (367)http://dx.doi.org/10.3389/fpsyg.2014.00367.

Nestor, A., Vettel, J.M., Tarr, M.J., 2013. Internal representations for face detection: An application of noise-based image classification to BOLD responses. Human Brain Mapping 34 (11), 3101–3115.

O'Craven, K.M., Kanwisher, N., 2000. Mental imagery of faces and places activates corresponding stimulus-specific brain regions. Journal of Cognitive Neuroscience 12 (6), 1013–1023.

Proverbio, A.M., Galli, J., 2016. Women are better at seeing faces where there are none: An ERP study of face pareidolia. Social Cognitive and Affective Neuroscience 11 (9), 1501–1512.

Righart, R., Andersson, F., Schwartz, S., Mayer, E., Vuilleumier, P., 2009. Top-down activation of fusiform cortex without seeing faces in prosopagnosia. Cerebral Cortex 20 (8), 1878–1890.

Rossion, B., Jacques, C., 2008. Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. Neuroimage 39 (4), 1959–1979.

Rossion, B., Gauthier, I., Tarr, M.J., Despland, P., Bruyer, R., Linotte, S., et al., 2000. The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: An electrophysiological account of face-specific processes in the human brain. Neuroreport 11 (1), 69–72.

Rossion, B., Joyce, C.A., Cottrell, G.W., Tarr, M.J., 2003. Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. Neuroimage 20 (3), 1609–1624.

Rossion, B., 2014. Understanding face perception by means of human electrophysiology. Trends in Cognitive Sciences 18 (6), 310–318.

Rossion, B., Campanella, S., Gomez, C.M., Delinte, A., Debatisse, D., Liard, L., et al., 1999. Task modulation of

brain activity related to familiar and unfamiliar face processing: An ERP study. Clinical Neurophysiology 110 (3), 449–462.

Rossion, B., Delvenne, J., Debatisse, D., Goffaux, V., Bruyer, R., Crommelinck, M., et al., 1999. Spatio-temporal localization of the face inversion effect: An event-related potentials study. Biological Psychology 50 (3), 173–189.

Saavedra, C., Olivares, E.I., Iglesias, J., 2012. Cognitive decline effects at an early stage: Evidence from N170 and VPP. Neuroscience Letters 518 (2), 149–153.

Smith, M.L., Gosselin, F., Schyns, P.G., 2012. Measuring internal representations from behavioral and brain data. Current Biology 22 (3), 191–196.

Sreenivasan, K.K., Katz, J., Jha, A.P., 2007. Temporal characteristics of top-down modulations during working memory maintenance: An event-related potential study of the N170 component. Journal of Cognitive Neuroscience 19 (11), 1836–1844.

Stahl, J., Wiese, H., Schweinberger, S.R., 2008. Expertise and own-race bias in face processing: An event-related potential study. Neuroreport 19 (5), 583–587.

Taylor, M.J., McCarthy, G., Saliba, E., Degiovanni, E., 1999. ERP evidence of developmental changes in processing of faces. Clinical Neurophysiology 110 (5), 910–915.

Wild, H.A., Busey, T.A., 2004. Seeing faces in the noise: Stochastic activity in perceptual regions of the brain may influence the perception of ambiguous stimuli. Psychonomic Bulletin & Review 11 (3), 475–481.

Zhang, H., Liu, J., Huber, D.E., Rieth, C.A., Tian, J., Lee, K., 2008. Detecting faces in pure noise images: A functional MRI study on top-down perception. Neuroreport 19 (2), 229–233.