Encoding of biologically significant information in the human brain: Face and biological motion perception

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Yi Jiang

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Adviser: Sheng He

May 2009

Acknowledgements

I am grateful that I have had the opportunity to study in the department of Psychology at the University of Minnesota and to work with Dr. Sheng He on human visual perception. It has been an eye-opening and life-changing experience for me. Now on my way to be a researcher, I appreciate that my progress, including the work presented here in my dissertation, would not have been possible without the tremendous help from the professors and the students who I have had the luck to interact with.

I would like to especially thank Dr. Sheng He for his painstaking supervision throughout the years. I am grateful to my other committee members Dr. Daniel Kersten, Dr. Gordon Legge, and Dr. Yuhong Yang for their thoughtful comments and suggestions on my dissertation. In addition, I would like to thank my colleagues Xiangchuan Chen, Patricia Costello, Fang Fang, Chengzhi Feng, Jay Hegdé, Miner Huang, Scott Murray, Robert Shannon, and Yang Zhang for all their assistance.

I thank my parents, brother, and all my friends for their love, encouragement and support on my pursuit of a doctoral degree in Cognitive Neuroscience. And I would like to give my special thanks to my girlfriend Wen Zhou for her invaluable support these years, both intellectually and psychologically.

To my parents, Guiqiao Jiang and Gui Liu,
my brother, Rui Jiang,
and my girlfriend, Wen Zhou

Abstract

Inherently social, humans communicate identity, emotion, and intention mainly via visual signals — faces and body motions. We are highly efficient in processing and recognizing such biologically salient cues, so much so that it seems effortless.

This dissertation presents four studies which employ both psychophysical and brain imaging techniques to probe the neural encoding of faces and biological motion. Study 1 behaviorally demonstrates that a substantial amount of information, including face orientation, can be processed in the absence of observers' conscious awareness. Study 2 and Study 3 further examine the cortical and sub-cortical processing of facial information that take place at the subconscious level. By rendering face images invisible through interocular suppression, distinct patterns of responses are revealed in FFA, STS, and the amygdala, with STS and the amygdala being selectively sensitive to facial-expression information. Study 4 focuses on the processing of local biological motion signals. A series of experiments show that such signals are processed automatically in the visual system independent of global form and global pattern of motion, and that dorsal occipito-parietal areas are the prime neural candidate for the "life motion detector".

Together, these studies indicate that the human visual system is sensitive to biologically significant information, which can be processed without awareness. The findings add to our understandings of the brain mechanisms underlying humans' superb processing of face and biological motion information.

<p style="text-align:center">Table of Contents</p>

List of Figures

Chapter 1

Introduction: Encoding of biologically significant information in the human brain

Humans are social creatures. Both faces (who the person is and what the facial expression means) and body motion (what the person is doing) are biologically salient, and are arguably the most important sources of social information in human interactions. As such, face and biological motion perception are highly developed visual skills, and are both mediated by distributed and specialized neural systems in the human brain. This dissertation focuses on the neural encoding of such biologically significant information. Here, Chapter 1 provides a brief review of the background and our current understanding of how the human visual system processes basic visual information as well as faces and biological motion.

Visual information processing in the human brain

Human visual perception involves cortical and sub-cortical information processing, organized into feed-forward, horizontal, within-area, and feedback connections with feed-forward connections relaying information from lower to higher visual cortical areas (Lamme & Roelfsema, 2000). With recent advances in neuroscience and neuroimaging methods, detailed retinotopic and functional properties of human primary visual cortex (V1) (Tootell et al., 1998) and extrastriate visual areas (Brewer, Liu, Wade, & Wandell, 2005; DeYoe et al., 1996; Dougherty et al., 2003; S. A. Engel, Glover, & Wandell, 1997; Hadjikhani, Liu, Dale, Cavanagh, & Tootell, 1998; Sereno, McDonald, & Allman, 1994) have been examined, as shown in Figure 1-1 (Tootell, Tsao, & Vanduffel, 2003).

Specialized areas in extrastriate visual cortex

In addition, a number of category-specialized areas have been identified in extrastriate visual cortex (see Fig. 1-1). Using functional magnetic resonance imaging (fMRI), Kanwisher et al. found an area in the fusiform gyrus that was significantly more active when subjects viewed faces than when they viewed assorted common objects (Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997). This area was consequently named the Fusiform Face Area (FFA) according to its functional specificity and selectivity. Similarly, the Parahippocampal Place Area (PPA) and the Extrastriate Body Area (EBA) were identified and found to respond selectively to scenes depicting places and images of human bodies, respectively, as compared with a wide range of control stimuli (Downing, Jiang, Shuman, & Kanwisher, 2001; Epstein &

Kanwisher, 1998). In a recent fMRI study, Downing et al. scanned subjects when they viewed scenes and 19 different object categories including faces and human bodies. Consistently, they found selectivity for faces in the FFA, for scenes in the PPA, and for bodies in the EBA. In addition, there were no regions in the occipitotemporal pathway that were strongly selective for other categories. Taken together, these results demonstrate both the strong selectivity of a small number of regions, such as the FFA, PPA, and EBA, and that true category selectivity is not a widely found property in human visual cortex (Downing, Chan, Peelen, Dodds, & Kanwisher, 2006).

Figure 1-1. Human visual cortex (Tootell, Tsao, & Vanduffel, 2003).
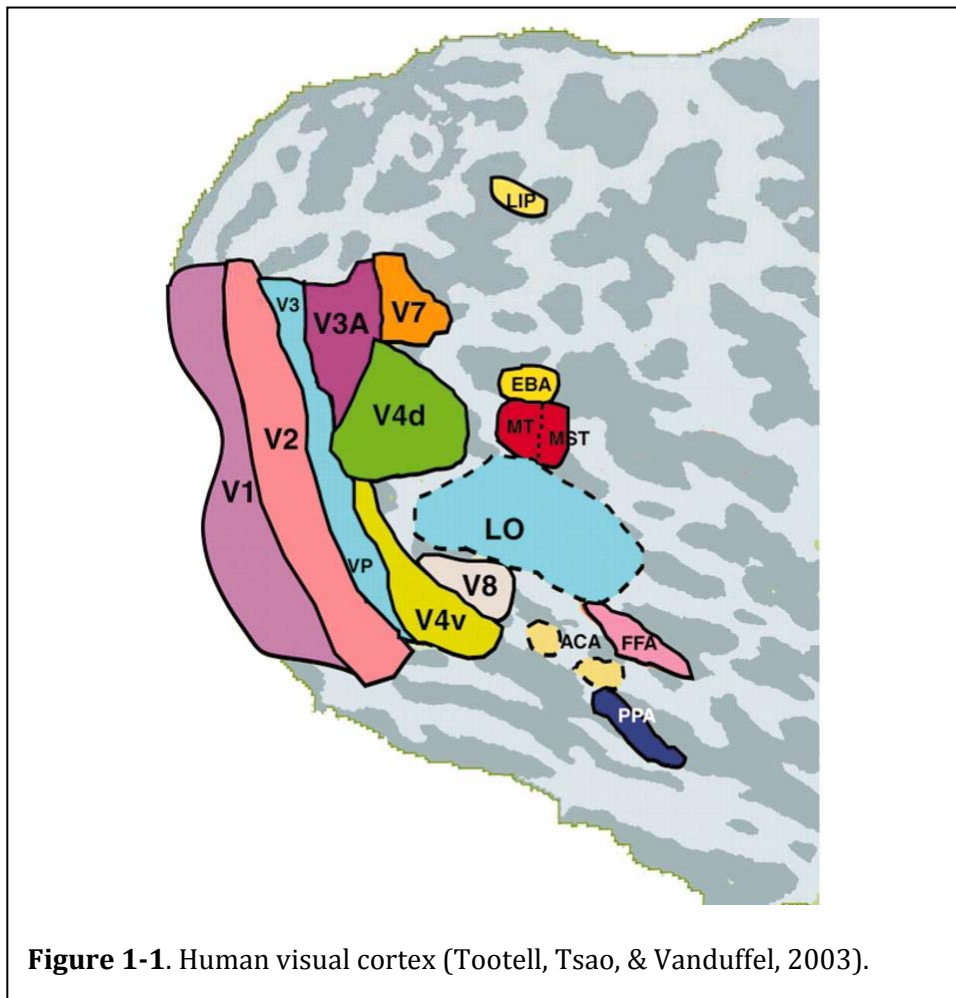
Among the specialized areas in extrastriate visual cortex, the FFA represents the best example of category selectivity. However, the field has also witnessed a long and hot debate on how to conceptualize the category selective responses observed in these regions. In one study, Gauthier et al. found that acquisition of expertise with novel objects led to increased activation in the right hemisphere face areas, and the same areas were also more activated in experts compared to novices during a passive viewing condition. Therefore, they suggested that expertise seems to be one of the factors that leads to specialization of the face area. However, more and more recent evidence supports the idea of stimulus category based, but not process based specialization in face perception (Yovel & Kanwisher, 2004).

Visual information processing in the absence of awareness

When we think of visual information processing, often we think about the processes that ultimately lead to our conscious perceptual experience, for example, seeing and recognizing shapes, objects, and people in our environment. However, it is no longer a surprise that a significant part of these processes are in fact going on "under the radar"- automatically and without subjective awareness. Some of these processes are the prelude to more conscious stages of information processing, while other processes lead to outcomes that are never meant to become a part of our conscious experience, such as much of the information used for guiding our motor action. Results obtained from patients who have suffered damage in their visual cortex have been pivotal in this area of research. Patient DF, a woman who suffered a lesion in a lateral occipital cortical region from carbon monoxide poisoning, couldn't tell you if a slot placed in front of her

4

was oriented horizontally or vertically, yet she could post a letter through that slot. DF couldn't tell you which of two chips were larger, yet when asked to pick them up, she opened her hand wider for the larger chip and narrower for the smaller chip (Goodale, Milner, Jakobson, & Carey, 1991; James, Culham, Humphrey, Milner, & Goodale, 2003). The patient reported that she was not able to perceive anything, functionally blind, yet she was remarkably good at processing many attributes (e.g., form, orientation, and size) of the stimuli presented in her blind field, a phenomenon called blindsight (Stoerig & Cowey, 1997; Weiskrantz, 1996). Although vision may appear to serve the purpose of giving us conscious visual experience, the example of blindsight shows us the crucial role of unconscious visual information processing. However, this is not only true in pathological conditions, but has also been demonstrated in normal subjects (Milner & Goodale, 1995).

Figure 1-2. Psychophysical techniques to render stimuli invisible. (a) Backward masking, (b) Visual crowding, and (c) Binocular rivalry (or interocular suppression) (Kim & Blake, 2005; Jiang & He, 2006).
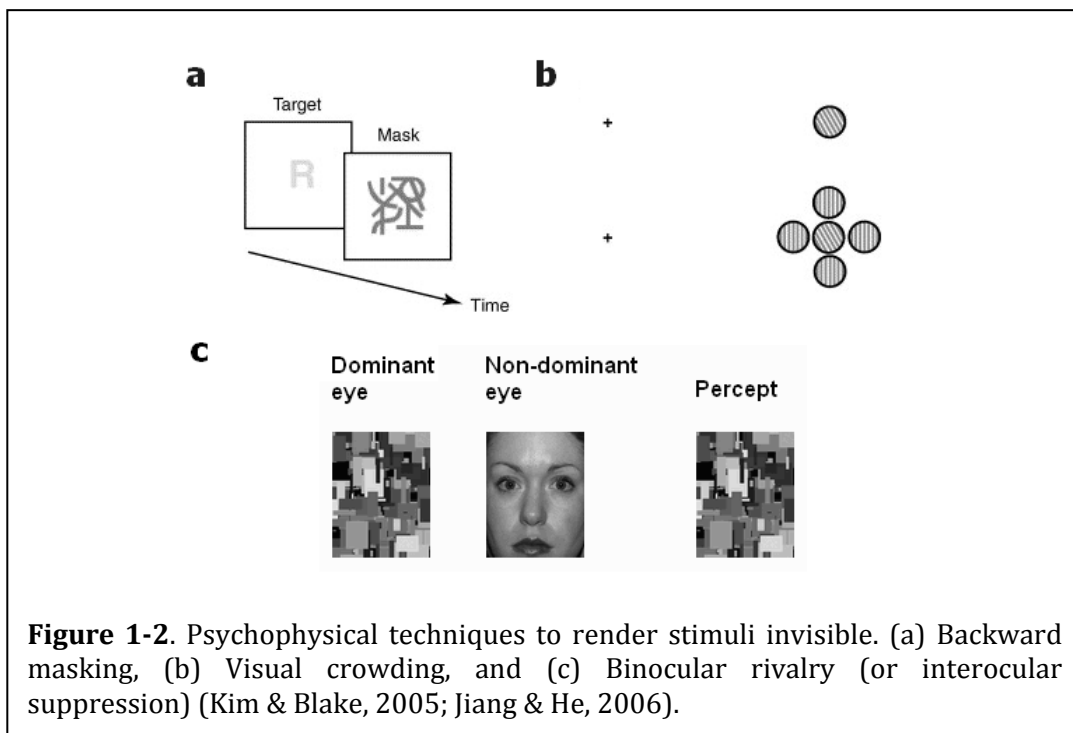
Studies with normal subjects usually examine behavioral consequences and record cortical activation when stimuli are rendered invisible. However, there are various psychophysical techniques that can be used for manipulating visual awareness. A widely used technique is visual backward masking (Bruno G. Breitmeyer, 1984). A briefly presented target stimulus is erased from visual awareness by subsequent presentation of a mask in close spatial and temporal proximity to the target (see Fig. 1-2a). Although it has not come to a consensus how the mask disrupts the processing of the target (feed-forward or feedback), visual words that are masked and cannot be seen may nevertheless facilitate the subsequent processing of related words, a phenomenon called masked priming (Marcel, 1983). It has also been shown that the color of a target masked to invisibility still exerts a color-specific influence on reaction times to the mask itself (B. G. Breitmeyer, Ro, & Singhal, 2004). Unlike masking, visual crowding (Bouma, 1970, 1973; He, Cavanagh, & Intriligator, 1997; Pelli, Palomares, & Majaj, 2004; Townsend, Taylor, & Brown, 1971) and binocular rivalry (Blake, 2001; Blake & Logothetis, 2002; Tsuchiya & Koch, 2005) work effectively for extended viewing periods. Visual crowding refers to the phenomenon that a normally visible figure may be unrecognizable when flanked by nearby stimuli, and it works primarily within the peripheral visual field (see Fig. 1-2b). Binocular rivalry is induced by presenting dissimilar monocular patterns to corresponding areas of the two eyes (see Fig. 1-2c). In such a situation, the observer's perception usually alternates between the two patterns, with one pattern in view and the other suppressed at a given time (Blake & Logothetis, 2002). One decade ago, it was shown that when a grating was blocked from awareness by crowding, it could nevertheless generate undiminished orientation selective

6

adaptation (He, Cavanagh, & Intriligator, 1996). Similarly, orientation-selective adaptation and other relatively low level visual aftereffects were demonstrated to occur when visibility of the adapting stimulus was disrupted by binocular rivalry (Blake & Fox, 1974; Lehmkuhle & Fox, 1975; Wade & Wenderoth, 1978; White, Petry, Riggs, & Miller, 1978). However, the idea of full-blown feature adaptation in the absence of awareness (either due to crowding or binocular rivalry) was recently modified (Blake, Tadin, Sobel, Raissian, & Chong, 2006), in that when non-saturating stimuli were used as adaptation stimuli, the strength of orientation and motion adaptation was indeed reduced by the removal of awareness.

There are a number of advantages of using binocular rivalry as a tool to manipulate stimulus visibility compared to other approaches (Kim & Blake, 2005). Binocular rivalry is quite versatile in its power to make a stimulus invisible, and its effect is relatively independent of specific stimulus features as well as the stimulus' retinal location. Tong and colleagues used a face and a house as rival targets to discover reciprocal activations in the FFA and the PPA coincident with perception of the face and the house (Tong, Nakayama, Vaughan, & Kanwisher, 1998). In a more recent study, Tsuchiya and Koch found that distinct Mondrian images flashed successively at approximately 10 Hz into one eye can reliably suppress an image presented to the other eye, called interocular suppression or continuous flash suppression (CFS) (Tsuchiya & Koch, 2005). The duration of perceptual suppression is at least ten times greater than that produced by binocular rivalry (Tsuchiya, Koch, Gilroy, & Blake, 2006). Fang and He used a very similar interocular suppression procedure and further reduced the contrast of the suppressed images to achieve a very long-lasting suppression effect, and

7

found that human dorsal cortical areas showed substantial responses to invisible objects, with stronger responses to tools than faces (Fang & He, 2005).

It remains unknown how much and what aspects of high level object information can be processed in the absence of awareness. Chapter 2 of this dissertation reports a psychophysical study using a novel interocular suppression paradigm. It shows that familiar and recognizable information (e.g., upright faces and recognizable words), even when suppressed and invisible, is processed differently from unfamiliar information (e.g., inverted faces and unrecognizable words). In other words, high-level visual information in the suppressed phase of binocular rivalry, including face orientation information, can be processed to the extent that object-related representations can be achieved.

Face perception: facial identity and facial expression

Perceiving facial identity and facial expression

Face perception is one of the most highly developed visual skills in humans. Infants prefer to look at faces at a very early age (Morton & Johnson, 1991). In a recent study, Sugita found that infant monkeys reared with no exposure to faces could still show a preference for human and monkey faces in photographs, and discriminated human faces as well as monkey faces. The result indicates the existence of an experience-independent ability for face processing (Sugita, 2008). More interestingly, after the monkeys were exposed first to either human or monkey faces for a month, they

selectively discriminated the exposed species of face and showed a marked difficulty in regaining the ability to discriminate the other nonexposed species of face. This plasticity of face processing can also be found in infants (Pascalis et al., 2005), suggesting an apparent sensitive period during which a broad but flexible face prototype develops into a concrete one for efficient processing of familiar faces.

Facial information plays a critical role in human social interaction. When we see a face, at least two main types of information are processed. First, a face is registered as a face and identified as belonging to a unique individual, establishing general facial category and facial identity information. Secondly, facial expressions are interpreted, which might pertain to a significant reward or an aversive event and could be crucial for social interaction (Posamentier & Abdi, 2003). The relative ease and speed with which facial identity and facial expression information are processed suggests that a highly specialized system or systems are responsible for face perception, be it domain specific (Kanwisher et al., 1997) or general (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999). With the development of functional brain imaging, the perception of faces has consistently been found to evoke activity in regions in the bilateral fusiform gyri, though more consistently on the right (Clark et al., 1996; Halgren et al., 1999; Haxby et al., 1994; Haxby et al., 1999; Hoffman & Haxby, 2000; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999; Kanwisher et al., 1997; McCarthy et al., 1997; Sergent, Ohta, & MacDonald, 1992). In addition to the face-selective fusiform region (FFA), functional imaging studies have identified other face-responsive regions, usually consistently located in the lateral inferior occipital gyrus (Occipital Face Area, or OFA) and the superior temporal sulcus (STS) (Chao, Haxby, & Martin, 1999; Halgren et al.,

1999; Haxby et al., 1999; Hoffman & Haxby, 2000; Kanwisher et al., 1997; Puce, Allison, Bentin, Gore, & McCarthy, 1998).

Bruce and Young proposed an influential model of face perception with separate functional routes for the recognition of facial identity and facial expression (Bruce & Young, 1986). More recently, Haxby and colleagues further proposed a model for the human neural system that mediates face perception with an emphasis on a distinction between the representation of invariant and changeable aspects of faces. The model has a hierarchical structure and is divided into a core system and an extended system (Fig. 1-3). Specifically, the core system is comprised of occipitotemporal regions in extrastriate visual cortex, including the inferior occipital gyrus, lateral fusiform gyrus, and STS, that mediate the visual analysis of faces. Within the core system, they suggested two functionally and neuroanatomically distinct pathways for the visual analysis of faces: one codes changeable facial properties (such as expression, lipspeech and eye gaze) and involves the STS, whereas the other codes invariant facial properties (such as identity) and involves the lateral fusiform gyrus. The extended system is comprised of regions from neural systems for other cognitive functions, such as the amygdala and orbitofrontal cortex, that can be recruited to act in concert with the regions in the core system to process the meaning of information gleaned from faces (Haxby, Hoffman, & Gobbini, 2000). These models share the idea of distinct functional modules for the visual analysis of facial identity and expression.

Figure 1-3. A model of the distributed human neural system for face perception. The model is divided into a core system, consisting of three regions of occipito-temporal visual extrastriate cortex, and an extended system, consisting of regions that are also parts of neural systems for other cognitive functions. Changeable and invariant aspects of visual facial configuration have distinct representations in the core system (Haxby, Hoffman, & Gobbini, 2000).
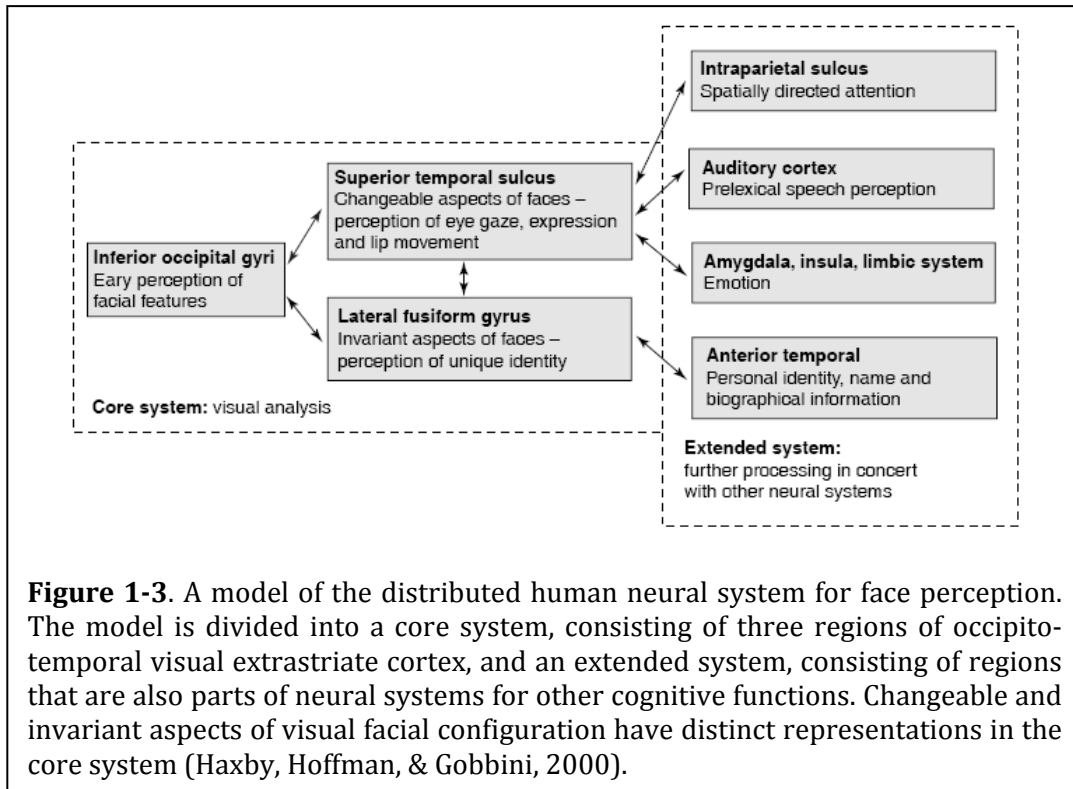
This idea is supported by behavioral studies of healthy participants and neuropsychological studies of patients with impaired face perception following brain damage. For example, the familiarity of a face does not affect the ability of a healthy participant to identify its expression and vice versa (Bruce, 1986; Calder, Young, Keane, & Dean, 2000; Campbell, Brooks, de Haan, & Roberts, 1996; Young, McWeeny, Hay, & Ellis, 1986). Humphreys et al. reported two patients: one is markedly impaired at perceiving facial identity but performs reasonably in some facial expression judgments, whereas the other patient is relatively good at perceiving facial identity but poor at judging the emotional expression of faces (Humphreys, Donnelly, & Riddoch, 1993). The findings are consistent with the existence of separate processes for encoding face identity and facial expression. Similar cases have also been shown in

many other patient studies, in which brain injury in humans can produce selective impairments in the recognition of facial identity or facial expression (Bruyer et al., 1983; Etcoff, 1984; Hornak, Rolls, & Wade, 1996; Tranel, Damasio, & Damasio, 1988; Young et al., 1995; Young, Newcombe, de Haan, Small, & Hay, 1993). Studies of nonhuman primates also support the existence of two neural systems for facial processing (Campbell, Heywood, Cowey, Regard, & Landis, 1990; Heywood & Cowey, 1992). In a neurophysiological study, neurons responsive to expression were found primarily in the cortex in the superior temporal sulcus, while neurons responsive to identity were found primarily in the inferior temporal gyrus (Hasselmo, Rolls, & Baylis, 1989). The significant difference in anatomical distribution of cell populations supports the theory of independent encoding of facial identity and facial expression in face-responsive neurons.

More recently, functional brain imaging studies have yielded less convergent findings. While some researchers found that the perception of facial identity and facial expression have different neural correlates (George et al., 1993; Narumoto, Okada, Sadato, Fukui, & Yonekura, 2001; Winston, Henson, Fine-Goulden, & Dolan, 2004), others have failed to find such a dissociation (Fairhall & Ishai, 2007; Ishai, Schmidt, & Boesiger, 2005; LaBar, Crupain, Voyvodic, & McCarthy, 2003; Vuilleumier, Armony, Driver, & Dolan, 2001). With an fMRI adaptation paradigm, it was found that repeating identity across face pairs led to a reduced fMRI signal in fusiform cortex and the posterior STS, whereas repeating emotional expression across pairs led to a reduced signal in a more anterior region of the STS (Winston et al., 2004). In addition, paying attention to the face enhanced the activity of the bilateral fusiform gyri and the right STS, whereas

selective attention to facial emotion specifically enhanced the activity of the right STS compared with attention to the face per se (Narumoto et al., 2001). These results provide neuroanatomical evidence for the distributed model of face processing and highlight the special role of the right STS in facial emotion recognition.

Although functional imaging investigations have come to a consensus that the FFA is consistently activated by facial identity information, the results have been less consistent with regard to the brain areas involved in facial expression processing (Ishai et al., 2005; LaBar et al., 2003; Vuilleumier et al., 2001). Specifically, Ishai et al. found that viewing emotional faces resulted in larger spatial extents of activation and higher amplitudes of the fMRI signal not only in the STS but also in the inferior occipital gyrus and fusiform gyrus, indicating that face perception is mediated by a distributed cortical network (Ishai et al., 2005). In an fMRI study, Vuilleumier et al. investigated attentional effects on facial emotion processing, and found that the activity in the fusiform gyrus was strongly modulated by attention and was greater for fearful than neutral faces (Vuilleumier et al., 2001). Similarly, LaBar et al. showed that the amygdala and fusiform gyrus responded greater to dynamic versus static emotional expressions. However, they also found that the activity in the STS could discriminate emotion morphs from identity morphs (LaBar et al., 2003). Based on these findings, Calder and Young argued that the separation of visual pathways for facial identity and expression is relative rather than absolute, and other possible models of facial identity and expression recognition, including one that has emerged from principal component analysis techniques, should be considered (Calder & Young, 2005).

Taken together, the functional roles of the FFA and the STS in face perception remain unclear, especially in relation to the emotional processing of faces (Ishai et al., 2005; LaBar et al., 2003; Narumoto et al., 2001; Vuilleumier et al., 2001). In a more recent study, Fairhall and Ishai used dynamic causal modeling to investigate the effective connectivity and the functional organization between and within the distributed cortical network (i.e., the core system and the extended system) for face perception. The researchers found that the core system is hierarchically organized in a predominantly feed-forward fashion, and that the fusiform gyrus exerts a dominant influence on the extended system. Moreover, emotional faces increased the coupling between the fusiform gyrus and the amygdala, whereas famous faces increased the coupling between the fusiform gyrus and the orbitofrontal cortex (Fairhall & Ishai, 2007).

Chapter 3 reports an fMRI study showing that by rendering face images invisible through interocular suppression, distinct patterns of responses in FFA, STS, and the amygdala can be better revealed, with STS and the amygdala being selectively sensitive to facial-expression information. The results support the idea that face perception involves two functionally (identity versus expression or invariant versus changeable) and anatomically (FFA, STS, and amygdala) distinct pathways, and suggest the distinct functions of FFA and STS may reflect primarily feed-forward visual-information processing.

Temporal dynamics of face perception

In addition to investigating the cortical areas involved in face perception, dynamic responses have also been studied with single unit, electroencephalography (EEG), and

14

magnetoencephalography (MEG) recordings, which may help us to gain a complete understanding of facial information processing in the brain. Sugase et al. recorded the activity of single neurons in the temporal cortex of macaque monkeys while presenting visual stimuli consisting of geometric shapes, and monkey and human faces with various expressions. They found that single neurons conveyed two different scales of facial information in their firing patterns, starting at different latencies. Global information, categorizing stimuli as monkey faces, human faces or shapes, was conveyed in the earliest part of the responses (Tovee, 1994). However, a significant portion of cells also convey fine information about facial identity or expression, but only after an additional delay following global information processing (Sugase, Yamane, Ueno, & Kawano, 1999).

Similarly, both ERP and MEG have characterized a response component called the N170 or M170 that occurs around 170 ms after stimulus onset, and has an amplitude about twice as large for faces than for a variety of control nonface stimuli (Eimer, 2000a; Sams, Hietanen, Hari, Ilmoniemi, & Lounasmaa, 1997). Although it was suggested that the N170 component reflects late stages in the structural encoding of faces (Eimer, 2000b), its selectivity and specificity for faces have also been examined, respectively. In an ERP experiment, Itier and Taylor presented subjects with upright, inverted faces, and seven object categories, and revealed an extra map found only for faces, reflecting an extra cluster of activity compared to objects. They suggested that while the N1 for objects seems to reflect the return to baseline from the P1, the N170 for faces reflects supplement activity. The electrophysiological specificity of faces could lie in the involvement of extra generators for face processing compared to objects

and the N170 for faces seems qualitatively different from the N1 for objects (Itier &

Taylor, 2004a). The selectivity of M170 for face perception was further examined in an

MEG study. Xu et al. found that cars did not elicit a higher M170 response (relative to

control objects) in car experts compared to control subjects, and the M170 amplitude

was correlated with successful face identification, but not with successful car

identification in car experts. These results indicate that the early face processing

mechanisms marked by the M170 are involved in the identification of faces in

particular, not in the identification of any objects of expertise (Xu, Liu, & Kanwisher,

2005). Moreover, Itier and Taylor performed source analyses for faces and seven object

categories at the latency of the N170. They found larger source intensities in the

posterior STS for faces compared to objects, and STS intensities were highly correlated

with N170 amplitude over both hemispheres only for faces. The results suggest that one

of the major neural sources for the face-sensitive N170 is the STS region (Itier &

Taylor, 2004b).

In addition to the N170/M170 component in facial processing, recent studies seem to

support a view of face recognition that involves at least two processing steps. In an

MEG study, Liu et al. found a face-selective MEG response occurring only 100 ms after

stimulus onset (M100), 70 ms earlier than previously reported N170/M170. Further, the

amplitude of this M100 response was correlated with successful categorization of

stimuli as faces, but not with successful recognition of individual faces, whereas the

M170 response was correlated with both processes. These results suggest that face

processing proceeds through two stages: an initial stage of face categorization, and a

later stage at which the identity of the individual face is extracted (Liu, Harris, &

Kanwisher, 2002).

Although the N170/M170 component has been usually associated with processing of facial identity (Itier & Taylor, 2004a; Liu et al., 2002; Xu et al., 2005), this component seems to be modulated by processing of facial expression as well (Batty & Taylor, 2003; Blau, Maurer, Tottenham, & McCandliss, 2007; Miyoshi, Katayama, & Morotomi, 2004), though it is still controversial (Eimer & Holmes, 2002). It has been found that the amplitude of the N170 evoked by fearful faces is larger compared to neutral faces, and source localization analyses confirmed the superior and middle-temporal regions for early processing of facial expressions (Batty & Taylor, 2003), highly similar to the face N170 itself (Blau et al., 2007). The results indicate that facial expression is processed in the early stages of face processing (Miyoshi et al., 2004). In a recent intracranial recording study, Krolak-Salmon et al. showed that specific responses to fearful faces, compared with neutral faces, were first recorded in the amygdala as early as 200 ms after stimulus onset, and then spread to occipito-temporal visual regions through a retrograde neuromodulation of the amygdala on the ventral visual system (Krolak-Salmon, Henaff, Vighetto, Bertrand, & Mauguiere, 2004; Morris, Friston et al., 1998; Pessoa, McKenna, Gutierrez, & Ungerleider, 2002).

Using an interocular suppression paradigm, Chapter 4 reports an ERP study showing that specific responses to invisible fearful faces (compared to invisible neutral and scrambled faces) is indexed by a significantly larger negative deflection starting at 220 ms along the STS. The results provide further evidence for cortical processing of facial information in the absence of awareness, and shed light on the temporal sequence of the

17

automatic extraction of facial expression information.

Biological motion perception

Perceiving biological motion

Similar to the existence of an experience-independent ability for face processing (Sugita, 2008), biological motion perception is also an intrinsic capacity of the visual system. Four-month-old infants stare at human motion sequences for longer durations than they do at the same number of dots undergoing random motions, and moreover, the preference for biological motion is not exhibited when infants view an inverted person (B. I. Bertenthal, Proffitt, & Kramer, 1987; Fox & McDaniel, 1982). In a more recent study, Simion et al. tested 2-day-old babies' discrimination after familiarization and their spontaneous preferences for biological vs. nonbiological animations. Newborns were shown to be able to discriminate between two different patterns of motion and, when first exposed to them, selectively preferred to look at the biological motion display. Similarly, newborns also looked longer at upright displays than upside-down displays (Simion, Regolin, & Bulf, 2008). This sensitivity is also evidenced in eight-month-old infants by ERP differences during the perception of biological motion vs. scrambled motion (Hirai & Hiraki, 2005).

People are remarkably adept at recognizing motions of biological entities in complex visual scenes, even when they are portrayed by a handful of point-lights attached to the head and major joints (Johansson, 1973). It has also been demonstrated that observers

can readily recognize the action (Dittrich, 1993; Norman, Payton, Long, & Hawkes, 2004), the gender (Kozlowski & Cutting, 1977, 1978; Mather & Murdoch, 1994; Sumi, 2000; Troje, 2002), and the identity information (Cutting & Kozlowski, 1977; Fani, Prasad, Harber, & Shiffrar, 2005; Troje, Westhoff, & Lavrov, 2005) conveyed by the point-light biological motion. Observers can also judge the emotional implication of an action when viewing point-light biological motion (Clarke, Bradshaw, Field, Hampson, & Rose, 2005; Dittrich, Troscianko, Lea, & Morgan, 1996; Pollick, Paterson, Bruderlin, & Sanford, 2001; Walk & Homan, 1984). In addition to body motion, facial expressions can be portrayed using point-light animation as well, and it has been shown that observers can identify the facial expression (Bassili, 1978) and even the gender of the actor (Hill, Jinno, & Johnston, 2003).

Perceiving biological motion is robust. Observers can easily recognize biological motion even when a point-light animation is embedded in an array of dynamic noise dots that far outnumber the dots defining the biological motion (B.I. Bertenthal & Pinto, 1994; Cutting, Moore, & Morrison, 1988; Ikeda, Blake, & Watanabe, 2005). Observers can also recognize human activity when a point-light animation is presented for less than one-tenth of a second (Johansson, 1973), when the dots are blurred or randomized in contrast polarity over time (Ahlstrom, Blake, & Ahlstrom, 1997), when the stereoscopic depth of the dots are scrambled and unrelated to their implied depth for the biological entity (Ahlstrom et al., 1997; Bulthoff, Bulthoff, & Sinha, 1998; Lu, Tjan, & Liu, 2006), or when the dots are placed on the intermediate positions of the joints (B.I. Bertenthal & Pinto, 1994). In sum, the human visual system seems to be especially well adapted for the perception of other people's actions (Blake & Shiffrar, 2007).

One of the most important characteristics of biological motion perception is that it is strongly impaired if the point-light displays are shown upside down (Ikeda et al., 2005; Pavlova & Sokolov, 2000; Sumi, 1984). In addition, the inversion effect seems to be relatively independent of the experimental task, as it impairs biological motion detection (B.I. Bertenthal & Pinto, 1994; Pavlova & Sokolov, 2000; Troje, 2003), direction discrimination (B.I. Bertenthal & Pinto, 1994), action recognition (Dittrich, 1993; Shipley, 2003), and even gender classification from biological motion (Barclay, Cutting, & Kozlowski, 1978). In a sense, biological motion perception resembles face perception, which is also highly susceptible to inversion (for a review, see (Valentine, 1988)). This, therefore, has led biological motion researchers to consider biological motion perception (biological motion inversion effect) a holistic process (Dittrich, 1993; Reed, Stone, Bozova, & Tanaka, 2003), similar to face perception (face inversion effect).

Additionally, the fact that the biological motion inversion effect persists in detection tasks using scrambled motion masks that render all local information useless strongly supports such a view (B.I. Bertenthal & Pinto, 1994; Shiffrar, Lichtey, & Heptulla Chatterjee, 1997), and several lines of evidence underscore the importance of visual form in the perception of biological motion. For example, when realistic photographs of a human body are sequentially presented at slow temporal rates, it is sufficient for observers to perceive the human action, even though such displays contain minimal motion information (Chatterjee, Freyd, & Shiffrar, 1996). Beintema and Lappe designed a variant of the point-light animation in which the positions of the dots were not confined to the joint but instead could appear anywhere along the limbs. Moreover, the

dots changed their positions along the limbs unpredictably from frame to frame. While the manipulations should not disrupt specification of the body form, they made it virtually impossible to perceive coherent motion of the individual dots defining an activity. Nonetheless, observers viewing these displays could judge with reasonable accuracy the walking direction (left versus right) in the biological motion. Thus the researchers suggested that biological motion is derived from dynamic form information, independent of local image motion (Beintema & Lappe, 2002). This result is consistent with a simple template-matching model in which an algorithm uses static postures monitored from walking humans as stored templates. The positions of the stimulus dots in each single frame of biological motion are then matched to these templates. With this approach, a high correlation was found between simulation results and psychophysical data, suggesting that the limited form information in point-light stimuli might be sufficient to perceive biological motion (Lange, Georg, & Lappe, 2006).

However, a more recent study by Troje and Westhoff proposed an alternative account. They showed that observers can readily retrieve information about direction from scrambled point-light displays of humans and animals. Even though all configural information is entirely disrupted, perception of these displays is still subject to a significant inversion effect (Troje & Westhoff, 2006). The observation is intriguing as it provides evidence that local biological motion information can be processed independent of global form in biological motion perception. They also found that the information about direction, as well as the associated inversion effect, is carried entirely by the local motion of the feet, consistent with previous findings, which demonstrate that for point-light walkers at least, the points defining the wrists and ankles are

crucially important when judging the direction of walking (Mather, Radford, & West, 1992), while the points defining the mid-limb joints (elbow and knees) and the torso (shoulder and hips) contribute significantly to detection of PL walkers embedded within noise (Pinto & Shiffrar, 1999). Taken together, then, these studies indicate that both global form and local motion play critical roles in the perception of biological motion.

Neurophysiological and neuropsychological evidence for biological motion perception

Similar to face perception, the visual and functional importance of biological motion perception is also instantiated in a specialized neural network. The first neurophysiological evidence for the existence of a brain area visually coding human movement came from single unit recording (Perrett et al., 1985). Recording from single cells in macaque cortex, neurons in the superior temporal sulcus (STS) are selectively responsive to human forms and motions, suggesting the role of the STS in integrating motion and form information (Oram & Perrett, 1994, 1996) converging from the dorsal and ventral visual streams, respectively (Felleman & Van Essen, 1991).

Selective visual deficits in patients with focal brain lesions also revealed a specialized neural mechanism for biological motion perception. For example, two patients with biparietal lobe lesions were identified in Schenk and Zihl's study to have difficulty in perceiving a point-light walker in noise but have no problem discriminating the direction of coherent dots in motion (Schenk & Zihl, 1997). Similarly, another patient could not recognize point-light depictions of biological motion but could detect movement of the dots and could see that not all were moving in the same direction (Cowey & Vaina, 2000). Conversely, Vaina et al. identified a patient with bilateral

lesions involving the posterior visual pathways, who had great difficulty in perceiving coherent motion in random noise, speed discrimination, and seeing two-dimensional form from relative speed of motion. However, on higher-order motion tasks the patient was able to recognize human actions from the point-light biological motion (Vaina, Lemay, Bienfang, Choi, & Nakayama, 1990). Therefore, this group of evidence points to the critical role of the parietal regions for biological motion perception, different from the classical motion-blind patients whose lesions included the visual cortical area V5 (Battelli, Cavanagh, & Thornton, 2003).

To support this view, Grossman et al. applied transcranial magnetic stimulation (TMS) to induce temporary "lesions" in normal individuals and subsequently measured sensitivity to biological motion sequences. For a short time, following TMS to a region of the scalp overlying posterior brain regions (the posterior STS), observers had more difficulty recognizing point-light biological motion sequences in noise, but only when sequences were shown in their upright orientation; sensitivity to inverted sequences, although generally poorer, was unaffected by TMS. No effect was found when TMS was applied over motion-sensitive area MT+/V5 on perception of biological motion. These findings underscore the involvement and the critical role of the STS in perception of human motion (E. D. Grossman, Battelli, & Pascual-Leone, 2005).

Specialized brain areas for biological motion perception

Recent brain imaging studies have examined the specialized brain areas potentially involved in biological motion perception. Most of the studies contrasted brain activations when observers viewed point-light biological motion sequences with

activation when viewing scrambled versions of the same motion sequences, with the latter serving as a baseline. Since these two kinds of motion sequences (intact and scrambled biological motion) comprise identical individual dot trajectories and differ only in the global spatiotemporal coherence of the dots portraying the biological motion information, the comparison between them will reveal a number of brain areas that are responsive to motion and/or form components embedded in biological motion (see Fig. 1-4). Those areas involve human motion complex (hMT+, or V5), lateral occipital complex (LOC), EBA (see (Michels, Lappe, & Vaina, 2005; Peelen, Wiggett, & Downing, 2006)), FFA (see (E. D. Grossman & Blake, 2002; Michels et al., 2005)), and OFA (see (E. D. Grossman & Blake, 2002)). More importantly, the posterior STS, particularly in the right hemisphere, has been consistently found to respond strongly to biological motion and to play a critical role in the analysis of biological motion, a trend evident in the results of most brain imaging studies (Beauchamp, Lee, Haxby, & Martin, 2003; Bonda, Petrides, Ostry, & Evans, 1996; Grezes et al., 2001; E. Grossman et al., 2000; E. D. Grossman & Blake, 2001, 2002; E. D. Grossman, Blake, & Kim, 2004; Michels et al., 2005; Santi, Servos, Vatikiotis-Bateson, Kuratate, & Munhall, 2003; Saygin, Wilson, Hagler, Bates, & Sereno, 2004; Thompson, Clarke, Stewart, & Puce, 2005; Vaina, Solomon, Chowdhury, Sinha, & Belliveau, 2001).

Figure 1-4. Regions of Interest (ROIs) involved in processing of biological related stimulus: pSTS (red), EBA (purple), LO (blue), FFA and OFA (orange) (Grossman & Blake, 2002).
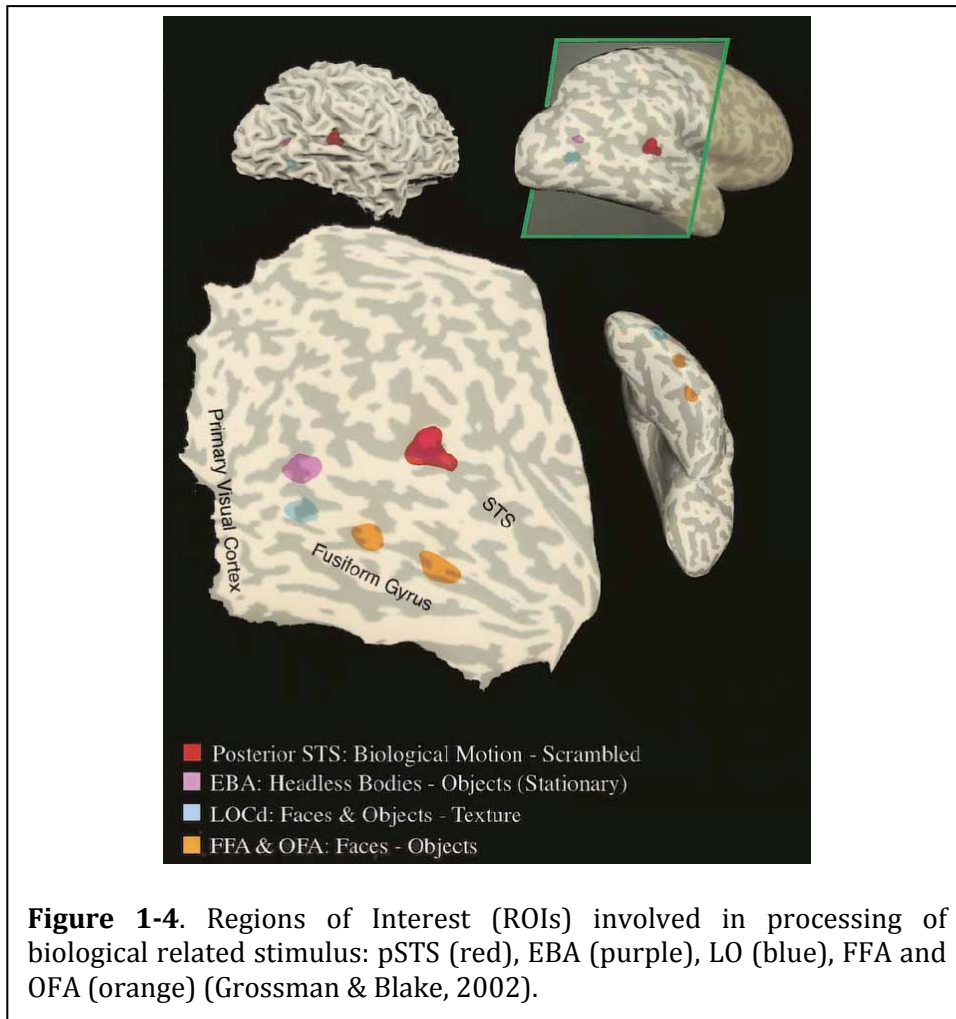
Interestingly, brain areas including the posterior STS were also found to respond robustly even when observers viewed biomechanically impossible movements of fingers (Costantini et al., 2005). Although these studies have established that biological motion activates the STS, the use of scrambled/random motion as controls does not make clear whether coordinated and meaningful nonbiological motion might also activate these same brain regions and thus call into question their specificity for processing biological motion. Pelphrey et al. addressed this issue by comparing responses in the STS with animations of human and robot walking figures and found

that the STS is specifically sensitive to biological motion, rather than complex nonbiological motion (Pelphrey et al., 2003). Moreover, Peuskens et al. compared biological motion with a three-dimensional rotation of a human figure, articulated motion and translation, and further confirmed the unique involvement of the STS in perceiving biological motion (Peuskens, Vanrie, Verfaillie, & Orban, 2005).

Despite the crucial role of the STS, the functional roles of the other brain areas within the network that respond to biological motion stimuli are less clear and their relationship to the STS remains in question. In a recent study, Peelen et al. applied pattern-analysis techniques to dissociate the overlapping functional brain areas that respond to biological motion. They found that biological motion selectivity was correlated positively, on a voxel-by-voxel basis, with body selectivity only. However, biological motion selectivity was negatively correlated with simple motion selectivity, indicating a possible dissociation between the responses in EBA and hMT+ to form and motion, respectively. They therefore concluded that biological motion engages areas involved in the analysis of the static human form through the process of structure-from-motion (Peelen et al., 2006).

Based on the results from the neurophysiological, psychophysical, and brain imaging studies on biological motion perception, Lange and Lappe proposed a computational two-stage-process model to elucidate the contributions of motion and form signals to biological motion perception and the computations in the underlying brain network. The first stage of the model analyzes only the form information in each sequential frame of the stimulus without knowledge of the temporal order and involves the ventral brain

areas, like the EBA and FFA. The second stage performs global motion analysis by explicitly analyzing the temporal order of the selected frames, and the pSTS seems necessary in order to derive and process such information (Lange & Lappe, 2006).

Chapter 5 reports a series of psychophysical and brain imaging experiments which examine the automatic processing of local biological motion information in the human brain. The results highlight the intrinsic sensitivity of the human brain to local life motion signals, and point to a specialized neural mechanism serving as a "life motion detector" that automatically processes biologically meaningful local motion signals independent of global configuration.

Chapter 2

Processing of invisible stimuli: Advantage of upright faces and recognizable words in overcoming interocular suppression

This chapter is a reproduction of Jiang, Y., Costello, P., & He, S. (2007). Processing of invisible stimuli: advantage of upright faces and recognizable words in overcoming interocular suppression. Psychological Science, 18(4), 349-355.

Familiar and recognizable stimuli enjoy an advantage of predominance during binocular rivalry, and this advantage is usually attributed to their enhanced processing during the dominant phase. However, do familiar and recognizable stimuli have an advantage in breaking suppression? Test images were gradually introduced to one eye to compete against a standard high-contrast dynamic noise pattern presented to the other eye. Results showed that an upright face took less time than an upside-down face to gain dominance against the identical suppression noise. Results also showed that for Chinese readers, Chinese characters were faster to gain dominance than Hebrew words, whereas for Hebrew readers, the reverse was true. These results suggest that familiar and recognizable information, even when suppressed and invisible, is processed differently from unfamiliar information. Apparently, high-level information about visual form does contribute to the strength of a stimulus during its suppressed phase.

INTRODUCTION

Binocular rivalry refers to the alternations in perception that occur when two different images are presented dichoptically to the two eyes. It is generally believed that rivalry results from the multiple stages of mutual inhibition between neural populations coding for the competing images features, with the neurons generating the dominant image at a given time inhibiting the neurons responding to the suppressed image. Previous studies have shown that perceptual switching occurs when the dominant signal adapts over time and eventually becomes weaker than the signal of the suppressed stimulus (Blake, 1989; Lehky, 1988; Mueller, 1990; Sugie, 1982; Wilson, Blake, & Lee, 2001). It has also been

suggested that dominance and suppression rely on distinct neural processes (Blake & Logothetis, 2002). In other words, neural processes that amplify the salience of a dominant target are not necessarily engaged during the suppression phase of rivalry.

Several studies have demonstrated that high-order organizational structures of an image influence its probability of dominance during rivalry. For example, during rivalry, an upright face generally prevails when paired with an inverted face (E. Engel, 1956). During piecewise rivalry, parts of an object can be grouped into one coherent image through global organizational processes (Kovacs, Papathomas, Yang, & Feher, 1996). It has also been found that the texture patches forming a "Dalmatian Dog" dominate during rivalry more than similar stimuli that cannot be perceptually grouped into a coherent and meaningful object (Yu & Blake, 1992). Remarkably, this effect was obtained even before observers were aware of the embedded "dog," which indicates that the greater dominance of a structured figure does not necessarily depend on actual recognition of that figure. However, all these effects can be interpreted as due to enhancement of the structurally meaningful object during its dominant phases, with or without the observers' awareness of the structure. It is difficult to infer whether the meaningful (familiar) stimuli are processed any differently from their meaningless (unfamiliar) control stimuli during the suppressed phases.

Evidence seems to suggest that the answer to this question is "no," that high-level information is not represented during the suppression phases of rivalry. For example, a number of studies have examined the adaptation and priming effects of a stimulus during rivalry. Low-level features such as stimulus orientation, spatial frequency, and

linear motion tend to show a preserved adaptation effect during rivalry (Blake & Fox, 1974; Lehmkuhle & Fox, 1975; Wade & Wenderoth, 1978), although new evidence shows that rivalry suppression can reduce the strength of low-level adaptation (Blake et al., 2006; Tsuchiya & Koch, 2005). However, object or semantic information about the suppressed stimulus is not able to generate an aftereffect (Moradi, Koch, & Shimojo, 2005) or priming effect, a result vividly summarized as "out of sight, out of mind" (Zimba & Blake, 1983). The different fates of low- and high-level information during rivalry suppression are consistent with the idea that some visual processing occurs before and other processing occurs after the neural site (or sites) of rivalry. This idea is also consistent with neurophysiological results showing that from primary visual cortex to extrastriate areas to inferior temporal cortex, neuronal responses become increasingly correlated with the alternating perception (Sheinberg & Logothetis, 1997). Neuroimaging studies paint a more complex picture, showing that brain activation is correlated with perceptual alternations from as early as V1 and lateral geniculate nucleus (Tong & Engel, 2001; Wunderlich, Schneider, & Kastner, 2005), and that this correlation carries all the way to the fusiform face and object-sensitive areas (Tong et al., 1998).

However, although some studies suggest that high-level information is not processed and represented during the suppression phase of rivalry, recent neuroimaging studies have demonstrated the contrary, at least for certain types of high-level information. For example, under interocular suppression, emotional faces generate stronger responses in the amygdala than neutral faces (Williams, Morris, McGlone, Abbott, & Mattingley, 2004) and nonface objects (Pasley, Mayes, & Schultz, 2004), and suppressed images of

tools can activate dorsal cortical areas (Fang & He, 2005). These findings suggest that considerable information, including object category information about the suppressed stimulus, is processed in cortical and subcortical structures. However, this possibility seems to be at odds with the earlier conclusion that object shape and semantic information do not seem to be extracted and represented during suppression, as they fail to generate an aftereffect or priming effect.

To directly measure the effect of higher-level information such as meaning and familiarity of stimuli during suppression, we adopted a single-trial paradigm measuring the time needed for a stimulus to break from suppression. The critical manipulation was the familiarity (upright vs. inverted face) or recognizability (words in native vs. unknown language) of the test image. Briefly, there were three key components to our approach. First, the test stimuli were competing against the same noise pattern. Second, in each trial, as soon as the observer detected the stimulus or any part of it, the trial stopped. This ensured that the key factor influencing the dependent variable (suppression duration) was operating while the stimulus remained invisible. Third, the dynamic suppression noise was presented immediately at full contrast, whereas the target stimuli were gradually ramped up. This ensured that the noise was the dominant percept at the beginning of each trial.

This approach provided several advantages over the commonly used paradigm, in which average dominance durations are measured for paired stimuli (target vs. control, e.g., an upright face against an inverted face) engaged in many cycles of rivalry. One of the advantages was that our design maximally reduced the influence of a nonexclusive

rivalry stage because the observers were asked to make a response as soon as they detected the target stimulus, whether the whole stimulus or just part of it. The second advantage was that the two conditions under comparison (e.g., upright face vs. inverted face) were not competing against each other, in which case it would be difficult if not impossible to determine whether reduced dominance duration of one object compared with the other was due to the effectiveness of the suppressed stimulus or the ineffectiveness of the suppressing stimulus (e.g., stronger upright face or weaker inverted face?). We arranged to have the two stimuli of interest compete against the same noise pattern. This ensured that the suppression times of the target and control stimuli could be interpreted more precisely.

METHODS

Subjects

Ten observers (4 male) participated in Experiment 1. Eight native Chinese speakers (4 male), 8 native English speakers (5 male), and 6 native Hebrew speakers (4 male) participated in Experiment 2. Although they are identified here as native speakers of Chinese, English, and Hebrew, the Chinese speakers and Hebrew speakers were expert readers of Chinese and Hebrew, respectively. Subjects identified as native English speakers could read neither Chinese nor Hebrew. All subjects had college degrees. They had normal or corrected-to-normal vision; their age range was from 23 to 40. Subjects

gave written informed consent in accordance with procedures and protocols approved by the human-subjects review committee of the University of Minnesota.

Stimuli and Procedure

Stimuli were generated with MATLAB and presented on a 19-in. Mitsubishi Diamond Pro monitor (1280 × 1024 at 100 Hz) using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The images presented to the two eyes were displayed side by side on the monitor and fused using a mirror stereoscope mounted on a chin rest. A frame (10.7°× 10.7°) that extended beyond the outer border of the stimulus and fixation point was presented to facilitate stable convergence of the two images. The viewing distance was 40 cm.

Figure 2-1a shows the general paradigm for the experimental condition. A central cross (0.8°× 0.8°) was always presented to each eye, serving as the fixation point. Briefly, at the beginning of each trial, a standard dynamic noise pattern was presented to one of the observer's eyes at full contrast, and then the test figure (an upright face, an inverted face, a Chinese character, or a Hebrew word) was presented to the other eye at a random location within the region corresponding to the location of the noise. The contrast of the test figure was ramped up gradually from 0 to 100% within a period of 1 s starting from the beginning of the trial and then remained constant until the observer made a button-press response to indicate the figure's location.

34

Figure 2-1. Schematic representation of the experimental paradigm. In the experimental condition (a), a test figure (e.g., an upright face, as shown) was gradually introduced to one eye to compete with a dynamic noise pattern presented to the other eye. The contrast of the test figure was linearly ramped up from 0 to 100% within a period of 1 s starting from the beginning of the trial, and then remained constant until the observer made a response to indicate the side on which the test figure appeared. In the control condition (b), a test image was presented directly on the noise background with contrast that increased gradually (at a slower rate than in the experimental condition). Subjects viewed the stimulus binocularly and responded to the appearance of the test image as soon as possible.

In Experiment 1, the test images were 10 upright faces and their inversions. Each test image subtended $2.1° \times 2.6°$ visual angle and was presented in a random position either to the left or to the right of fixation. The horizontal distance between the center of the test image and fixation ranged from 1.9° to 3.9°, and the vertical center of the test image was anywhere between 2.9° above and 2.9° below fixation. At the very beginning of each trial, observers perceived the noise patch and were unaware which side contained the test image. To measure the time it took for the test image to overcome the suppression noise and become dominant, we asked observers to press the left or the right arrow key on a standard keyboard to indicate the side of fixation on which the test image appeared. They were told that they should respond to the appearance of any part

of the test image as soon as possible and that they did not need to know the specific content of the image.

Experiment 2 was the same as Experiment 1 except that the test images were 40 Chinese characters and 40 Hebrew words (all nouns, semantically matched between Chinese and Hebrew). Because the Hebrew words had much wider horizontal extension than the Chinese characters, each stimulus was presented either directly above or directly below fixation. The vertical distance between the centers of the words or characters and fixation was fixed at 1.2°. In this experiment, the observers made "up" and "down," instead of "left" and "right," responses.

Each subject in Experiment 1 viewed a total of 120 trials, 60 with upright faces and 60 with inverted faces. Experiment 2 contained 160 trials, 80 with Chinese characters and 80 with Hebrew words. The stimuli were presented in a randomized sequence. Response times (RTs) were calculated based on correct trials only, but very few trials were excluded because accuracy was above 99% for each subject. To reject data outliers, we also excluded trials in which the RT was longer than 10 s (this value was more than 3 standard deviations away from the sample mean). We reasoned that if the test image did not overcome the suppression noise and become dominant within 10 s, then the obtained RT would be likely to reflect some unknown and uncontrolled factors. Overall, fewer than 1% of the trials were excluded from analyses.

To test whether the results obtained in the experimental (rivalry) condition could be explained simply by different recognition speeds or different detection criteria corresponding to the different types of stimuli (upright vs. inverted faces, characters vs.

words), we also ran control conditions in which the same test stimuli were blended into the dynamic noise pattern and their contrast was ramped up gradually. Figure 2-1b shows the paradigm for this condition. Observers viewed the stimuli binocularly (nonrivalry), rather than dichoptically. Their task was exactly the same as in the corresponding experimental condition. Their perceptual experience in this control condition also mimicked the rivalry situation, in which the faces or the words overcame suppression.

The control and experimental conditions were run in separate blocks. Because observers could detect the targets before the test stimuli reached full contrast in the nonrivalry condition, the time course of contrast ramping was modified so that detection time would be in the same range as the suppression time in the experimental condition. The main control conditions were performed with the contrast ramped up at a rate of 10% increment per second, but we also ran the control conditions at larger and smaller contrast increments. Results from different ramping rates were consistent with each other.

RESULTS

We studied the effects of face orientation (upright vs. inverted) and language expertise (Chinese characters vs. Hebrew words for Chinese and Hebrew speakers) on the ability of stimuli to break from noise suppression and become dominant.

Experiment 1: Upright Faces Versus Inverted Faces

In Experiment 1, we tested upright faces and their inverted versions, and we found a significant face-inversion effect: An upright face took less time than an upside-down face to gain dominance against the identical suppression noise (1.36 s vs. 1.76 s), $t(9) = 5.53$, $p < .0005$, $p_{rep} > .99$, $d = 0.92$ (see Fig. 2-2). This result implies that the suppressed face images were processed to the level where an upright face and upside-down face could be distinguished (i.e., face representation was achieved). One might argue that the difference between upright and inverted faces could have been due to disparate recognition speeds for the two distinct types of test figures as they emerged from noise. In other words, against a noise background, an upright face might be detected more easily than an inverted face, and this difference might have been responsible for the current result. The control condition, in which the same upright and inverted faces were blended into the dynamic noise pattern and their contrast was ramped up gradually, was designed to test this possibility. Observers' perceptual experience in this control condition mimicked their perceptual experience in the rivalry situation, in which the faces overcame suppression.

Figure 2-2. Suppression and detection times for upright and inverted faces in Experiment 1. The left side of the figure shows the suppression times for 10 individual observers, as well as the averaged suppression times. The two bars on the right show the detection times in the control condition, in which the face images were presented binocularly on the noise background. Asterisks indicate a significant difference between upright and inverted faces, ** $p < .001$, $p_{rep} > .985$.

Results from the control condition showed that there was no significant difference in RT between upright faces and inverted faces (1.78 s vs. 1.82 s), $t(4) = 2.03$, $p > .1$ (see Fig. 2-2). There was a significant interaction between face orientation (upright vs. inverted) and test condition (experimental vs. control), $F(1, 13) = 11.47$, $p < .005$, $p_{rep} = .97$, $\eta^2_p = .47$. This pattern of results indicates that the advantage of familiar (upright) faces was specific to the interocular competition, and was not a general advantage in detecting upright faces.

Experiment 2: Chinese Characters Versus Hebrew Words

Faces, especially upright faces, convey rich information that is important for social interactions. Humans are experts at processing facial information, and it is likely that a significant component of human expertise in recognizing faces results from adaptive pressure over the long history of evolution. For this reason, face processing may be privileged, and results of the first experiment may or may not generalize to other categories of objects. In Experiment 2, we explored whether learned familiar visual forms (i.e., words for which people have acquired expertise) also enjoy an advantage in achieving dominance. Two types of test stimuli (Chinese characters and Hebrew words) were dichoptically presented with the dynamic noise to three groups of observers (Chinese, Hebrew, and English speakers). If the result we observed in Experiment 1 was not due solely to special face-processing mechanisms, then words that are familiar and recognizable to a subject would be expected to have an advantage in overcoming suppression and gaining dominance.

Results indeed showed a significant Observer Group × Stimulus Type interaction, $F(2, 19) = 5.00$, $p < .02$, $p_{rep} = .93$, $\eta^2_p = .34$ (see Fig. 2-3): For Chinese speakers, the Chinese characters emerged from suppression sooner than the Hebrew words, whereas for Hebrew speakers, the Hebrew words were faster to gain dominance than the Chinese characters. For English speakers who knew neither Chinese nor Hebrew, the Hebrew words also became dominant faster than the Chinese characters, but the Hebrew words' advantage was larger for Hebrew speakers than for English speakers. Because the Hebrew words occupied a wider horizontal region than the Chinese characters, we believe the effect in English speakers was due to the geometric properties of the stimuli.

Figure 2-3. Mean suppression and detection times for Chinese characters and Hebrew words in the experimental (rivalry) condition (left) and control (nonrivalry) condition (right) of Experiment 2. Results are shown separately for Chinese, English, and Hebrew speakers.
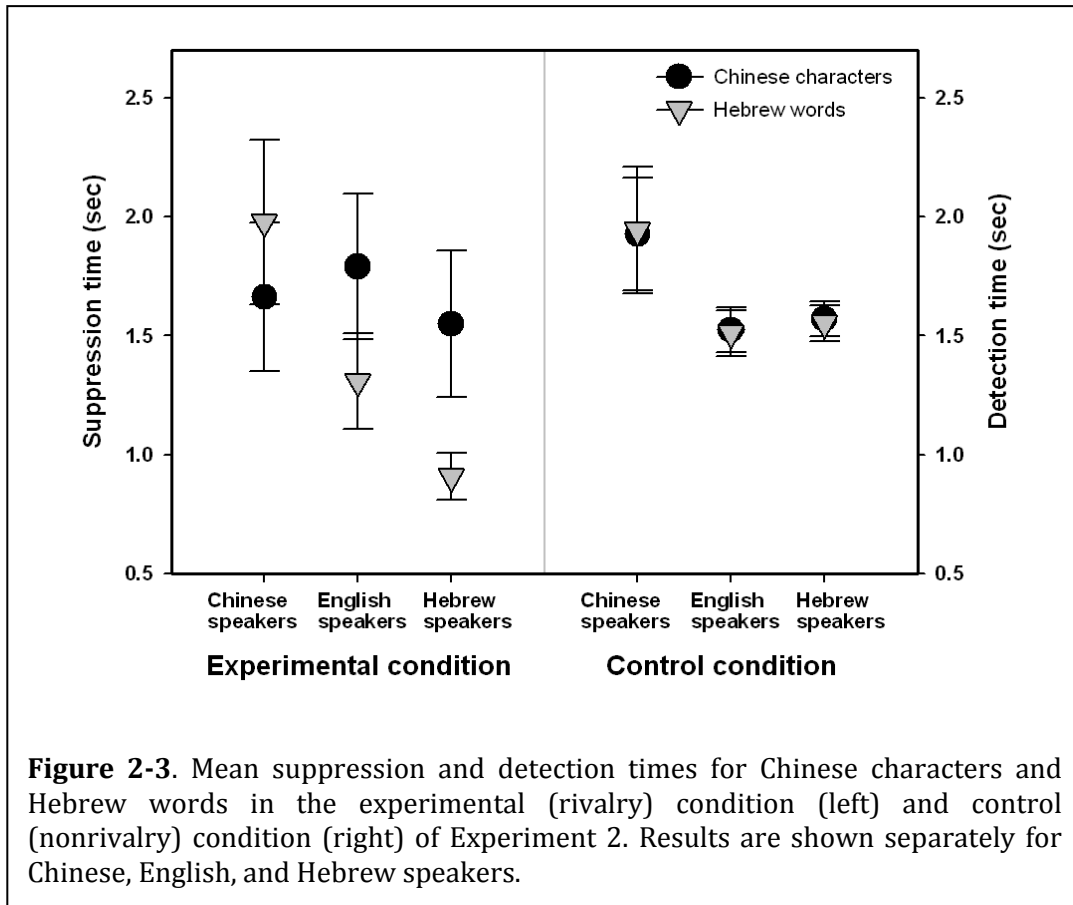
In a control condition similar to the one in Experiment 1, we tested whether these results can be explained simply by different recognition speeds or different detection criteria corresponding to the different types of test stimuli. We directly presented the same Chinese characters and Hebrew words on the noise background, gradually ramping up their rate of contrast. For each group of subjects (Chinese, English, and Hebrew speakers), the Chinese characters and Hebrew words were detected equally fast, $F(2, 9) = 0.20$, $p > .8$, although on average English and Hebrew speakers had shorter RTs than Chinese speakers (see Fig. 2-3).

Taken together with previous findings regarding binocular rivalry, our results suggest that familiar and recognizable forms, even when suppressed and invisible, are differentially processed compared with unfamiliar and unrecognizable forms and have an advantage in attaining dominance.

DISCUSSION

Early studies showed that when the strength of a stimulus is increased (e.g., by increasing its contrast), the duration of its suppression during binocular rivalry usually decreases (Levelt, 1968). Results from the current study show that changes in high-level shape and form information of the stimuli (i.e., face orientation and language familiarity) also alter their suppression time. Our first experiment shows that when upright and inverted faces are tested against a common suppressing noise, upright faces are "stronger" stimuli. As mentioned in the introduction, it would not be surprising if upright faces were stronger than inverted faces in the dominant phases of rivalry. However, our procedure specifically targeted suppression duration, and the observation that face orientation affected the stimuli's ability to break suppression suggests that at the site (or sites) of rivalry competition, upright and inverted faces are represented differently, with upright faces being stronger. Experiment 2 shows that the effect of high-level information on a stimulus's ability to break suppression is not restricted to faces. In that experiment, the familiarity and recognizability of words contributed to their strength during suppression. For Chinese observers, Chinese characters took less time to become dominant than Hebrew words, but the reverse was true for English and

42

Hebrew speakers, with Hebrew words being fastest to break suppression for Hebrew speakers.

If we took a simplistic view of our data, we might infer that face-orientation representation and visual word-form processing occur before the neural site of rivalry. A recent study suggests that the fusiform face area (FFA) is the neural correlate of the face-inversion effect (Yovel & Kanwisher, 2005), and there is evidence (though controversial) that there is a specialized visual word-form area (VWFA) in the ventral extrastriate areas (McCandliss, Cohen, & Dehaene, 2003; Price & Devlin, 2004). Do our results imply that the site for interocular suppression is later than the FFA and VWFA? The answer is not so simple. Although it is likely that interocular competition starts at V1 when the two eyes' input converges, it is believed that competition is a multistage process (Freeman, 2005; Nguyen, Freeman, & Alais, 2003). In any case, results of the current study suggest that some information from the suppressed image still reaches high-level visual areas (e.g., FFA and VWFA). Further, the information that reaches the high-level areas is strong enough to make a difference in the ability of the stimulus to overcome suppression.

Our study by itself cannot distinguish between two possibilities: First, binocular rivalry may be a process with multistage competition, with some information of the suppressed object still available at high-level stages. Second, object-related information from the suppressed image may not survive interocular competition at the cortical level, but may be processed subcortically and reach cortical object-selective areas via subcortical projections (e.g., through superior colliculus or pulvinar). Furthermore, if the

suppressed information has high social or emotional significance (e.g., is relevant to rewards and dangers), it may be able to reach cortical regions via projections through the amygdala. The amygdala responds more strongly to emotional than to neutral images during the suppression phases of rivalry (Pasley et al., 2004; Williams et al., 2004).

Regardless of how information from the suppressed image reaches object-selective regions, the key suggestion from the current results is that such information does reach those areas. In a recent imaging study using a similar interocular suppression paradigm, we showed that even when observers were completely unaware of the nature of the pictures presented, the FFA still reliably showed greater activation for invisible faces than for invisible scrambled faces (Jiang & He, 2006). These findings suggest that suppressed and invisible faces can still be represented in face-specific cortical areas. We believe such object-related representations, however degraded and rudimentary, can and do influence rivalry dynamics via feedback to the early processing stages, such that the feedback signal enhances (strengthens) the input signal associated with a coherent or familiar stimulus. It is possible that the findings reported here are unique to the specific type of interocular suppression we investigated, namely, continuous flash suppression, and cannot be generalized to more typical rivalry conditions in which two stimuli are of similar strength and alternate in dominance. However, existing evidence suggests that, if anything, suppression from continuous flash suppression is more effective than traditional rivalry suppression (Tsuchiya & Koch, 2005; Tsuchiya et al., 2006).

The current study provides strong evidence that substantial information in the suppression phase of binocular rivalry can be processed to the extent that object-related representations can be achieved, either because of incomplete suppression over the multiple stages of rivalry competition or through direct subcortical projections. These object-related representations, in turn, can strengthen the signal of suppressed images, resulting in shortened suppression durations for familiar objects.

Chapter 3

Cortical responses to invisible faces: Dissociating subsystems for facial-information processing

This chapter is a reproduction of Jiang, Y., & He, S. (2006). Cortical responses to invisible faces: dissociating subsystems for facial-information processing. Current Biology, 16(20), 2023-2029.

Perceiving faces is critical for social interaction. Evidence suggests that different neural pathways may be responsible for processing face identity and expression information. By using functional magnetic resonance imaging (fMRI), we measured brain responses when observers viewed neutral, fearful, and scrambled faces, either visible or rendered invisible through interocular suppression. The right fusiform face area (FFA), the right superior temporal sulcus (STS), and the amygdala responded strongly to visible faces. However, when face images became invisible, activity in FFA to both neutral and fearful faces was much reduced, although still measurable; activity in the STS was robust only to invisible fearful faces but not to neutral faces. Activity in the amygdala was equally strong in both the visible and invisible conditions to fearful faces but much weaker in the invisible condition for the neutral faces. In the invisible condition, amygdala activity was highly correlated with that of the STS but not with FFA. The results in the invisible condition support the existence of dissociable neural systems specialized for processing facial identity and expression information. When images are invisible, cortical responses may reflect primarily feed-forward visual-information processing and thus allow us to reveal the distinct functions of FFA and STS.

RESULTS AND DISCUSSION

Humans are social creatures. One of the most important sources of information in human interaction is the face. As such, face perception is one of the most highly developed visual skills in humans. Bruce and Young (1986) proposed an influential model of face perception with separate functional routes for the recognition of facial

identity and facial expression (Bruce & Young, 1986). More recently, Haxby and colleagues (2000) further suggested two functionally and neurologically distinct pathways for the visual analysis of faces: One codes changeable facial properties (such as expression, lipspeech, and eye gaze) and involves the superior temporal sulcus (STS), whereas the other codes invariant facial properties (such as identity) and involves the lateral fusiform gyrus (Haxby et al., 2000). These models share the idea of distinct pathways for the visual analysis of facial identity and expression. It should be noted that although evidence from behavioral and neuropsychological studies of patients with impaired face perception after brain damage and studies of nonhuman primates support the existence of two systems of facial processing (Campbell et al., 1990; Hasselmo et al., 1989; Heywood & Cowey, 1992; Humphreys et al., 1993; Tranel et al., 1988; Young et al., 1995), recent neuroimaging studies in normal human observers yield a less consistent picture because considerable overlap has been found in activation patterns in response to different face-processing tasks (Ishai et al., 2005; LaBar et al., 2003; Narumoto et al., 2001; Vuilleumier et al., 2001).

Here, we examined this issue by presenting participants with faces containing neutral and fearful expressions. These images were either presented binocularly (visible) or dichoptically with strong suppression noise and thus rendered invisible due to interocular suppression. The inclusion of faces with neutral and fearful expressions and the manipulation of stimulus-visibility states provide us with a unique opportunity to examine the distinct loci and pathways for processing invariant and changeable facial properties with and without feedback influence from a conscious representation of facial information. We hypothesize that facial identity and expression information is

initially registered and processed in relatively independent subsystems following feed-forward pathways. Neural subsystems in the brain are highly interconnected. Initially distinct facial processes interact with one another through both lateral interactions as well as feedback modulations; thus, the overall neural response in different brain areas may not reflect the relatively independent analysis for different types of facial information.

Rendering Face Images Invisible

In conventional binocular rivalry, two different images are dichoptically shown to the two eyes, and the observer's percept alternates between one image and the other. By making one of the two competing stimuli much stronger (high contrast, dynamic, and full of contours, etc.) than the other, one can make the strong stimulus dominant for much longer durations. In an extreme case, a static image in one eye can be completely suppressed for quite a long time by continuous flashing of random Mondrian images (around 10 Hz) to the corresponding location in the other eye; such is the so-called continuous flash suppression (CFS) (Tsuchiya & Koch, 2005). We adopted this CFS procedure and further reduced the contrast of the suppressed images to achieve a very long-lasting suppression effect (Fang & He, 2005).

In the invisible condition, observers viewed a composite of dynamic red Mondrain patterns and static green faces through red-green anaglyph glasses, rendering the face images invisible. In the visible condition, the face images were presented to both eyes without the Mondrian patterns (Figure 3-1). The contrast of the face images was adjusted for each individual observer to ensure that the intact and scrambled faces were

fully suppressed during the entire session. All observers underwent behavioral experiments in two separate sessions (before and after they were scanned) in which they were asked to make a Two-Alternative Forced Choice (2AFC) decision on which interval (first or second) the intact face was presented (see Experimental Procedures for details). Each observer performed at chance level in both prescan and postscan sessions. After each invisible condition scan, observers were also asked whether they saw any intact or scrambled faces or any parts of faces other than the Mondrian patterns. No observers reported seeing faces or parts of faces.

Figure 3-1. Sample stimuli and procedure used in the current study. (A) In the invisible condition, the intact face images with neutral and fearful expressions and the scrambled face images presented to the nondominant eye can be completely suppressed from awareness by dynamic Mondrian patterns presented to the dominant eye because of interocular suppression. The suppression effectiveness was verified by objective behavioral experiments. (B) The visible condition was the same as the invisible condition except that the Mondrian patterns were not presented; instead, both eyes viewed the same face or scrambled face stimuli.

Regions of Interest (ROIs)

Face-selective ROIs were defined independently with a block-design scan. Observers passively viewed the images of faces and nonface objects. Face-selective ROIs were defined as areas that responded more strongly to faces than nonface objects ($p < 104$, confirmed with a Bonferroni correction, p(Bonf) $< .05$). Three cortical areas, including the bilateral FFA and the right STS, were consistently found in all observers (Figure 3-

51

2A). Significant face-specific activation in left STS was seen in four of the six observers. Because the pattern of results in the left FFA (all six observers) and the left STS (four of the six observers) are the same as the more robust ROIs on the right side, we decided to present the results from the right FFA and right STS only.

The amygdala could be activated by emotional information even when stimuli were not perceived by observers (Morris, Ohman, & Dolan, 1998, 1999; Pasley et al., 2004; Whalen et al., 1998; Williams et al., 2004). To assess the relationship between amygdala activity and activity in FFA and STS, we also identified each observer's bilateral amygdalae with an independent scan contrasting neutral and fearful faces to scrambled faces (Figure 3-3A). Amygdala locations were very consistent across the six individual participants, with the following mean Talairach coordinates: left amygdala (-20±2, 0±1, and -14±1) and right amygdala (20±2, 1±2, and -16±2).

fMRI Responses to Visible and Invisible Face Images across ROIs

<u>The Effect of Awareness States and Facial Expressions</u>

FFA had strong activations to the visible faces, for both the neutral ($t(5)$ = 10.5 and $p <$ .0005) and the fearful expression ($t(5)$ = 15.8 and $p <$ .0005), with scrambled faces as the baseline (Figure 3-2B). FFA activations from the fearful faces were slightly but significantly stronger than activation from the neutral faces ($t(5)$ = 2.83 and p<.04). Although this enhanced response in the visible condition to the fearful faces could be due to increased attention to fearful faces compared to neutral faces, it is still consistent with recent functional imaging studies showing that the response in the fusiform gyrus

was stronger to emotional faces than neutral faces (Breiter et al., 1996; Ishai et al., 2005; Sabatinelli, Bradley, Fitzsimmons, & Lang, 2005; Vuilleumier et al., 2001). In the invisible (suppressed) condition where observers perceived the dynamic noise and were not aware of the face pictures presented, the FFA responses to faces were much reduced, consistent with previous findings (Tong et al., 1998). However, the residual responses in FFA to faces were still reliable for both invisible neutral faces ($t(5) = 3.68$ and $p < .02$) and invisible fearful faces ($t5 = 4.07$ and $p < .01$). Furthermore, no difference was found between the fMRI responses to the invisible neutral and fearful faces in FFA ($t(5) = 1.02$ and $p > .3$) (Figure 3-2C).

The STS was also significantly activated in response to the visible neutral faces ($t(5) = 2.90$ and $p < .04$) and the visible fearful faces ($t(5) = 3.25$ and $p < .03$) (Figure 3-2B). In contrast to the activation patterns in FFA, the STS showed similar fMRI signals between neutral and fearful face conditions ($t(5) = 0.84$ and $p > .4$). More surprisingly, the STS only responded to invisible fearful faces ($t(5) = 3.13$ and $p < .03$) but not to invisible neutral faces ($t(5) = 0.57$ and $p > .5$) (Figure 3-2C).

Figure 3-2. fMRI responses of face-selective areas to both visible and invisible face images. (A) Face-selective areas (FFA and STS) were identified with an independent scan and depicted on the inflated right hemisphere of a representative observer. (B) Results for the visible condition. Each panel shows the time course averaged from six observers with scrambled faces as the baseline as well as the BOLD amplitude for each individual. Results from the left hemispheres are similar to the data shown here for the right hemispheres. Both the FFA and the STS had strong activations to visible neutral (blue curves and bars) and fearful (red curves and bars) faces. (C) Results for the invisible condition. Each panel shows the time course averaged across six observers and BOLD amplitude for each individual. Even when observers were not aware of the nature of the pictures presented in this condition, the FFA still showed substantial activation for both invisible neutral and fearful faces, whereas the STS only responded to invisible fearful faces. Error bars stand for SE.

The bilateral amygdalae showed strong activation to both the visible neutral faces ($t(11) = 5.93$ and $p < .0005$) and the visible fearful faces ($t(11) = 5.40$ and $p < .0005$), with the fearful faces evoking a stronger signal than the neutral faces ($t(11) = 2.40$ and $p < .04$).

When the faces were suppressed and invisible, the bilateral amygdalae still responded significantly to the invisible neutral faces ($t(11) = 2.72$ and $p < .02$) and the invisible fearful faces ($t(11) = 6.26$ and $p < .0005$) (Figure 3-3A). The signal of the invisible fearful faces was significantly stronger than that of the invisible neutral faces ($t(11) = 6.23$ and $p < .0005$), even though observers did not consciously perceive the faces at all; such findings are consistent with recent brain imaging studies (Morris, Ohman et al., 1998; Morris et al., 1999; Pasley et al., 2004; Whalen et al., 1998; Williams et al., 2004).

The Interaction among Awareness States, ROIs, and Facial Expressions

Comparing the pattern of results across FFA and STS allowed us to test the validity of the idea that there are distinct pathways for facial-identity and facial-expression processing (Bruce & Young, 1986; Haxby et al., 2000). A significant interaction among awareness state (visible versus invisible), ROI (FFA versus STS), and facial expression (neutral versus fearful) was found ($F(1,10) = 13.2$ and $p < .005$). Further analyses revealed that the significant interaction was due to the changes in activity pattern across visibility states in STS ($F(1,5) = 88.1$ and $p < .0005$) but not FFA ($F(1,5) = 1.89$ and $p > .2$), suggesting dissociable functions of the FFA and the STS in facial processing. In addition to the interaction analysis, another way to highlight the differential sensitivity between different ROIs to the visibility of the stimuli is to quantify the suppression by a residual-activity index, RI, defined as the ratio of activity in the invisible condition compared to the visible condition, where 0 represents complete suppression and 1 represents no suppression. Specifically, there was an equal suppression of activity in

FFA for both neutral and fearful faces (RI = 26% and RI = 27%, respectively), whereas there was a contrasting reduction of activity in STS for neutral (RI ≈ 0%) and fearful faces (RI = 57%). There was a mild reduction of amygdalae responses for neutral faces (RI = 66%) but no reduction at all for fearful faces (RI ≈ 100%).

A correlational analysis between the amygdala and both the FFA and the STS showed that bilateral amygdalae activity was highly correlated with STS activity in the invisible condition only (left amygdala: r = .767 and $p < .005$; right amygdala: r = .821 and $p < .001$). However, amygdala activity was not significantly correlated with FFA activity (left amygdala: r = .359 and $p > .2$; right amygdala: r = .236 and $p > .4$) (Figure 3-3B). To rule out the possibility that the significant correlations found here were due to a general pattern of brain activity, we also analyzed the correlation between the amygdala and the primary visual cortex (V1) but could not find any significant correlations (left amygdala: r = .064 and $p > .8$; right amygdala: r = .008 and $p > .9$). It is worth emphasizing that all of the significant correlations were found in the invisible condition; there were no significant correlations in the visible condition.
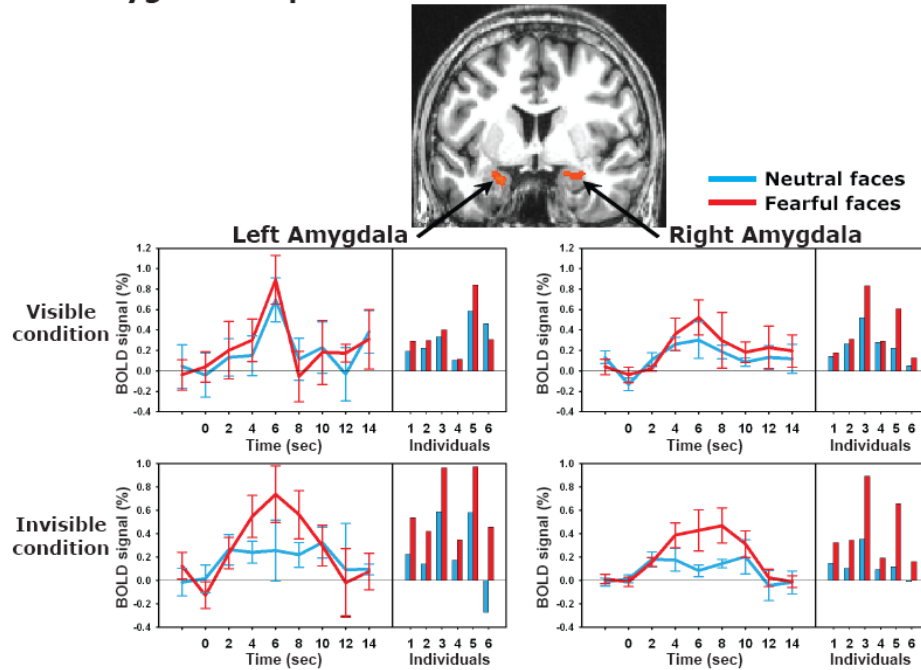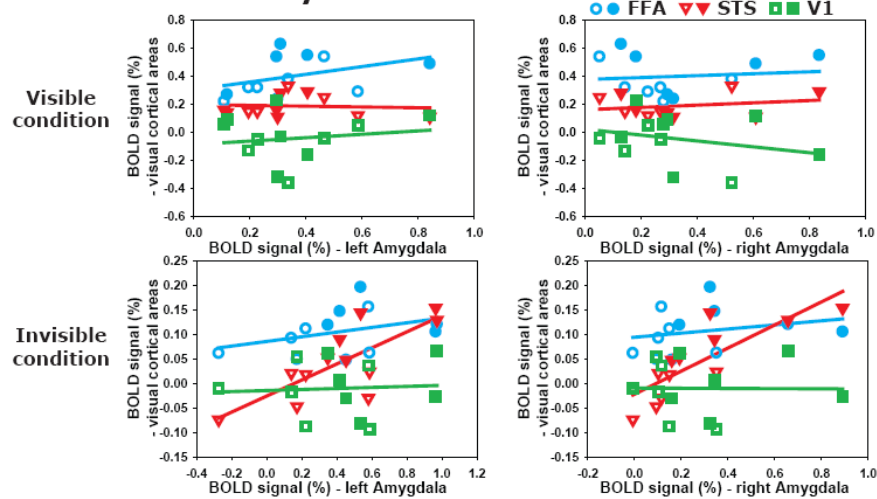
Figure 3-3. fMRI responses of bilateral amygdalae to both visible and invisible face images. (A) The bilateral amygdalae were identified with an independent scan. The average time course and individual BOLD amplitude are shown in each panel. For both visible and invisible conditions, the amygdala showed stronger responses to fearful faces (red curves and bars) than neutral faces (blue curves and bars). (B) Correlations of BOLD signals between amygdala and other visual areas (FFA: blue circles, STS: red triangles, and V1: green squares). Activities generated by the fearful faces are shown with the filled symbols, whereas activities from neutral faces are shown with the open symbols. There were highly significant correlations between activity in the amygdala and STS in the invisible condition only (red lines in the bottom panels). Error bars stand for SE.

Relationships to Other Studies

Our results show that in FFA, the fMRI signals were much reduced when faces were rendered invisible, for both the neutral and fearful faces. This result is consistent with earlier fMRI findings with similar binocular-rivalry techniques (Fang & He, 2005; Tong et al., 1998). On the other hand, FFA activity was still significantly greater for invisible faces than scrambled faces: Whereas observers could not tell whether a scrambled or a nonscrambled face was presented, their FFA reacted more strongly to the intact faces. The reliable residual activity seen in FFA when faces were invisible is consistent with Moutoussis and Zeki's discovery that invisible faces due to dichoptic color fusion could activate face-specific brain areas (Moutoussis & Zeki, 2002). On the contrary, Pasley and colleagues used binocular suppression to render stimuli invisible, but they did not find significant brain activation in the object-selective inferior temporal cortex to unperceived fearful faces compared to unperceived nonface objects (Pasley et al., 2004). Because meaningful visual images (e.g., houses) were used to suppress face images, it is possible that the relatively weak brain responses to unperceived faces could not be distinguished from the fMRI signals to houses. We recently performed a psychophysical experiment showing that upright faces took less time to gain dominance compared to upside-down faces against the identical suppression noise during binocular rivalry (Jiang, Costello, & He, 2007), and such a finding implies that the suppressed face images are processed at the level where the brain can tell the upright face from the inverted face, most likely FFA (Grill-Spector, Knouf, & Kanwisher, 2004; Yovel & Kanwisher, 2005).

58

In our view, the more surprising and novel aspect of the current finding is that STS showed a robust response to the invisible fearful faces but did not show a response to the invisible neutral faces. Contrasted with the FFA's similar responses to fearful and neutral invisible faces, this pattern of results supports a dissociation of function between FFA and STS. Such dissociation of function is highlighted by a significant interaction between ROI (the right FFA versus the right STS) and facial expression (neutral versus fearful) in the invisible condition ($F1,5 = 8.52$ and $p < .04$). Our observation provides further support for the dissociable neural systems that are specialized for facial-structural encoding versusfacial-expression analysis (Bruce & Young, 1986) or invariant versus changeable aspects of faces (Haxby et al., 2000). Why are the distinct functions for facial-information processing between FFA and STS only revealed in the invisible condition? We hypothesize that the absence of awareness of the facial information maximally reduced, if not abolished, cortical feedback modulations. Thus, the observed BOLD responses reflect more feed-forward visual-information processing without the influence of feedback modulation from a conscious representation of faces, and such an isolation of the feed-forward processmade it possible for us to reveal the distinct functions of FFA and STS by using fMRI.

A remaining question is when face images were suppressed interocularly, how did the information arrive at the face-sensitive cortical sites? It is possible that interocular suppression is incomplete at the early cortical areas and face information leaks through at the site of interocular competition. This idea is consistent with the current belief that binocular rivalry is better characterized as a multistage process, occurring at multiple cortical regions (Blake & Logothetis, 2002). Alternatively, face information could also

travel through subcortical pathways and bypass the cortical site of interocular suppression to eventually reach FFA and STS. The amygdala is believed to play a critical role in processing threat-related stimuli, particularly facial expressions of fear and anger (Adolphs, Tranel, Damasio, & Damasio, 1994; Breiter et al., 1996; Calder et al., 1996; Morris et al., 1996). It has been suggested that the amygdala receives visual inputs via both cortical and subcortical pathways with the subcortical pathway conveying crude but rapid signals before awareness and also facilitating early detection of threat (Williams et al., 2004). Furthermore, studies with both blindsight patients and normal observers support the existence of a subcortical neural pathway (colliculo-thalamo-amygdala) capable of processing fear-related stimuli independent of both the striate cortex and visual awareness (Morris, DeGelder, Weiskrantz, & Dolan, 2001; Morris, Ohman et al., 1998; Morris et al., 1999). Our study provides further support that processing of emotional information in the amygdala is, to a large degree, automatic. However, although still controversial (Vuilleumier et al., 2001; Williams, McGlone, Abbott, & Mattingley, 2005), our findings do not exclude the possibility that deprivation of attentional resources (with a very heavy load) could still modulate the amygdala's response to emotional stimuli (Pessoa, McKenna et al., 2002). The observation that STS only responded to invisible fearful faces but not neutral faces and the strong correlation between STS and amygdala activity imply a close relationship between STS and the amygdala. However, our data are consistent with communication in either direction between the amygdala and STS.

In summary, by rendering face images invisible through interocular suppression, we demonstrated distinct patterns of responses in FFA, STS, and the amygdala, with STS

and the amygdala being selectively sensitive to the facial-expression information. Our results support the idea that face perception involves two functionally (identity versus expression or invariant versus changeable) and anatomically (FFA, STS, and amygdala) distinct pathways. The current study also highlights the usefulness of the interocular suppression paradigm, which we believe is effective in isolating the feed-forward processes for close examination.

EXPERIMENTAL PROCEDURES

Participants

Six healthy observers (three male) participated in the current experiment. Observers had normal or corrected-to-normal vision and gave written, informed consent in accordance with procedures and protocols approved by the human-subjects review committee of the University of Minnesota.

Stimuli and Procedure

Stimuli were generated with MATLAB based on the NimStim face stimulus set (The Research Network on Early Experience and Brain Development) and presented through a LCD projector onto a rear projection screen located behind the participant's head inside the magnet bore. The screen was viewed with an angled mirror positioned on the head-coil. The scrambled faces were made by segmentation of the face images into 18×24 square grids and random rearrangement of the grid elements. A central point ($0.3° \times$

0.3°) was always presented to each eye and served as the fixation point. The presentation of a frame (5.7° × 7.5°) that extended beyond the outer border of the stimulus and fixation point facilitated stable convergence of the two eyes' images. The viewing distance was 102 cm.

In the fMRI scanner, observers viewed the face images (neutral, fearful, and scrambled faces) through red-green anaglyph eyeglasses. The visible and invisible conditions were run in separate scans. In the invisible condition, dynamic Mondrian patterns (10Hz) were presented to subject's dominant eye through the red filter, and the intact or the scrambled face was simultaneously presented to the other eye through the green filter for the same period. The faces and the Mondrian patterns extended 4.2° × 5.7°, and each trial was presented for 2 s followed by a 2 s intertrial interval (ITI) in which only the fixation and the outer frame were presented. The contrast of the face images was adjusted for each individual observer to ensure that the intact and scrambled faces were fully suppressed and truly invisible for the length of the experiment. After each scan of the invisible condition, observers were also asked whether they saw any faces or parts of faces other than the Mondrian patterns. Observers also performed a 2AFC test so that the effectiveness of the suppression could be checked (see below). The visible condition was exactly the same as the invisible condition except that the Mondrian patterns were replaced with the same face stimulus presented to the other eye so that observers could tell whether a neutral, fearful, or scrambled face was presented during the block.

In both conditions, neutral, fearful, and scrambled faces were presented randomly with one block consisting of 48 trials. Observers were asked to detect an occasional 3-fold

size change of the fixation point. Functional scans in the visible and the invisible conditions were run two and four times, respectively, and the order of the visible and invisible scans was randomized. The order of the three test stimulus types was also counterbalanced within each scan with M sequences (Buracas & Boynton, 2002). These are pseudorandom sequences that have the advantage of being perfectly counterbalanced among the three stimulus types so that trials from each kind of test stimulus were preceded equally often by trials for each of the other kinds ofstimuli.

fMRI Data Acquisition and Analysis

MRI data were collected on a 3T Siemens Trio outfitted with an 8-channel phase-array coil. Echoplanar data were acquired with standard parameters (28 axial slices, 3.0 mm thick; field of view, 220×220 mm^2; matrix, 64×64; repetition time, TR, 2000 ms; echo time, TE, 30 ms; and flip angle, 75°). The first four volumes were discarded to allow for magnetization equilibration. The lowest slice of the 28 functional slices was positioned just beneath the ventral surface of the temporal lobes. This ensured coverage of most of the visual and temporal cortex. A T1-weighted anatomical volume (3D MPRAGE; $1 \times 1 \times 1$ mm^3 resolution) was acquired for localization and visualization of the functional data.

After motion correction (SPM99, http://www.fil.ion.ucl.ac.uk/spm), the functional data were coregistered with the anatomical scan with BrainVoyager QX (Brain Innovation). After regions of interest (ROIs) were defined (see below), time courses from each ROI were extracted and imported into MATLAB for further analyses. For each scan, we averaged the signal intensity across the 16 trials in each condition at each of 9 time

points (from 2 s to 14 s). These event-related time courses of signal intensity were then converted to percent signal change by subtraction of the corresponding value for the scrambled face condition and then division by that value. The resulting time course for each condition was then averaged across scans and observers. The mean of three successive time points around the peak in the event-related averages served as the measured BOLD response for each condition and each individual.

Localizing Face-Specific Areas

A separate scan was used to localize face-processing areas in the occipital-temporal cortex. During the localizer scan, observers viewed alternating 20 s blocks of intact face images (both neutral and fearful faces with full contrast) and nonface object images. A boxcar response model smoothed with a hemodynamic response function was used so that voxels with a significantly higher signal ($p < 10^{-4}$, and the significance was further confirmed with Bonferroni correction for multiple comparisons, p(Bonf) < .05) could be identified in response to the face images compared to the nonface object images, and these brain areas were defined as regions of interest (ROIs). Because the amygdala is centrally involved in emotional-information processing even in the absence of awareness and is known to respond to faces (Fitzgerald, Angstadt, Jelsone, Nathan, & Phan, 2006), we further identified each observer's bilateral amygdalae with this independent face localizer scan contrasting neutral and fearful faces to scrambled faces.

Objective Measures of the Suppression Effectiveness

Because the interpretation of the study depends critically on the suppressed images being truly invisible, we also checked the suppression effectiveness in a criterion-free way. All participants underwent 2AFC experiments in separate sessions both before and after the fMRI experiment. The experimental situation (contrast, luminance, and viewing angle, etc.) was made to be fully comparable with that in the functional imaging experiments. The same set of intact face images and scrambled controls were used in this behavioral measurement. For each trial, there were two successive temporal intervals (2 s each, with a 500 ms blank gap between them). The intact face (either neutral or fearful expression) could be presented randomly in the first or the second interval, and the scrambled face was presented in the other interval. Observers pressed one of two buttons to indicate whether the face was presented in the first or the second interval. Each observer underwent 200 trials (100 trials before and 100 trials after the fMRI experiment), and all performed at chance level. To further test the possibility that the face stimuli could be fused with the Mondrian pattern during the initial brief period of stimulus presentation (Wolfe, 1983) but remain invisible due to the masking effect of the subsequent dynamic Mondrian patterns, we performed the same 2AFC experiment but with a brief presentation duration (100 ms). Observers still performed at chance level under this condition. Therefore, these 2AFC measurements confirmed that the suppressed images were truly invisible throughout the whole presentation.

Chapter 4

Dynamics of processing invisible faces in the brain: Automatic neural encoding of facial expression information

This chapter is a reproduction of Jiang, Y., Shannon, R. W., Vizueta, N., Bernat, E. M., Patrick, C. J., & He, S. (2009). Dynamics of processing invisible faces in the brain: automatic neural encoding of facial expression information. NeuroImage, 44(3), 1171-1177.

The fusiform face area (FFA) and the superior temporal sulcus (STS) are suggested to process facial identity and facial expression information respectively. We recently demonstrated a functional dissociation between the FFA and the STS as well as correlated sensitivity of the STS and the amygdala to facial expressions using an interocular suppression paradigm [Jiang, Y., He, S., 2006. Cortical responses to invisible faces: dissociating subsystems for facial-information processing. Curr. Biol. 16, 2023–2029.]. In the current event-related brain potential (ERP) study, we investigated the temporal dynamics of facial information processing. Observers viewed neutral, fearful, and scrambled face stimuli, either visibly or rendered invisible through interocular suppression. Relative to scrambled face stimuli, intact visible faces elicited larger positive P1 (110–130 ms) and larger negative N1 or N170 (160–180 ms) potentials at posterior occipital and bilateral occipito-temporal regions respectively, with the N170 amplitude significantly greater for fearful than neutral faces. Invisible intact faces generated a stronger signal than scrambled faces at 140–200 ms over posterior occipital areas whereas invisible fearful faces (compared to neutral and scrambled faces) elicited a significantly larger negative deflection starting at 220 ms along the STS. These results provide further evidence for cortical processing of facial information without awareness and elucidate the temporal sequence of automatic facial expression information extraction.

INTRODUCTION

Facial information plays a critical role in human social interaction, and face perception is one of the most highly developed visual skills in humans. When we see a face, at least two main types of information are processed. On the one hand, a face is registered as a face and identified as belonging to a unique individual, establishing general facial category and facial identity information. On the other hand, facial expressions are evaluated in terms of their potential relevance to rewarding or aversive events or their social significance (Posamentier & Abdi, 2003). The relative ease and speed with which facial identity and facial expression information are processed suggests that a highly specialized system or systems are responsible for face perception, be it domain specific (Kanwisher et al., 1997) or general (Gauthier et al., 1999).

Bruce and Young (1986) proposed an influential model of face perception in which separate functional routes were posited for the recognition of facial identity and facial expression. More recently, Haxby et al. (2000) further suggested two functionally and neuroanatomically distinct pathways for the visual analysis of faces: one codes changeable facial properties (such as expression, lip movement and eye gaze) and involves the superior temporal sulcus (STS), whereas the other codes invariant facial properties (such as identity) and involves the lateral fusiform gyrus (FG) (Haxby et al., 2000). These models share the idea of distinct functional modules for the visual analysis of facial identity and expression.

Although evidence from behavioral and neuropsychological studies of patients with impaired face perception following brain damage and studies of nonhuman primates support the existence of two neural systems for facial processing (Campbell et al., 1990;

Hasselmo et al., 1989; Heywood & Cowey, 1992; Humphreys et al., 1993; Tranel et al., 1988; Young et al., 1995), recent brain imaging studies have yielded less convergent findings; thus, the functional roles of the FG and the STS remain unclear, especially in relation to the emotional processing of faces (Fairhall & Ishai, 2007; Ishai et al., 2005; LaBar et al., 2003; Narumoto et al., 2001; Vuilleumier et al., 2001).

In a recent fMRI experiment, we examined the loci and pathways for processing invariant and changeable facial properties with and without influence from a conscious representation of facial information (Jiang & He, 2006). In the experiment, we presented participants with faces depicting neutral and fearful expressions. These images were either presented binocularly (visible) or dichoptically with strong suppression noise, and thus rendered invisible due to interocular suppression (see Fig. 3-1). With this paradigm, we found that the right fusiform face area (FFA), the right superior temporal sulcus (STS), and the amygdala responded strongly to visible faces. However, when face images were rendered invisible, activity in the FFA to both neutral and fearful faces was greatly reduced, though still measurable; activity in the STS was robust to invisible fearful faces but not to invisible neutral faces. The results in the invisible condition support the existence of dissociable neural systems specialized for processing facial identity and expression information (Jiang & He, 2006), and provide further insight into the automatic processing of facial expression information under interocular suppression (Pasley et al., 2004; Williams et al., 2004).

Interestingly, in the visible condition, the differentiation between neutral and fearful faces suggested a coupling between activity in the amygdala and the FFA rather than

69

the STS (see also Fairhall and Ishai, 2007), given that both the FFA and the amygdala showed significantly greater responses to fearful than neutral faces. However, in the invisible condition, activity in the amygdala was highly correlated with that of the STS but not the FFA (Jiang & He, 2006). This strong correlation between STS and amygdala activity, combined with the finding that the STS only responded to invisible fearful faces but not neutral faces, suggests a possible connection between the STS and the amygdala that may contribute to the processing of facial expression information even in the absence of awareness. However, due to the limited temporal resolution of fMRI, we could not further investigate the temporal dynamics of facial information processing with and without influence from a conscious representation, and we could not distinguish whether facial expression information was communicated from the amygdala to the STS or vice versa.

In the current study, we recorded event-related brain potentials (ERPs) to examine the dynamics of facial information processing (and in particular, facial expression information) while observers viewed neutral, fearful, and scrambled face stimuli, either visibly or rendered invisible through interocular suppression (see Fig. 4-1). In brief, we found that 1) in the visible condition, intact faces, compared with scrambled face stimuli, evoked significantly larger P1 and N1 (N170) components at posterior occipital and bilateral occipito-temporal regions respectively, with facial expression (fearful faces vs. neutral faces) reflected by a stronger N1 (N170) response; 2) in the invisible condition, face-related components were significantly reduced compared with the visible condition. There remained small but significant responses to faces versus scrambled faces between 140 ms and 200 ms after stimulus onset in posterior occipital

areas, and more importantly, a facial expression specific response was evident starting at 220 ms and was associated with activity in the STS.
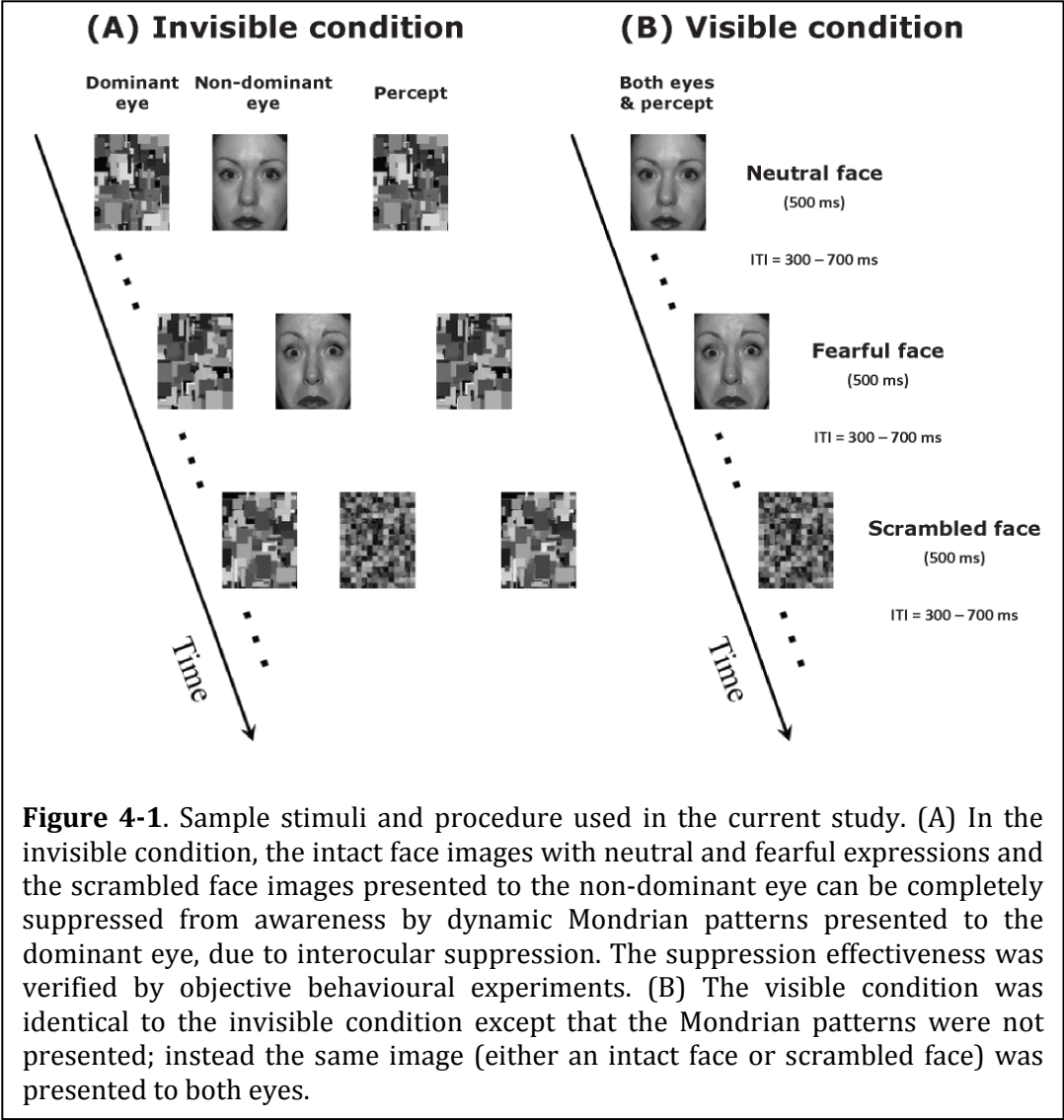
Figure 4-1. Sample stimuli and procedure used in the current study. (A) In the invisible condition, the intact face images with neutral and fearful expressions and the scrambled face images presented to the non-dominant eye can be completely suppressed from awareness by dynamic Mondrian patterns presented to the dominant eye, due to interocular suppression. The suppression effectiveness was verified by objective behavioural experiments. (B) The visible condition was identical to the invisible condition except that the Mondrian patterns were not presented; instead the same image (either an intact face or scrambled face) was presented to both eyes.

The current study thus provides further evidence for cortical processing of facial information in the absence of awareness and sheds new light on the temporal sequence of facial expression information extraction.

MATERIALS AND METHODS

Participants

Eighteen healthy observers (7 male) participated in the current ERP experiment. Observers had normal or corrected-to-normal vision and provided written, informed consent in accordance with procedures and protocols approved by the human subjects review committee of the University of Minnesota. They were naive to the purpose of the experiment.

Stimuli and Procedure

Stimuli selected from the NimStim face stimulus set (The Research Network on Early Experience and Brain Development) were configured as described below using the MATLAB software package (The MathWorks, Inc.) and displayed on a 19-in. Mitsubishi Diamond Pro monitor (1024 × 768 at 100 Hz) using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The images presented to the two eyes were displayed side-by-side on the monitor and perceptually fused using a mirror stereoscope mounted to the monitor. A central point (0.45° × 0.45°) was always presented to each eye serving as the fixation. A frame (8.1° × 10.8°) that extended beyond the outer border of the stimulus was presented to facilitate stable convergence of the two eyes' images. The viewing distance was 36.8 cm.

In conventional binocular rivalry, two different images are displayed dichoptically to the two eyes, and the observer's percept alternates between one image and the other. By continuously flashing random Mondrian images (around 10 Hz) to one eye, a static low-

contrast image in the other eye can be completely suppressed for an extended duration, known as continuous flash suppression (Tsuchiya & Koch, 2005). Pilot data showed that the spectrum power of the early ERP components to faces was around 10 Hz, thus we used 20 Hz dynamic noise in the current study to reduce the noise contribution to critical ERP components while still achieving effective interocular suppression.

Observers viewed the face images (neutral, fearful, and scrambled faces) through a mirror stereoscope. The scrambled face stimuli were constructed by segmenting the face images into 18×24 square grids and randomly rearranging the grid elements. The visible and invisible conditions were run in separate blocks. In the invisible condition, dynamic Mondrian patterns (20 Hz) were presented to the subject's dominant eye, and the face stimuli were simultaneously presented to the other eye for the same period, rendering the face images invisible. The faces and the Mondrian patterns extended $6.7°$ $\times 9.4°$ and each trial was presented for 500 ms, followed by a randomized intertrial interval (ITI) from 300 ms to 700 ms in which only the fixation and the outer frame were presented. The contrast of the face images was adjusted for each individual participant to ensure that the images were effectively suppressed and truly invisible for the length of the experiment (Jiang & He, 2006; Tsuchiya et al., 2006). The observers were asked to perform a moderately demanding fixation task (detecting a slight change in the contrast of the fixation point) throughout each run to facilitate stable fixation and to help maintain their attention on the center of the display. The fixation point changed at random time points (on average once every 1 s), independent of the face stimulus presentation. During each trial of the invisible condition, observers were also asked to report whether they saw any faces or any parts of faces other than the Mondrian patterns

by pressing another key, and these trials were discarded from further analysis. On average, 6.21% of the invisible trials were excluded. The visible condition was exactly the same as the invisible condition except that the Mondrian patterns were replaced with the same face stimulus presented to the other eye, so that observers could tell whether a neutral, fearful, or scrambled face stimulus was presented during the block.

In both conditions, neutral, fearful, and scrambled face stimuli were presented randomly with one block consisting of 144 trials. Each observer took part in 3 or 4 visible blocks and 9 or 12 invisible blocks. After the ERP session, a subset of the observers also performed a two-alternative forced-choice (2AFC) test to check the effectiveness of the suppression (see below).

ERP data recording and analysis

Electroencephalograms (EEG) were recorded from 64 scalp electrodes embedded in a NeuroScan Quik-Cap. Electrode positions included the standard 10–20 system locations and additional intermediate positions (Fig. 4-2). Four bipolar facial electrodes, positioned on the outer canthi of each eye and in the inferior and superior areas of each orbit, monitored horizontal and vertical EOG (HEOG and VEOG), respectively. The skin resistance of each electrode was adjusted to less than 5 kΩ. EEG was continuously recorded at a rate of 1000 Hz using CPz as an online reference. The signal was amplified using Synamps amplifiers and band-pass filtered online at 0.05–200 Hz. We epoched and averaged the EEG signal 200 ms prior to and 1000 ms after stimulus onset, and the pre-stimulus interval was used for baseline correction. Prior to averaging, epochs were screened for eye blinks and other artifacts. Epochs contaminated by eye

blinks, eye movements, or muscle potentials exceeding ± 50 µV at any electrode were excluded from averaging. Overall, less than 10% of the epochs were excluded from analyses. Averaged ERPs were baseline corrected by subtracting the averaged amplitude of the baseline and were re-referenced offline using a mean reference.

Both peak amplitudes and latencies were measured to examine effects associated with face type and awareness state conditions. Peak latencies were measured relative to stimulus onset. ERP peak amplitudes and latencies for face stimuli were subjected to repeated measures ANOVAs with face type and awareness state as within-subjects factors. To isolate the electrophysiological activity related to the processes of face detection and facial expression discrimination, difference waves were obtained by subtracting ERPs to the scrambled face stimuli from ERPs to the neutral face stimuli, and by subtracting ERPs to the neutral face stimuli from ERPs to the fearful face stimuli, respectively.

Objective measures of the suppression effectiveness

Since the interpretation of the study critically depends on the suppressed images being truly invisible, we also checked the suppression effectiveness in a criterion-free way. Six of the participants completed a two-alternative forced-choice (2AFC) task following the ERP experiment. The experimental context (contrast, luminance, viewing angle, etc.) was made to be fully comparable with that in the ERP experiment. The same set of intact and scrambled face images were used in this behavioral measurement. Each trial consisted of two successive temporal intervals (500 ms each, with a 500-ms blank gap between them). The intact face image (either a neutral or fearful expression) could be

presented randomly in the first or the second interval, and the scrambled face was presented in the other interval. Observers pressed one of two buttons to indicate whether the intact face was presented in the first or the second interval. Each observer underwent 100 trials, and all performed at chance level (50%). To further test the possibility that the face stimuli could be fused with the Mondrian pattern during the initial brief period of stimulus presentation (Wolfe, 1983), but remain invisible due to the masking effect of the subsequent dynamic Mondrian patterns, we performed the same 2AFC experiment but with a brief presentation duration (100 ms). Observers still performed at chance level under this condition. Therefore, these 2AFC measurements confirmed that the suppressed images were truly invisible during the presentation.

RESULTS AND DISCUSSION

ERP responses to visible and invisible face images

ERP responses to face stimuli were characterized by an early positive component (P1, peaking at 110–130 ms after stimulus onset) over the posterior occipital cortex (Fig. 4-2, lower center), followed by a negative wave (N1 or N170, peaking at 160–180 ms after stimulus onset) over bilateral temporal regions (Fig. 4-2, upper right and upper left). When face images were rendered invisible, the face stimuli elicited similar positive (P1) and negative (N1) components as those in the visible condition (Fig. 4-2). However, amplitudes of the ERP components were significantly reduced for the invisible fearful and the neutral face presentations compared with the visible condition (P1: $F(1, 17) = 9.56$, $p < 0.007$; N1: $F(1, 17) = 14.1$, $p < 0.002$). It should also be noted

that although the peak latency of the P1 component did not differ significantly for the visible versus the invisible conditions (120 ms vs. 126 ms), the peak latency of the negative ERP component (N1) was significantly delayed for the invisible compared to the visible condition (225 ms vs. 173 ms, $F(1, 17) = 96.2$, $p < 0.001$).

Figure 4-2. Depiction of the 64-channel scalp electrode montage and the grand average ERPs evoked by visible and invisible face images. Relative to scrambled faces, neutral and fearful faces elicited both larger P1 and larger N1 (N170) components at posterior occipital (waveforms averaged from electrodes OZ, POZ, O1, and O2 were shown in the lower center) and bilateral temporal regions (waveforms averaged from electrodes P6/P5, P8/P7, TP8/TP7, and T8/T7 were shown in the upper right and upper left), respectively, with the amplitude of the N1 stronger for fearful than neutral faces. However, when face images were rendered invisible, processing of face detection (intact faces vs. scrambled faces) was reflected only in activity at 140–200 ms over posterior occipital areas (lower center, highlighted in blue) while fearful faces (compared to neutral and scrambled faces) produced a significantly larger negative deflection starting at 220 ms (N1) along superior temporal regions (upper right, highlighted in blue). The small insets show the voltage topographies for the P1 and N1 components.

Differentiating intact faces from scrambled faces

To better isolate the electrophysiological activity related to the processing of face detection, difference waves were obtained by subtracting ERPs to the scrambled face stimuli from ERPs to the neutral face stimuli. Difference waves and voltage topographies of each difference-wave component are plotted in Fig. 4-3A. As can be seen from Fig. 4-2 and Fig. 4-3A, within the visible condition both the neutral and the fearful faces (relative to the scrambled face stimuli) elicited a larger P1 component (Pd1, neutral faces: $F(1, 17) = 15.9$, $p < 0.001$; fearful faces: $F(1, 17) = 15.1$, $p < 0.001$) and a larger N1 (N170) component (Nd1, neutral faces: $F(1, 17) = 24.6$, $p < 0.001$; fearful faces: $F(1, 17) = 38.0$, $p < 0.001$). Interestingly, when face images were rendered invisible, differential brain responses to intact faces versus scrambled faces were reflected only in positive activity (140–200 ms) over posterior occipital sites following the peak of the P1 component (Pd1, neutral faces: $F(1, 17) = 13.5$, $p < 0.002$; fearful faces: $F(1, 17) = 10.1$, $p < 0.006$). In contrast with the visible condition, the neutral and scrambled face stimuli were not distinguished from each other in the N1 component ($F(1, 17) = 0.09$, $p > 0.7$) in the invisible condition.

Processing of facial expression

To better isolate the brain response related to facial expression processing, difference waves were obtained by subtracting ERPs to the neutral face stimuli from ERPs to the fearful face stimuli (Fig. 4-3B). While in the visible condition the fearful faces were not differentiated from the neutral faces in the P1 component ($F(1, 17) = 0.005$, $p > 0.9$), the fearful faces evoked a significantly stronger response than the neutral faces in the N1 (N170) component (Nd1, $F(1, 17) = 7.16$, $p < 0.02$). Following the N1 component,

there was no difference between the neutral and the fearful faces ($F(1, 17) = 1.42$, $p >$ 0.2). When face images were rendered invisible, the affective component of facial expression (fearful faces vs. neutral faces) yielded a significantly larger negative deflection (N1) along lateral superior temporal areas commencing 220 ms after stimulus onset (Nd1, $F(1, 17) = 7.64$, $p < 0.02$). The differentiation of the N1 (Nd1) component between face detection (Fig. 4-3A) and facial expression processing (Fig. 4-3B) in the visible and invisible conditions is consistent with the pattern of our fMRI results for the STS, which evidenced comparable responses to the neutral and fearful faces in the visible condition, but responses to the fearful faces only in the invisible condition. The fact that this response pattern was evident only for the electrodes close to the lateral superior temporal regions supports the idea that the N1 effect originated from the STS.

However, it is also possible that the ERP amplitude differences in distinctive face processes (face detection and facial expression) in the invisible condition might be due to increased eye movements associated with neutral face stimuli compared with scrambled face stimuli (Pd1 in Fig. 4-3A), or increased eye movements for fearful face stimuli than for neutral face stimuli (Nd1 in Fig. 4-3B). To test this possibility, we extracted data from HEOG and VEOG channels and calculated the difference waves for both face detection and facial expression processing. As illustrated in Fig. 4-3 (bottom panels), the HEOG and VEOG waveforms did not show a significant effect at or before the latencies of the P1 (Pd1) and N1 (Nd1) components ($p > 0.2$). These results suggest that the observed effects of face detection and facial expression processing under interocular suppression are not simply due to increased eye movements for neutral or

fearful face stimuli, but indeed reflect cortical processing of invisible facial information.

GENERAL DISCUSSION

In the current study, we investigated the temporal dynamics of facial information extraction. Recent ERP and MEG experiments have shown that face specific processing can occur as early as 170 ms (N170/M170 component), associated with face detection and categorization (Anaki, Zion-Golumbic, & Bentin, 2007; Bentin, Allison, Puce, Perez, & McCarthy, 1996; Itier & Taylor, 2004a), and/or emotional expression recognition (Blau et al., 2007; Miyoshi et al., 2004). The N170/M170 component has been demonstrated to be a face-selective response in the human brain, though its function, characteristics, and source origins are still controversial (Batty & Taylor, 2003; Harris & Nakayama, 2007, 2008; Itier, Herdman, George, Cheyne, & Taylor, 2006; Itier & Taylor, 2004a, 2004b; Sagiv & Bentin, 2001; Xu et al., 2005).

Consistent with previous studies, we found that both neutral and fearful faces evoked strong P1 and N1 (N170) responses compared with scrambled face stimuli. Furthermore, the N1 (N170) amplitude, which is associated with activity in bilateral temporal regions, was significantly larger for fearful than neutral faces. By rendering face images invisible through interocular suppression, our data show that intact faces were differentiated from scrambled face stimuli between 140 ms and 200 ms (P1) over posterior occipital regions, whereas encoding of facial expression (fearful faces vs. neutral faces) started just after 220 ms (N1) and could be observed in the superior temporal sites (STS) only. Although it is unclear whether the ERP component sensitive

80

to intact face versus scrambled face stimuli reflects early visual processing of general objects or a face-specific response (Itier & Taylor, 2004a), the facial expression related activity in the STS clearly supports cortical processing of facial information in the absence of awareness. Indeed, it has been shown that when upright and inverted faces were paired with identical suppression noise in a modified binocular rivalry paradigm, upright faces took less time to gain dominance compared to inverted faces (Jiang et al., 2007). This observation was replicated and extended to emotional faces (i.e., fearful faces became dominant faster than neutral faces) (Yang, Zald, & Blake, 2007). These findings suggest that suppressed face images are processed to the level where the brain can tell an upright (or emotional) face from an inverted (or neutral) face, suggesting that facial information (both face detection and facial expression) can indeed be processed to some extent in the absence of explicit awareness.

How did information arrive at the STS when face images were suppressed interocularly? It is possible that interocular suppression is incomplete at early cortical areas and facial information leaks through at the site of interocular competition. This idea is consistent with the current view that binocular rivalry may be a process involving multistage competition (Blake & Logothetis, 2002; Freeman, 2005; Nguyen et al., 2003; Tong et al., 1998) and that the Magnocellular pathway may be less vulnerable to rivalry competition (He, Carlson, & Chen, 2004). To investigate this possibility, we extracted the data in the OFA from our previous neuroimaging study (Jiang & He, 2006), since the OFA is believed to be the first stage in face processing models (Calder & Young, 2005; Haxby et al., 2000). Compared with the visible condition, interocular suppression eliminated activity to faces in the OFA whereas

amygdala activity was robust in both the visible and invisible conditions (see Fig. 3-3). The fact that the OFA did not show a significant response to the invisible faces indeed casts some doubts on this cortical-leaking possibility. However, it is still possible that activity in the OFA to invisible faces reflects an early visual process on a time course that is too short to be revealed by fMRI.

Alternatively, invisible facial information may travel through sub-cortical pathways to the amygdala and bypass the cortical site of interocular suppression to eventually reach the STS. Indeed, it has been reported that during suppressed periods of binocular rivalry, emotional faces still generate a greater response in the amygdala relative to neutral faces and non-face objects (Pasley et al., 2004; Williams et al., 2004), and our previous fMRI results also showed that amygdala activity was less dependent on the conscious states of face perception (Jiang & He, 2006). Moreover, a recent intracranial recording study found that specific responses to fearful faces, compared with neutral faces, were first recorded in the amygdala as early as 200 ms after stimulus onset, and then spread to occipito-temporal visual regions including the STS (Krolak-Salmon et al., 2004). Collectively, our current data in conjunction with the strong correlation between activity in the STS and the amygdala found in our fMRI study (Jiang & He, 2006) suggests that the signal related to fearful faces in the STS was modulated by the amygdala (Krolak-Salmon et al., 2004; Morris, Friston et al., 1998; Pessoa, Kastner, & Ungerleider, 2002).

Figure 4-3. Difference waves and voltage topographies for distinctive face processes. (A) The difference-wave components for early face detection (neutral faces vs. scrambled faces) were significant at Pd1 and Nd1 components at posterior occipital and bilateral temporal regions in the visible condition whereas only Pd1 was significant in the invisible condition. (B) The difference-wave components specific to the neural encoding of facial expression (fearful faces vs. neutral faces) were indexed by an Nd1 component at lateral superior temporal sites in both the visible and invisible conditions. Moreover, the difference waves obtained from HEOG and VEOG in the invisible condition (bottom panels) show that there was no significant effect of eye movements at or before the latencies of the Pd1 and Nd1 components.

In summary, the current results provide further evidence for cortical processing of facial information in the absence of awareness and suggest that different neural subsystems for encoding facial information can be better revealed when faces are rendered invisible to observers, possibly due to the reduced modulatory effects from conscious representation. We suggest that cortical and sub-cortical visual pathways jointly contribute to facial information processing, conveying the fine detail and the crude emotion information of faces, respectively.

Chapter 5

Detecting life motion: Automatic processing of local biological motion information

This chapter is a reproduction of Jiang, Y., Zhang, Y., & He, S. (submitted). Detecting

life motion: Automatic processing of local biological motion information.

Biological motion consists of meaningful global form and global motion pattern information, as well as unique dynamics of local motion which could also provide potentially important cues to the existence of living entities around us. Most previous studies have examined the global aspects of biological motion, but the functional significance and neural mechanism of local motion processing is less understood. In the current study, local motion components of upright and inverted point-light walkers were isolated by spatially scrambling the positions of individual dots. Naïve observers could not distinguish whether a set of seemingly random moving dots came from an upright or an inverted walker, yet their spatial attention was automatically attracted to the upright local motion signals, suggesting that the dynamics of local motion contains important cues for life motion. Furthermore, initially suppressed scrambled moving dots from an upright walker had an advantage over those from an inverted walker in interocular competition against an identical suppression noise, suggesting that local biological motion information can be automatically extracted in the absence of awareness. A companion fMRI experiment revealed that the human dorsal occipitoparietal regions responded more strongly to scrambled upright than scrambled inverted point-lights, suggesting that these regions are sensitive to local dynamic properties of biological motion. Together these results highlight the functional importance and intrinsic sensitivity of the human brain to local life motion signals, and suggest a specialized brain mechanism serving as a "life motion detector" to automatically process biologically meaningful local motion signals independent of any global configuration.

INTRODUCTION

People are remarkably adept at recognizing the motion of biological entities in complex visual scenes, even when they are portrayed by a handful of point-lights attached to the head and major joints (Johansson, 1973). It has also been demonstrated that observers can readily recognize the action (Dittrich, 1993; Norman et al., 2004), gender (Kozlowski & Cutting, 1977, 1978; Mather & Murdoch, 1994; Sumi, 2000; Troje, 2002), and identity information (Cutting & Kozlowski, 1977; Fani et al., 2005; Troje et al., 2005) conveyed by point-light biological motion. Moreover, this perceptual ability is not restricted to human patterns but can be extended to point-light animations depicting the motions of other legged animals (Mather & West, 1993). However, perception of biological motion is strongly impaired if the point-light displays are shown upside down (Ikeda et al., 2005; Pavlova & Sokolov, 2000; Proffitt & Bertenthal, 1990). In addition, the global inversion effect seems to be relatively independent of the experimental task, as it impairs biological motion detection (B.I. Bertenthal & Pinto, 1994; Pavlova & Sokolov, 2000; Troje, 2003), direction discrimination (B.I. Bertenthal & Pinto, 1994), action recognition (Dittrich, 1993; Shipley, 2003), and even gender classification from biological motion (Barclay et al., 1978). It has been shown that 2-day-old human newborns also prefer upright biological motion over inverted biological motion, suggesting that there might be a intrinsic sensitivity to biological motion in the visual system (Simion et al., 2008).

In addition to the global components, the local motion components of biological motion have special properties as well, in that each individual point moves with a specific

spatiotemporal pattern (Shipley, 2003) and in a characteristic trajectory (Bidet-Ildei, Orliaguet, Sokolov, & Pavlova, 2006). However, the processing of local motion information pertaining to biological motion has received little attention. Recent studies by Troje and colleagues showed that observers can retrieve information about the walking direction of point-light displays of humans and animals when all of the points are spatially scrambled and the global configural information is removed (Troje & Westhoff, 2006), and the discrimination seems to primarily rely on the motion of the feet (Chang & Troje, 2009; Saunders, Suchan, & Troje, in press) (see also (Mather et al., 1992)) and is correlated with perception of animacy (Chang & Troje, 2008). This observation is intriguing and suggests that local biological motion independent of global configuration contains specific information. However, what are the functional significance of and the neural mechanism for processing local biological motion information in the brain? Is there a specialized brain mechanism serving as a potential "life motion detector" such that it informs the organism about the presence of life motion before and independent of the recognition of whole biological entity? Indeed, for such a mechanism to be functionally important, the detection of "life motion signal" should have meaningful consequences (Johnson, 2006). One possible outcome, we hypothesize, is that detecting the presence of life motion, prior to the recognition of life form, should lead the observer to direct attention to the site to enhance information processing from there. Moreover, for the life motion detector to function as an alerting mechanism, it should be able to process local biological motion signals automatically and even independent of awareness. This also leads to the prediction that the processing of local and global components in biological motion perception may be

supported by different neural mechanisms, and potentially in distinct regions of the brain.

To address these questions and test these hypotheses, we performed a series of psychophysical and neuroimaging experiments and examined the functional properties as well as the neural correlate of processing local biological motion information. The key comparison in the current study is the information processing of local moving dots scrambled from upright walkers vs. those scrambled from inverted walkers (Fig. 5-1a). Presumably, the dynamics of randomly positioned moving dots from an upright walker is consistent with the local motion trajectories of key parts of the human figure, but the motion of dots scrambled from an inverted walker is not consistent with motion of any normal component of the human body. In other words, human observers would have extensive experience with and exposure to the upright local motion signals, but relatively little experience with the inverted local motion trajectories. In a psychophysical experiment, we used a modified version of the Posner cueing paradigm (Jiang, Costello, Fang, Huang, & He, 2006; Posner, Snyder, & Davidson, 1980) to investigate whether the presence of local biological motion signals in the visual field could direct observers' spatial attention. Further, we adopted an interocular suppression paradigm (Jiang et al., 2007) to investigate whether local biological motion information (i.e., upright vs. inverted), independent of global form and global motion information, could be automatically processed in the absence of awareness. In a companion functional magnetic resonance imaging (fMRI) experiment, we further examined the neural correlate of processing local biological motion components in the brain.

METHODS

Participants

Twelve and eight observers participated in the psychophysical experiments (Exp. 1 and Exp. 2), respectively. Another six observers participated in the fMRI experiment (Exp. 3), and three of them also completed the fMRI control experiment. Observers had normal or corrected-to-normal vision and gave written, informed consent in accordance with procedures and protocols approved by the human subjects review committee of the University of Minnesota. They were naïve to the purpose of the experiments.

Stimuli and Procedure

Stimuli were generated and displayed using MATLAB (Mathworks, Inc.) together with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Point-light biological motion sequences were created by videotaping a walking actor. The segments were digitized and the head and joint positions in each frame were encoded as motion vectors with initial starting positions (Vanrie & Verfaillie, 2004). Scrambled biological motion sequences were created by randomizing the starting positions of each point within the region approximately covered by the intact biological motion sequences. In the scrambled biological motion sequences, the local motion components remained unchanged and only the global form and global pattern information was entirely removed. Inverted biological motion sequences were created by simply turning all of the motion sequences upside-down, including both the intact and the scrambled versions. The visual stimuli used in the psychophysical and brain imaging experiments

90

were 8 intact and 8 scrambled biological motion sequences (with different viewpoints from left to right) and their inverted counterparts (Fig. 5-1a). Linear motion sequences were used in the fMRI experiment as a control to localize the biological motion selective ROIs (see below for more detail). It should be noted that scrambled biological motion sequences were all randomly generated for each trial and for each subject in all psychophysical and brain imaging experiments in order to avoid any potential confounding factors from spatial scrambling.

In the psychophysical experiments (Exp. 1 and Exp. 2), visual stimuli were presented on a 19" Mitsubishi Diamond Pro monitor (1280×1024 at 100 Hz), and the viewing distance was 40 cm. In Exp. 1, each trial began with fixation on a central cross (0.8°×0.8°) within a frame (24. 5°×24. 5°) that extended beyond the outer border of the stimulus. Observers viewed a pair of intact or scrambled biological motion sequences, one was upright and the other was inverted (Fig. 5-1b). In the scrambled biological motion condition, observers could not distinguish which was upright and which was inverted. Each motion sequence subtended approximately 4.0°×6.8° in visual angle and was displayed for 500 ms, and the horizontal distance between the centers of this pair of motion stimuli was 10.0°. After stimulus presentation, there was a 100 ms inter-stimulus interval (ISI) in which only the fixation was displayed, followed by a small Gabor patch (2.5°×2.5°, 4.8 cpd) that was presented briefly (100 ms) as a probe in the position that either the upright or inverted biological motion sequence previously occupied. The Gabor patch was tilted one degree clockwise or counter-clockwise, and observers were required to press one of two buttons to indicate their perceived orientation of the Gabor patch regardless of the side of presentation. The scrambled and

91

intact conditions were run in separate blocks, and there were 72 trials for the scrambled condition and 72 trials for the intact condition. The Gabor probe was randomly presented in the position of the preceding upright biological motion in half of the trials. Presentation of the upright biological motion to the left or the right side of the central fixation was also balanced and randomized.

In Exp. 2, a central cross (0.8°×0.8°) was always presented to each eye serving as the fixation, and the two eyes' stimuli were displayed side-by-side on the monitor and fused using a mirror stereoscope mounted on a chinrest. A frame (12. 4°×12. 4°) that extended beyond the outer border of the stimulus was presented to facilitate stable convergence of the two eyes' images. At the beginning of each trial, a standard high contrast dynamic noise pattern was presented to the observer's one eye, and a test motion sequence was then introduced to the other eye. Each test motion sequence subtended approximately 1.6°×2.7° in visual angle and the contrast was smoothly ramped up from 0 to 100% within a period of 1.5 s starting from the beginning of each trial, and then remained constant until the observer made a response. The test motion sequence was presented in a random position either to the right or left side of fixation. The horizontal distance between the center of the test stimulus and the fixation ranged from 2.0° to 3.7°, and the vertically centered test stimulus was anywhere between 2.5° above or below fixation. With this arrangement, observers would always perceive the noise patch at the beginning of each trial and were unaware of which side contained the test motion sequence. To measure the time it took for the test motion sequence to overcome the suppression noise and gain dominance, observers were asked to make a response by pressing the left or the right arrow key on a standard keyboard to indicate on which side

of fixation the test motion sequence appeared. The observers were instructed to respond to the appearance of any part of the test motion stimulus as soon as possible without the need to know what type of stimulus was presented. For each participant in Exp. 2 there were a total of 160 trials for upright and inverted biological motion sequences and their spatially scrambled counterparts, with 40 trials for each condition. The stimuli were presented in a randomized sequence, and the assignment of the test stimuli and suppression noise to the two eyes was also randomized for each trial. Response times were calculated based on correct trials only.

In the fMRI experiment (Exp. 3), visual stimuli were presented through a LCD projector onto a rear projection screen located behind the participant's head inside the magnet bore. The screen was viewed with an angled mirror positioned on the head-coil. The motion sequences were exactly the same as those used in the psychophysical experiments, and another control motion sequence (linear motion) was added in the fMRI experiment to localize the biological motion selective ROIs. A red central cross (0.4°×0.4°) was always presented serving as the fixation. The viewing distance was 102 cm.

Each test motion sequence subtended 4.6°×7.8°. For each condition, stimuli were presented in a block design comprised of a 12 s probe (each motion sequence was presented for 1 s with an intertrial interval of 0.5 s) followed by 12 s of static random-dot controls (12 displays with 1 s for each) serving as baseline. The size and number of the static control dots were the same as those in the motion sequence to avoid any salient change between blocks. The 24 s cycle time was chosen to minimize the effect

of interference between the undershoot at the end of one hemodynamic response with the start of the following one. A total of 15 blocks (with the order of each condition truly randomized) were presented in each run, and an additional 8 s of static random-dot controls was added to the start of each scan to minimize transient magnetic saturation effects. Thus, the total scan time for each run was 368 s. In each block of stimulus presentation, all motion sequences were randomly shifted vertically and horizontally within a range of 1.2° around the central fixation to avoid local adaptation. The participants were asked to fixate at the central cross and do a one-back task to detect whether the current motion sequence was exactly the same as the previous one. To test if the activation in the visual dorsal areas in response to biological motion stimuli was due to the specific task, we conducted another fMRI control experiment with the same setup except that the observers were asked to judge whether the center-of-mass of the point-lights was presented above or below the fixation regardless of the type of biological motion.

fMRI data acquisition and analysis

MRI data were collected on a 3T Siemens Trio outfitted with an 8-channel phase-array coil at the Center for Magnetic Resonance Research (CMRR), University of Minnesota. Echoplanar data were acquired with standard parameters (28 axial slices, 3.0 mm thick; field of view, 220 × 220 mm2; matrix, 64 × 64; repetition time, TR, 2000 ms; echo time, TE, 30 ms; flip angle, 75°). The first four volumes were discarded to allow for magnetization equilibration. A T1 weighted anatomical volume (3D MPRAGE; 1 × 1 ×1 mm3 resolution) was acquired for localization and visualization of the functional data.

After motion correction (SPM99, http://www.fil.ion.ucl.ac.uk/spm), the functional data were coregistered with the anatomical scan with BrainVoyager QX (Brain Innovation). After regions of interest (ROIs) were defined (see below), time courses from each ROI were extracted and imported into MATLAB (Mathworks, Inc.) for further analyses. For each scan, we averaged the signal intensity across blocks for each condition at each of 12 time points (from -2 s to 20 s). These time courses for each condition were then averaged across scans. The BOLD signal amplitude was calculated using average signals between 8 and 12 s in each time course (condition) for each subject, as the signal usually took 6–8 s to rise to full magnitude.

Regions of interest (ROIs) identification

Each observer was first scanned with a standard retinotopic mapping procedure viewing a rotating wedge and an expanding ring (S. A. Engel et al., 1997). This allowed us to identify the borders of different visual areas (V1, V2, V3/VP, V3A, and V4) on a computationally flattened brain. FFA, EBA/FBA, and LO were defined independently with a block-design scan (localizer scan). Observers passively viewed the images of faces, human bodies, objects, and their scrambled controls. The face-selective ROI (FFA) was defined as the area that responded more strongly to faces than non-face objects; the body-selective ROIs (EBA/FBA) were defined as the areas that responded more strongly to human bodies than objects; the object-selective ROI (LO) was defined as the area that responded more strongly to intact objects than scrambled objects (all p values < 10-4, and the significance was further confirmed with Bonferroni correction for multiple comparisons, p(Bonf) < 0.05). Three cortical areas were consistently found

in all observers, and the FFA and EBA/FBA are also more robust in the right hemisphere than the left. Because the pattern of results in the left hemisphere is the same as the more robust ROIs on the right side, we decided to present the results from the right hemisphere only. The motion-selective ROI (hMT+) was identified independently with another block-design scan interleaved between motion stimuli and static controls. Since some of the adjacent ROIs could have overlapped activation voxels, ROIs only included those voxels that were exclusively selective for one category.

Contrasting between intact upright biological motion sequences and the linear motion controls showed multiple brain areas that overlapped with the ROIs defined with the localizer scans (Fig. 5-4). In addition to EBA, hMT+, and LO, biological motion sequences also evoked strong fMRI responses relative to linear motion sequences in posterior superior temporal sulcus (pSTS), consistent with previous studies (Grezes et al., 2001; E. Grossman et al., 2000; E. D. Grossman & Blake, 2001, 2002; E. D. Grossman et al., 2004; Peuskens et al., 2005; Santi et al., 2003; Saygin et al., 2004; Thompson et al., 2005; Vaina et al., 2001). Since the FFA and FBA didn't show consistent responses to biological motion sequences, we decided to restrict our further analysis to the rest of the ROIs. Moreover, the visual dorsal pathway including V3A and posterior intraparietal sulcus (pIPS) showed stronger responses to linear motion than biological motion, and these areas were also identified as ROIs. For simplicity purposes, we later refer to these regions as V3A/pIPS.

Objective measures of biological motion orientation discrimination

Following Exp. 1a, an objective Two-Alternative Forced Choice (2AFC) task was used to measure biological motion orientation discrimination in a criterion-free way. The stimuli and procedure in this 2AFC experiment (Exp. 1b) was exactly the same as those in Exp. 1a except that, after each stimulus presentation, observers were asked to make a forced choice response as to which side the upright biological motion stimulus was presented instead of judging the orientation of the Gabor patch. Seven, ten, or thirteen of the thirteen point-lights were randomly chosen and spatially scrambled in each trial, and all observers performed at chance level (50%) when all of the thirteen point-lights were scrambled.

All participants completed another 2AFC experiment after both the psychophysical (Exp. 2) and fMRI experiments (Exp. 3) as well. The same set of upright and inverted scrambled biological motion sequences were used in this 2AFC measurement. Each trial consisted of two successive temporal intervals (1 s each, with a 500-ms blank gap between them). The upright scrambled biological motion sequence could be presented randomly in the first or second interval, while the inverted scrambled biological motion sequence was presented in the other interval. Observers pressed one of two buttons to indicate whether the upright biological motion was presented in the first or second interval, and all participants performed at chance level when all of the point-lights were spatially scrambled (Exp. 2: accuracy = 52%, $t8 = 0.67$, $p > 0.5$; Exp. 3: accuracy = 48%, $t6 = -0.27$, $p > 0.7$). Therefore, the 2AFC measurement confirmed that observers could not explicitly recognize the local biological motion components when global configuration was entirely removed.

RESULTS AND DISCUSSION

Psychophysical and neuroimaging experiments were performed to investigate the functional significance (Exp. 1), automaticity (Exp. 2), and neural sites (Exp. 3) of processing local motion components in biological motion.
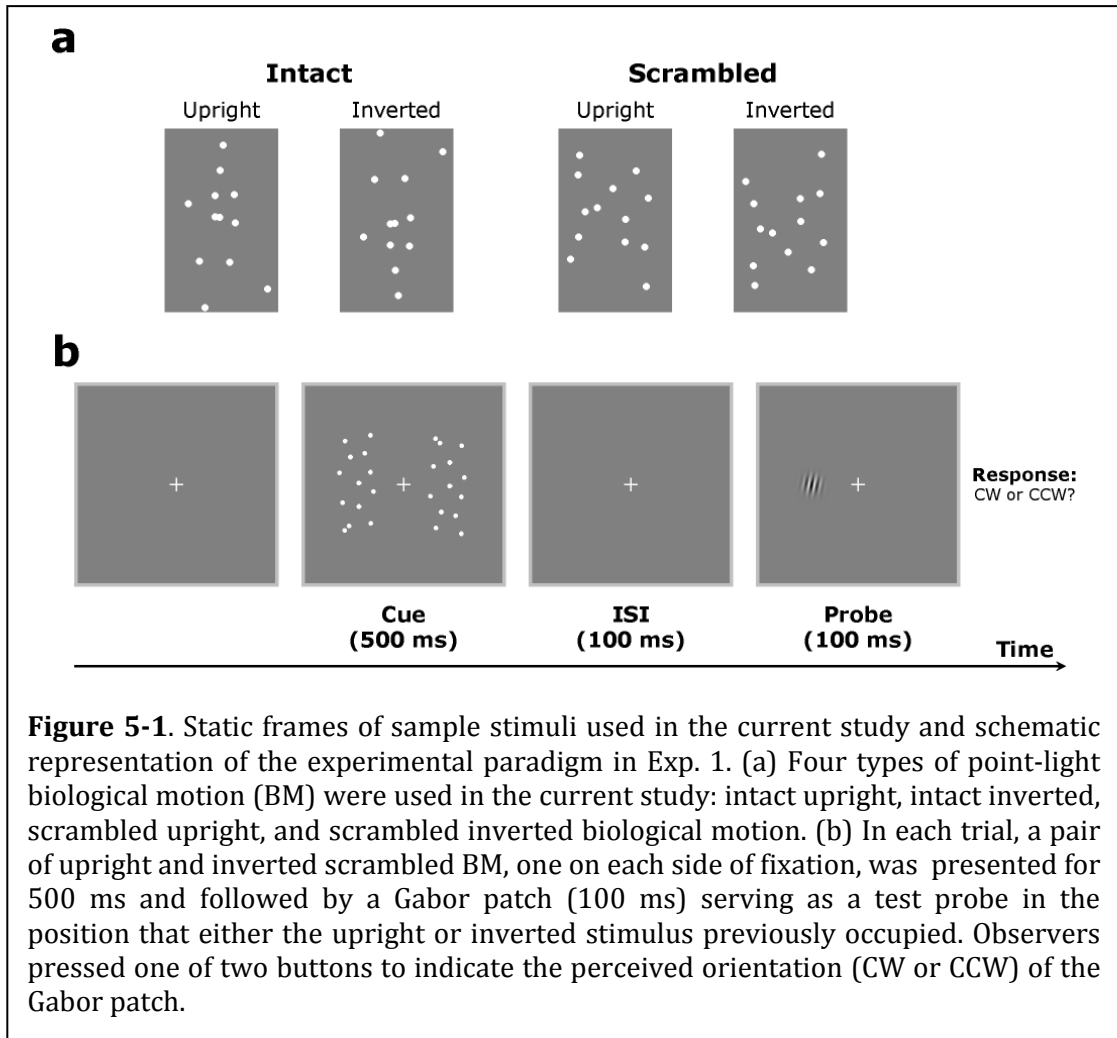
Experiment 1. Bottom-up attention capture from local biological motion: Life motion detector?

Experiment 1a: Upright local biological motion attracts attention

Salient events in a visual scene can attract visual spatial attention and subsequently enhance information processing at the attended location (Eriksen & Hoffman, 1973; Posner & Cohen, 1984; Posner et al., 1980). Presumably, humans are sensitive to the presence of life motion around us due to long-term evolutionary processes (e.g., predator and prey), and a potential life motion cue in the visual field should automatically attract our attention. However, in order for a "cue" to attract visual spatial attention, the "cue" needs to stand out from other distractors in the scene (Lang, Bradley, & Cuthbert, 1997). Given that none of the naïve observers could distinguish whether a set of scrambled moving dots came from an upright or inverted walker (see Exp. 1b for an objective measure of scrambled biological motion discrimination), it

98

provides us a good opportunity to examine whether local biological motion information can be extracted automatically and direct observers' attention.

To assess whether spatial attention can be influenced by the presence of local biological motion, two scrambled motion sequences, one from the upright walker and the other from the inverted walker, were briefly presented (500 ms) on the left and right sides of fixation respectively, followed by a Gabor patch (100 ms) that served as a test probe in the position that either the upright or inverted scrambled biological motion previously occupied. Observers had to press one of two keys to indicate the orientation (tilted 1 deg clockwise or counter-clockwise) of the briefly presented Gabor patch (see Fig. 5-1b and Experimental Procedures for details). If the performance on the Gabor orientation discrimination task depended on whether the Gabor patch was presented on the upright or inverted biological motion side, it would indicate that the preceding local motion signals had affected the distribution of observers' spatial attention. Thus the attentional effect is indexed by the accuracy difference between the two conditions: the Gabor probe presented on the side of the preceding upright scrambled biological motion vs. the Gabor probe presented on the side of the preceding inverted scrambled motion control.

Figure 5-1. Static frames of sample stimuli used in the current study and schematic representation of the experimental paradigm in Exp. 1. (a) Four types of point-light biological motion (BM) were used in the current study: intact upright, intact inverted, scrambled upright, and scrambled inverted biological motion. (b) In each trial, a pair of upright and inverted scrambled BM, one on each side of fixation, was presented for 500 ms and followed by a Gabor patch (100 ms) serving as a test probe in the position that either the upright or inverted stimulus previously occupied. Observers pressed one of two buttons to indicate the perceived orientation (CW or CCW) of the Gabor patch.
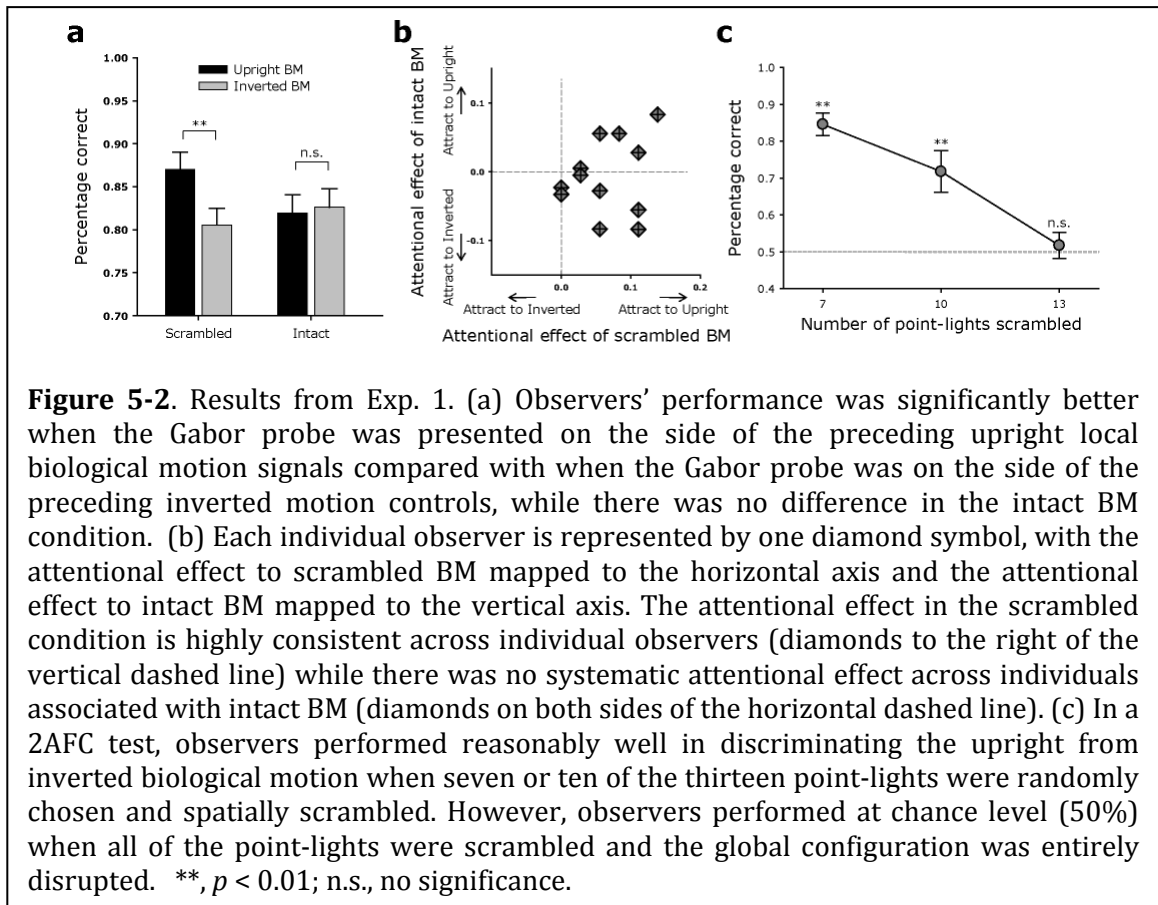
Results show that observers' performance was significantly better when the Gabor probe was presented on the side of the preceding upright local biological motion signals compared with when the Gabor probe was on the side of the preceding inverted motion controls (87.0% vs. 80.6%; $t(11) = 4.84$, $p < 0.001$; Fig. 5-2a). Furthermore, this pattern of results was highly consistent across individual observers (see Fig. 5-2b for individual data). In other words, observers' spatial attention was more attracted to the upright local biological motion signals even though observers could not tell which set of moving dots were scrambled from an upright walker and which were from an inverted walker,

100

suggesting a specialized brain mechanism that serves to automatically detect life motion signals (Chang & Troje, 2008; Troje & Westhoff, 2006). With the same paradigm, we also tested if there is a similar attentional effect associated with intact upright vs. intact inverted biological motion stimuli. Somewhat surprisingly, across observers, attention was not consistently biased by either intact upright or intact inverted biological motion (81.9% vs. 82.6%; $t(11) = -0.44$, $p > 0.6$; Fig. 5-2a). Therefore, the significant interaction between biological motion configuration (intact vs. scrambled) and biological motion orientation (upright vs. inverted) suggests that the presence of global form and global motion information disrupted the attentional effect arising from local motion signals ($F(1, 11) = 15.5$, $p < 0.002$), consistent with the psychophysical experiment described below (Exp. 2).

Experiment 1b: Biological motion orientation discrimination

With both upright and inverted local biological motion signals presented in the visual field, observers' spatial attention was attracted more to the upright motion signal while still oblivious to which one was upright and which one was inverted (Exp. 1a). When all point-lights were spatially scrambled, observers' inability to discriminate the upright vs. inverted biological motion was confirmed by an objective Two-Alternative Forced Choice (2AFC) task. In this 2AFC experiment, everything was exactly the same as Exp. 1a, but after each stimulus presentation, observers were asked to make a forced choice indicating which side the upright motion signal appeared instead of discriminating the Gabor orientation. In addition, the number of the scrambled point-lights was also manipulated (see Experimental Procedures for details). We found that all of the

observers could perform reasonably well in discriminating the upright from the inverted

biological motion sequences when seven or ten of the thirteen point-lights were

randomly chosen and spatially scrambled (with an average accuracy of 85% and 72%

respectively, $p < 0.01$). However, none of the observers could perform above chance

level (50%) when all 13 point-lights were spatially scrambled (the condition used in

Exp. 1a), removing all of the global form and global motion information (Fig. 5-2c,

$t(11) = 0.70$, $p > 0.5$). The lack of explicit knowledge about the local motion's

uprightness provides an interesting and sharp contrast to the results reported in Exp. 1a

that the attentional system was able to differentiate the upright local motion signal (i.e.,

life motion signal) from the inverted control motion signal.

Figure 5-2. Results from Exp. 1. (a) Observers' performance was significantly better when the Gabor probe was presented on the side of the preceding upright local biological motion signals compared with when the Gabor probe was on the side of the preceding inverted motion controls, while there was no difference in the intact BM condition. (b) Each individual observer is represented by one diamond symbol, with the attentional effect to scrambled BM mapped to the horizontal axis and the attentional effect to intact BM mapped to the vertical axis. The attentional effect in the scrambled condition is highly consistent across individual observers (diamonds to the right of the vertical dashed line) while there was no systematic attentional effect across individuals associated with intact BM (diamonds on both sides of the horizontal dashed line). (c) In a 2AFC test, observers performed reasonably well in discriminating the upright from inverted biological motion when seven or ten of the thirteen point-lights were randomly chosen and spatially scrambled. However, observers performed at chance level (50%) when all of the point-lights were scrambled and the global configuration was entirely disrupted. **, $p < 0.01$; n.s., no significance.

Experiment 2. Local biological motion information can be processed automatically

Observers' spatial attention was attracted automatically to local biological motion signals without their explicit knowledge about whether the motion was from an upright or inverted walker (Exp. 1). These results suggest a specialized brain mechanism that may serve as an automatic "life motion detector", which presumably functions to inform the rest of the brain about the presence of life motion signals (Chang & Troje, 2008; Troje & Westhoff, 2006). Thus it makes sense for the life motion detector to operate automatically. But how automatic is this "detector"? Can it distinguish life motion signals from non-life motion signals in the absence of awareness? To answer this question, we adopted an interocular suppression paradigm which was successful in revealing automatic processing of face orientation and word familiarity (Jiang et al., 2007), as well as facial expression (Yang et al., 2007).

Figure 5-3a shows the general paradigm for the experimental condition. At the beginning of each trial, a pair of identical dynamic noise patterns was presented to one eye at full contrast and the test stimulus (e.g., an intact upright biological motion sequence as shown here) was presented to the other eye within the corresponding region covered by the noise, but the exact location of the test stimulus was randomized for each trial. The contrast of the test stimulus was ramped up gradually over the period of one and one-half seconds. The critical manipulations were biological motion orientation (upright vs. inverted) and global configuration information (intact vs. scrambled), and the key measure was the time for the stimulus to break from noise suppression and begin to gain dominance. Observers were asked to detect the location of the test motion

sequence and press either a "left" or "right" button to indicate whether the test stimulus appeared to the left or right of fixation. They were asked to respond to the initial appearance of the test stimulus (whole or part) as soon as possible, and were not required to know what type of stimulus was presented. This ensured that the key factor influencing the dependent variable (suppression duration) was operating while the stimulus remained invisible.

With this paradigm, we found a significant local biological motion inversion effect: upright scrambled biological motion took less time to gain dominance compared to inverted scrambled biological motion against the identical suppression noise (1.42 s vs. 1.78 s; $t(7) = 2.48$, $p < 0.05$; Fig. 5-3b), an effect that was highly consistent across all individual observers (see Fig. 5-3c for individual data). Similar to Exp. 1, there was no significant inversion effect when the global configuration remained intact (1.53 s vs. 1.44 s; $t(7) = 1.73$, $p > 0.1$; Fig. 5-3b). Of course there would be an inversion effect for the intact biological motion signals in most explicit tasks on biological motion perception, as shown in a number of previous studies, such as detecting biological motion from noise or discriminating walking directions (Pavlova & Sokolov, 2000; Proffitt & Bertenthal, 1990; Sumi, 1984; Troje & Westhoff, 2006). Our experimental design (detecting target position) focused on the automatic processing of the stimuli, which is likely the reason that the orientation of the global configuration did not make a significant difference.
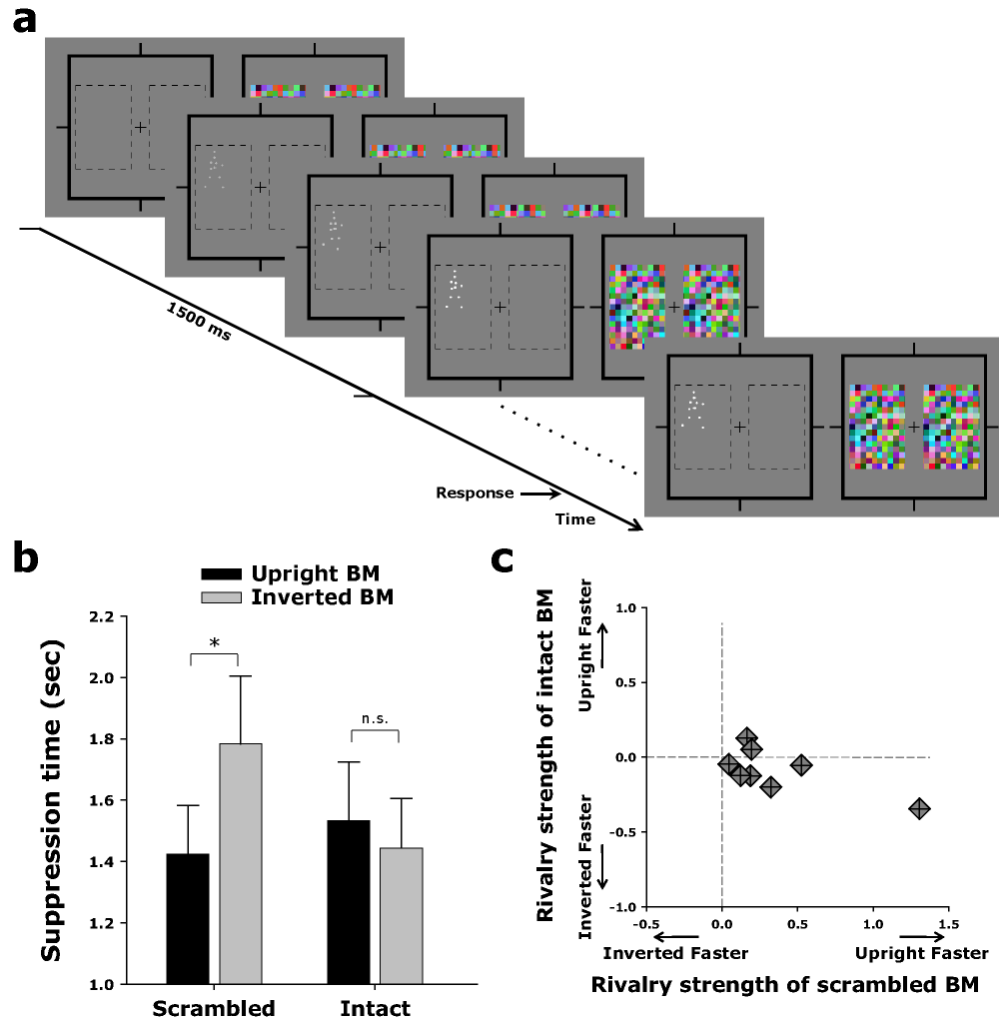
Figure 5-3. Experimental paradigm and results from Exp. 2. (a) A test biological motion stimulus (e.g., an intact upright BM as shown) was gradually introduced to one eye to compete with a dynamic noise pattern presented to the other eye. The dynamic noise was presented at full contrast from the beginning, but the contrast of the test motion stimulus was smoothly ramped up from 0 to 100% within a period of 1.5 s starting from the beginning of the trial, and then remained constant until the observer made a response to indicate on which side the test stimulus appeared. (b) Scrambled upright BM took significantly less time to overcome interocular suppression than scrambled inverted BM, while there was no difference in the intact BM condition. (c) The advantage of scrambled upright BM in interocular competition is highly consistent across individual observers (diamonds all to the right of the vertical dashed line) while there was no effect associated with intact BM (diamonds around the horizontal dashed line). * $p < .05$; n.s., no significance.

Apparently, the automatic processing of local biological motion properties can be better revealed when global configuration information is removed, consistent with our findings from Exp. 1 showing that the attentional effect of biological motion arises from local motion signals without global configural information. Taken together, these results suggest that the human visual system is sensitive to and makes important use of the local motion properties of biological motion, and the mechanism for processing local biological motion information could operate independent of observers' awareness. Our results are consistent with recent notion that biological motion perception can be an intrinsic capacity of the visual system (Johnson, 2006; Simion et al., 2008; Troje & Westhoff, 2006).
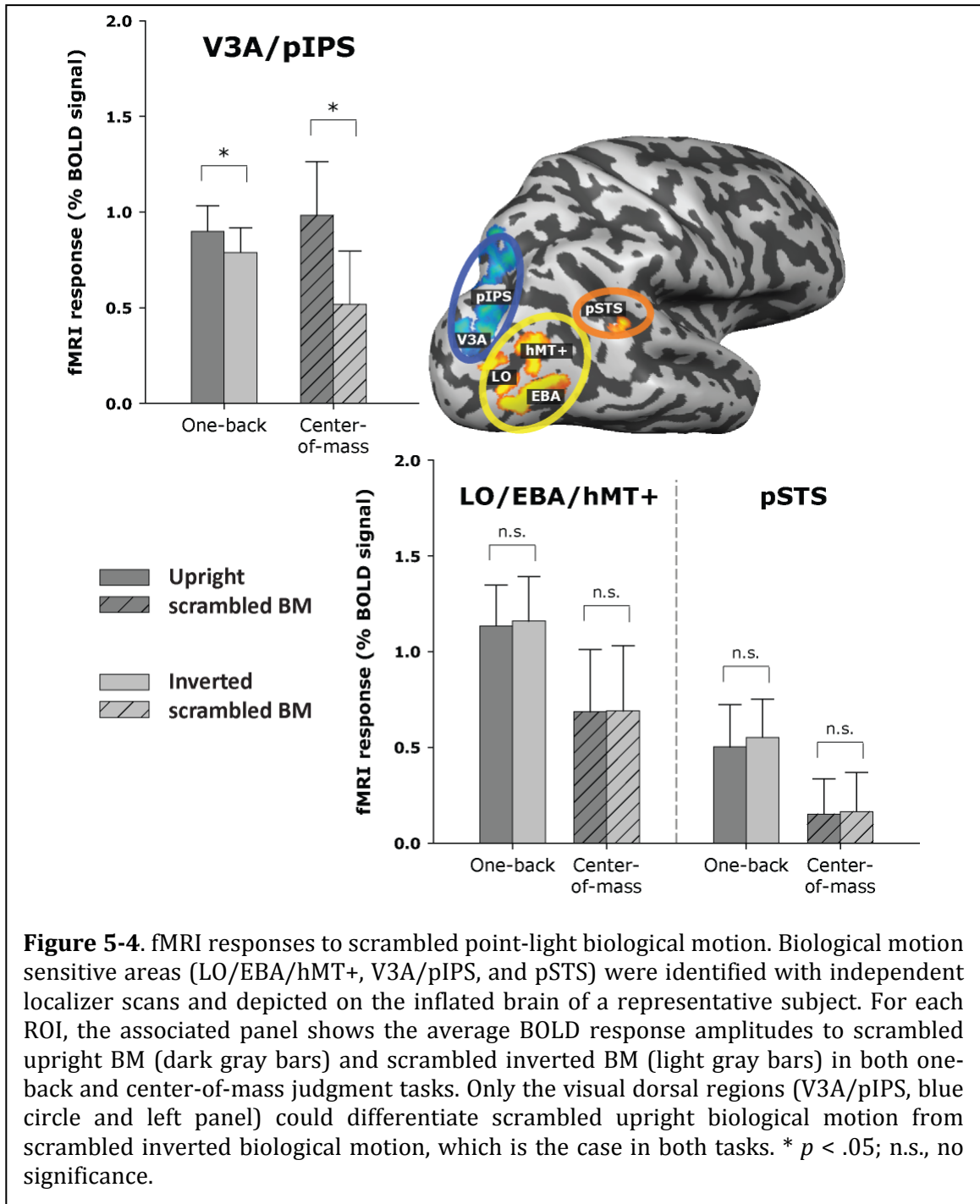
Experiment 3. Cortical response to local biological motion information

Results from the psychophysical experiments (Exp. 1 and Exp. 2) suggest the existence of a life motion detection mechanism. Hence, it is important to investigate its neural mechanism and in particular, the cortical regions for processing local motion properties associated with biological motion.

The visual and functional importance of biological motion perception is instantiated in a specialized neural network (E. D. Grossman & Blake, 2002; Peelen et al., 2006). Recent neuroimaging studies have examined the specialized brain areas involved in biological motion perception that detect either motion or form information, including the human motion complex (hMT+ or V5), lateral occipital complex (LOC), extrastriate body area

(EBA (Downing et al., 2001; E. D. Grossman & Blake, 2002; Michels et al., 2005)), fusiform body area (FBA (Peelen & Downing, 2005; Peelen et al., 2006; Schwarzlose, Baker, & Kanwisher, 2005)), and fusiform face area (FFA (E. D. Grossman & Blake, 2002; Kanwisher et al., 1997; Michels et al., 2005)). More importantly, the posterior superior temporal sulcus (pSTS), particularly in the right hemisphere, has consistently been found to respond strongly to biological motion and to play a critical role in the perception of biological motion (Grezes et al., 2001; E. Grossman et al., 2000; E. D. Grossman & Blake, 2001, 2002; E. D. Grossman et al., 2004; Peuskens et al., 2005; Santi et al., 2003; Saygin et al., 2004; Thompson et al., 2005; Vaina et al., 2001). However, none of the existing studies targeted the neural sites specifically sensitive to local motion components in biological motion.

Therefore, we used fMRI to identify potential cortical regions that are sensitive to local biological motion information. With a typical contrast between a coherent point-light walker and a scrambled control, as is often done in published studies, specific sensitivity to local motion information could not be revealed. However, given the psychophysical observations reported in earlier sections, our goal became simply to localize cortical regions that could differentiate between the two scrambled motion patterns: scrambled upright biological motion vs. scrambled inverted biological motion. These two types of stimuli are perceptually indistinguishable to naïve observers, yet they were differentiated by their visual spatial attention system (Exp. 1), and had different strength in binocular rivalry competition (Exp. 2). Certainly the difference between these two types of scrambled motion was registered in the brain, but which part(s) of the brain?

Figure 5-4. fMRI responses to scrambled point-light biological motion. Biological motion sensitive areas (LO/EBA/hMT+, V3A/pIPS, and pSTS) were identified with independent localizer scans and depicted on the inflated brain of a representative subject. For each ROI, the associated panel shows the average BOLD response amplitudes to scrambled upright BM (dark gray bars) and scrambled inverted BM (light gray bars) in both one-back and center-of-mass judgment tasks. Only the visual dorsal regions (V3A/pIPS, blue circle and left panel) could differentiate scrambled upright biological motion from scrambled inverted biological motion, which is the case in both tasks. * $p < .05$; n.s., no significance.

Among all of the predefined ROIs (see Experimental Procedures for ROIs identification), dorsal visual areas including V3A and posterior intraparietal sulcus (pIPS) showed significantly higher activations to upright biological motion than

inverted biological motion when global form and global motion information was completely removed (Fig. 5-4, $t(5) = 4.03$, $p < 0.01$), while there was no significant difference between scrambled upright and scrambled inverted biological motion for the other ROIs (EBA: $t(5) = 0.80$, $p > 0.4$; hMT+: $t(5) = 0.54$, $p > 0.6$; LO: $t(5) = 0.86$, $p > 0.4$; pSTS: $t(5) = -1.72$, $p > 0.1$). Indeed, contrasting between scrambled upright and scrambled inverted biological motion also revealed stronger activations to the upright local biological motion signals in the same visual dorsal areas. This result implies that human dorsal cortical regions are sensitive to local "life motion" information contained in scrambled biological motion. Previous animal, neuropsychological, and imaging studies have demonstrated that the dorsal stream can process visual information and guide action without accompanying conscious knowledge (Fang & He, 2005; Goodale et al., 1991; James et al., 2003; Kluever & Bucy, 1938; Milner & Goodale, 1995; Ungerleider & Mishkin, 1982). In our psychophysical experiments, we found that local motion properties in biological motion could be processed automatically and independent of conscious representation of the stimulus, consistent with known functional properties of visual dorsal regions.

Since a one-back task was used in the fMRI experiment, observers had to pay close attention to the trajectories of moving dots to perform the task. It is possible that the attentional demand in the task leads to the localization of the dorsal areas as the region differentiating the two scrambled biological motion signals, as the parietal cortex plays an important role in visual attention (Wojciulik & Kanwisher, 1999). To test whether the dorsal parietal region is automatically engaged in processing local biological motion signals, we conducted a control fMRI experiment with the same setup except that

observers were asked to simply judge whether the center-of-mass of the point-lights was presented above or below the fixation point. This task is irrelevant with regard to biological motion perception per se, and also significantly reduces the attentional demand for the stimuli. Results from this control experiment replicated the response pattern observed with the one-back task: visual dorsal areas showed stronger activation to scrambled upright biological motion compared to scrambled inverted biological motion, while there was no difference for the other ROIs (see Fig. 5-4). Therefore, these results further suggest that human visual dorsal areas are capable of automatically distinguishing local biological motion from non-biological motion, making the dorsal parietal areas the prime neural candidate for the life motion detector (Battelli et al., 2003).

CONCLUSIONS

Motion of biological entities consists of both local and global components. Most previous studies have focused on the global aspects of biological motion perception while local information processing has received much less attention. As such, recent computational models have emphasized only the contributions of global form and global motion information to biological motion perception (Lange & Lappe, 2006). Significantly, the current psychophysical experiments showed that the presence of upright local biological motion information can automatically attract observers' spatial attention, and that upright local motion has an advantage in interocular competition over inverted local motion in the absence of observers' awareness. These findings were true

even though observers were not able to explicitly distinguish between the two sets of seemingly random moving dots. The companion fMRI experiment further suggests that the human dorsal cortical region is sensitive to local biological motion properties – the dorsal cortex could distinguish between upright and inverted biological motion, even when global configurations of the stimuli were removed with spatial scrambling.

In summary, the current study highlights that local biological motion signals can be processed automatically by the visual system, independent of global form and global motion information, and dorsal occipitoparietal areas are the prime neural candidate for this life motion detector. The automatic detection and processing of biologically meaningful local motion information is evolutionarily significant, and presumably serves as a life motion detector to facilitate survival of the organism by directing attention to life motion signals and leading to the engagement of more specialized mechanisms for processing the global aspects of the input (Johnson, 2006; Troje & Westhoff, 2006).

Summary

This dissertation presents four studies which employ both psychophysical and brain imaging techniques to probe the neural encoding of face and biological motion information.

Study 1 reports a psychophysical study using a novel interocular suppression paradigm. It shows that familiar and recognizable information (e.g., upright faces and recognizable words), even when suppressed and invisible, is processed differently from unfamiliar information (e.g., inverted faces and unrecognizable words). In other words, high-level visual information in the suppressed phase of binocular rivalry, including face orientation information, can be processed to the extent that object-related representations can be achieved.

Study 2 and Study 3 report fMRI and ERP studies showing that by rendering face images invisible through interocular suppression, distinct patterns of responses in FFA, STS, and the amygdala can be better revealed, with STS and the amygdala being selectively sensitive to facial-expression information. The results support the idea that face perception involves two functionally (identity versus expression or invariant versus changeable) and anatomically (FFA, STS, and amygdala) distinct pathways, and shed light on the temporal sequence of the automatic extraction of facial expression information.

Study 4 reports a series of psychophysical and brain imaging experiments which examine the automatic processing of local biological motion information in the human brain. The results highlight the intrinsic sensitivity of the human brain to local life motion signals, and point to a specialized neural mechanism serving as a "life motion detector" that automatically processes biologically meaningful local motion signals independent of global configuration.

Taken together, these studies indicate that the human visual system is sensitive to biologically significant information, which can be processed without awareness. The findings add to our understandings of the brain mechanisms underlying humans' superb processing of face and biological motion information.

Bibliography

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature, 372*(6507), 669-672.

Ahlstrom, V., Blake, R., & Ahlstrom, U. (1997). Perception of biological motion. *Perception, 26*(12), 1539-1548.

Anaki, D., Zion-Golumbic, E., & Bentin, S. (2007). Electrophysiological neural mechanisms for detection, configural analysis and recognition of faces. *Neuroimage, 37*(4), 1407-1416.

Barclay, C. D., Cutting, J. E., & Kozlowski, L. T. (1978). Temporal and spatial factors in gait perception that influence gender recognition. *Percept Psychophys, 23*(2), 145-152.

Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. *J Exp Psychol Hum Percept Perform, 4*(3), 373-379.

Battelli, L., Cavanagh, P., & Thornton, I. M. (2003). Perception of biological motion in parietal patients. *Neuropsychologia, 41*(13), 1808-1816.

Batty, M., & Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Brain Res Cogn Brain Res, 17*(3), 613-620.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2003). FMRI responses to video and point-light displays of moving humans and manipulable objects. *J Cogn Neurosci, 15*(7), 991-1001.

Beintema, J. A., & Lappe, M. (2002). Perception of biological motion without local image motion. *Proc Natl Acad Sci U S A, 99*(8), 5661-5663.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *J Cogn Neurosci, 8*, 551-565.

Bertenthal, B. I., & Pinto, J. (1994). Global processing of biological motions. *Psychol Sci, 5*, 221-225.

Bertenthal, B. I., Proffitt, D. R., & Kramer, S. J. (1987). Perception of biomechanical motions by infants: implementation of various processing constraints. *J Exp Psychol Hum Percept Perform, 13*(4), 577-585.

Bidet-Ildei, C., Orliaguet, J. P., Sokolov, A. N., & Pavlova, M. (2006). Perception of elliptic biological motion. *Perception, 35*(8), 1137-1147.

Blake, R. (1989). A neural theory of binocular rivalry. *Psychol Rev, 96*(1), 145-167.

Blake, R. (2001). A primer on binocular rivalry, including current controversies. *Brain Mind, 2*, 5-38.

Blake, R., & Fox, R. (1974). Adaptation to invisible gratings and the site of binocular rivalry suppression. *Nature, 249*(456), 488-490.

Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nat Rev Neurosci, 3*(1), 13-21.

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annu Rev Psychol, 58*, 47-73.

Blake, R., Tadin, D., Sobel, K. V., Raissian, T. A., & Chong, S. C. (2006). Strength of early visual adaptation depends on visual awareness. *Proc Natl Acad Sci U S A, 103*(12), 4783-4788.

Blau, V. C., Maurer, U., Tottenham, N., & McCandliss, B. D. (2007). The face-specific N170 component is modulated by emotional facial expression. *Behav Brain Funct, 3*, 7.

Bonda, E., Petrides, M., Ostry, D., & Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *J Neurosci, 16*(11), 3737-3744.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226*(5241), 177-178.

Bouma, H. (1973). Visual interference in the parafoveal recognition of initial and final letters of words. *Vision Res, 13*(4), 767-782.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis, 10*(4), 433-436.

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., et al. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron, 17*(5), 875-887.

Breitmeyer, B. G. (1984). *Visual masking : an integrative approach.* Oxford; New York: Oxford University Press.

Breitmeyer, B. G., Ro, T., & Singhal, N. S. (2004). Unconscious color priming occurs at stimulus- not percept-dependent levels of processing. *Psychol Sci, 15*(3), 198-202.

Brewer, A. A., Liu, J., Wade, A. R., & Wandell, B. A. (2005). Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nat Neurosci, 8*(8), 1102-1109.

Bruce, V. (1986). Influences of familiarity on the processing of faces. *Perception, 15*(4), 387-397.

Bruce, V., & Young, A. (1986). Understanding face recognition. *Br J Psychol, 77 (Pt 3)*, 305-327.

Bruyer, R., Laterre, C., Seron, X., Feyereisen, P., Strypstein, E., Pierrard, E., et al. (1983). A case of prosopagnosia with some preserved covert remembrance of familiar faces. *Brain Cogn, 2*(3), 257-284.

Bulthoff, I., Bulthoff, H., & Sinha, P. (1998). Top-down influences on stereoscopic depth-perception. *Nat Neurosci, 1*(3), 254-257.

Buracas, G. T., & Boynton, G. M. (2002). Efficient design of event-related fMRI experiments using M-sequences. *Neuroimage, 16*(3 Pt 1), 801-813.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nat Rev Neurosci, 6*(8), 641-651.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *J Exp Psychol Hum Percept Perform, 26*(2), 527-551.

Calder, A. J., Young, A. W., Rowland, D., Perrett, D. I., Hodges, J. R., & Etcoff, N. L. (1996). Facial emotion recognition after bilateral amygdala damage: differentially severe impairment of fear. *Cognit Neuropsychol, 13*, 699-745.

Campbell, R., Brooks, B., de Haan, E., & Roberts, T. (1996). Dissociating face processing skills: decision about lip-read speech, expression, and identity. *Q J Exp Psychol A, 49*(2), 295-314.

Campbell, R., Heywood, C. A., Cowey, A., Regard, M., & Landis, T. (1990). Sensitivity to eye gaze in prosopagnosic patients and monkeys with superior temporal sulcus ablation. *Neuropsychologia, 28*(11), 1123-1142.

Chang, D. H., & Troje, N. F. (2008). Perception of animacy and direction from local biological motion signals. *J Vis, 8*(5), 3 1-10.

Chang, D. H., & Troje, N. F. (2009). Acceleration carries the local inversion effect in biological motion perception. *J Vis, 9*(1), 19 11-17.

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat Neurosci, 2*(10), 913-919.

Chatterjee, S. H., Freyd, J. J., & Shiffrar, M. (1996). Configural processing in the perception of apparent biological motion. *J Exp Psychol Hum Percept Perform, 22*(4), 916-929.

Clark, V. P., Keil, K., Maisog, J. M., Courtney, S., Ungerleider, L. G., & Haxby, J. V. (1996). Functional magnetic resonance imaging of human visual cortex during face matching: a comparison with positron emission tomography. *Neuroimage, 4*(1), 1-15.

Clarke, T. J., Bradshaw, M. F., Field, D. T., Hampson, S. E., & Rose, D. (2005). The perception of emotion from body movement in point-light displays of interpersonal dialogue. *Perception, 34*(10), 1171-1180.

Costantini, M., Galati, G., Ferretti, A., Caulo, M., Tartaro, A., Romani, G. L., et al. (2005). Neural systems underlying observation of humanly impossible movements: an FMRI study. *Cereb Cortex, 15*(11), 1761-1767.

Cowey, A., & Vaina, L. M. (2000). Blindness to form from motion despite intact static form perception and motion detection. *Neuropsychologia, 38*(5), 566-578.

Cutting, J. E., & Kozlowski, L. T. (1977). Recognition of friends by their walk: gait perception without familiarity cues. *Bull Psychon Soc, 9*, 353-356.

Cutting, J. E., Moore, C., & Morrison, R. (1988). Masking the motions of human gait. *Percept Psychophys, 44*(4), 339-347.

DeYoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., et al. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc Natl Acad Sci U S A, 93*(6), 2382-2386.

Dittrich, W. H. (1993). Action categories and the perception of biological motion. *Perception, 22*(1), 15-22.

Dittrich, W. H., Troscianko, T., Lea, S. E., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception, 25*(6), 727-738.

Dougherty, R. F., Koch, V. M., Brewer, A. A., Fischer, B., Modersitzki, J., & Wandell, B. A. (2003). Visual field representations and locations of visual areas V1/2/3 in human visual cortex. *J Vis, 3*(10), 586-598.

Downing, P. E., Chan, A. W., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain specificity in visual cortex. *Cereb Cortex, 16*(10), 1453-1461.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science, 293*(5539), 2470-2473.

Eimer, M. (2000a). Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clin Neurophysiol, 111*(4), 694-705.

Eimer, M. (2000b). The face-specific N170 component reflects late stages in the structural encoding of faces. *Neuroreport, 11*(10), 2319-2324.

Eimer, M., & Holmes, A. (2002). An ERP study on the time course of emotional face processing. *Neuroreport, 13*(4), 427-431.

Engel, E. (1956). The role content in binocular resolution. *Am J Psychol, 69*(1), 87-91.

Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb Cortex, 7*(2), 181-192.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*(6676), 598-601.

Eriksen, C. W., & Hoffman, J. E. (1973). The extent of processing of noise elements during selective encoding from visual displays. *Percept Psychophys, 14*, 155-160.

Etcoff, N. L. (1984). Selective attention to facial identity and facial emotion. *Neuropsychologia, 22*(3), 281-295.

Fairhall, S. L., & Ishai, A. (2007). Effective Connectivity within the Distributed Cortical Network for Face Perception. *Cereb Cortex, 17*(10), 2400-2406.

Fang, F., & He, S. (2005). Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nat Neurosci, 8*(10), 1380-1385.

Fani, L., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing people from their movements. *J Exp Psychol Hum Percept Perform, 31*, 210-220.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex, 1*(1), 1-47.

Fitzgerald, D. A., Angstadt, M., Jelsone, L. M., Nathan, P. J., & Phan, K. L. (2006). Beyond threat: amygdala reactivity across multiple expressions of facial affect. *Neuroimage, 30*(4), 1441-1448.

Fox, R., & McDaniel, C. (1982). The perception of biological motion by human infants. *Science, 218*(4571), 486-487.

Freeman, A. W. (2005). Multistage model for binocular rivalry. *J Neurophysiol, 94*(6), 4412-4420.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat Neurosci, 2*(6), 568-573.

George, M. S., Ketter, T. A., Gill, D. S., Haxby, J. V., Ungerleider, L. G., Herscovitch, P., et al. (1993). Brain regions involved in recognizing facial emotion or identity: an oxygen-15 PET study. *J Neuropsychiatry Clin Neurosci, 5*(4), 384-394.

Goodale, M. A., Milner, A. D., Jakobson, L. S., & Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature, 349*(6305), 154-156.

Grezes, J., Fonlupt, P., Bertenthal, B., Delon-Martin, C., Segebarth, C., & Decety, J. (2001). Does perception of biological motion rely on specific brain regions? *Neuroimage, 13*(5), 775-785.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nat Neurosci, 7*(5), 555-562.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., et al. (2000). Brain areas involved in perception of biological motion. *J Cogn Neurosci, 12*(5), 711-720.

Grossman, E. D., Battelli, L., & Pascual-Leone, A. (2005). Repetitive TMS over posterior STS disrupts perception of biological motion. *Vision Res, 45*(22), 2847-2853.

Grossman, E. D., & Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision Res, 41*(10-11), 1475-1482.

Grossman, E. D., & Blake, R. (2002). Brain Areas Active during Visual Perception of Biological Motion. *Neuron, 35*(6), 1167-1175.

Grossman, E. D., Blake, R., & Kim, C. Y. (2004). Learning to see biological motion: brain activity parallels behavior. *J Cogn Neurosci, 16*(9), 1669-1679.

Hadjikhani, N., Liu, A. K., Dale, A. M., Cavanagh, P., & Tootell, R. B. (1998). Retinotopy and color sensitivity in human visual cortical area V8. *Nat Neurosci, 1*(3), 235-241.

Halgren, E., Dale, A. M., Sereno, M. I., Tootell, R. B., Marinkovic, K., & Rosen, B. R. (1999). Location of human face-selective cortex with respect to retinotopic areas. *Hum Brain Mapp, 7*(1), 29-37.

Harris, A., & Nakayama, K. (2007). Rapid face-selective adaptation of an early extrastriate component in MEG. *Cereb Cortex, 17*(1), 63-70.

Harris, A., & Nakayama, K. (2008). Rapid adaptation of the m170 response: importance of face parts. *Cereb Cortex, 18*(2), 467-476.

Hasselmo, M. E., Rolls, E. T., & Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behav Brain Res, 32*(3), 203-218.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn Sci, 4*(6), 223-233.

Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., & Grady, C. L. (1994). The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. *J Neurosci, 14*(11 Pt 1), 6336-6353.

Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., & Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. *Neuron, 22*(1), 189-199.

He, S., Carlson, T. A., & Chen, X. (2004). Parallel pathways and temporal dynamics in binocular rivalry. In D. Alais & R. Blake (Eds.), *Binocular rivalry and perceptual ambiguity*. Boston, MA: MIT Press.

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature, 383*(6598), 334-337.

He, S., Cavanagh, P., & Intriligator, J. (1997). Attentional resolution. *Trends Cogn Sci, 1*, 115-121.

Heywood, C. A., & Cowey, A. (1992). The role of the 'face-cell' area in the discrimination and recognition of faces by monkeys. *Philos Trans R Soc Lond B Biol Sci, 335*(1273), 31-37; discussion 37-38.

Hill, H., Jinno, Y., & Johnston, A. (2003). Comparing solid-body with point-light animations. *Perception, 32*(5), 561-566.

Hirai, M., & Hiraki, K. (2005). An event-related potentials study of biological motion perception in human infants. *Brain Res Cogn Brain Res, 22*(2), 301-304.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nat Neurosci, 3*(1), 80-84.

Hornak, J., Rolls, E. T., & Wade, D. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. *Neuropsychologia, 34*(4), 247-261.

Humphreys, G. W., Donnelly, N., & Riddoch, M. J. (1993). Expression is computed separately from facial identity, and it is computed separately for moving and static faces: neuropsychological evidence. *Neuropsychologia, 31*(2), 173-181.

Ikeda, H., Blake, R., & Watanabe, K. (2005). Eccentric perception of biological motion is unscalably poor. *Vision Res, 45*(15), 1935-1943.

Ishai, A., Schmidt, C. F., & Boesiger, P. (2005). Face perception is mediated by a distributed cortical network. *Brain Res Bull, 67*(1-2), 87-93.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proc Natl Acad Sci U S A, 96*(16), 9379-9384.

Itier, R. J., Herdman, A. T., George, N., Cheyne, D., & Taylor, M. J. (2006). Inversion and contrast-reversal effects on face processing assessed by MEG. *Brain Res, 1115*(1), 108-120.

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

Itier, R. J., & Taylor, M. J. (2004b). Source analysis of the N170 to faces and objects. *Neuroreport, 15*(8), 1261-1265.

James, T. W., Culham, J., Humphrey, G. K., Milner, A. D., & Goodale, M. A. (2003). Ventral occipital lesions impair object recognition but not object-directed grasping: an fMRI study. *Brain, 126*(Pt 11), 2463-2475.

Jiang, Y., Costello, P., Fang, F., Huang, M., & He, S. (2006). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *Proc Natl Acad Sci U S A, 103*(45), 17048-17052.

Jiang, Y., Costello, P., & He, S. (2007). Processing of invisible stimuli: advantage of upright faces and recognizable words in overcoming interocular suppression. *Psychol Sci, 18*(4), 349-355.

Jiang, Y., & He, S. (2006). Cortical responses to invisible faces: dissociating subsystems for facial-information processing. *Curr Biol, 16*(20), 2023-2029.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Percept Psychophys, 14*, 195-204.

Johnson, M. H. (2006). Biological motion: a perceptual life detector? *Curr Biol, 16*(10), R376-377.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci, 17*(11), 4302-4311.

Kim, C. Y., & Blake, R. (2005). Psychophysical strategies for rendering the normally visible "invisible". *Trends Cogn Sci, 9*, 381-388.

Kluever, H., & Bucy, P. C. (1938). An analysis of certain effects of bilateral temporal lobectomy in the rhesus monkey, with special reference to "psychic blindness". *J. Psychol, 5*, 203-211.

Kovacs, I., Papathomas, T. V., Yang, M., & Feher, A. (1996). When the brain changes its mind: interocular grouping during binocular rivalry. *Proc Natl Acad Sci U S A, 93*(26), 15508-15511.

Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point light display. *Percept Psychophys, 21*, 575-580.

Kozlowski, L. T., & Cutting, J. E. (1978). Recognizing the gender of walkers from point-lights mounted on ankles: some second thoughts. *Percept Psychophys, 23*, 459.

Krolak-Salmon, P., Henaff, M. A., Vighetto, A., Bertrand, O., & Mauguiere, F. (2004). Early amygdala reaction to fear spreading in occipital, temporal, and frontal cortex: a depth electrode ERP study in human. *Neuron, 42*(4), 665-676.

LaBar, K. S., Crupain, M. J., Voyvodic, J. T., & McCarthy, G. (2003). Dynamic perception of facial affect and identity in the human brain. *Cereb Cortex, 13*(10), 1023-1033.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci, 23*(11), 571-579.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). Motivated attention: affect, activation, and action. In P. J. Lang, R. F. Simons & M. Balaban (Eds.), *Attention and Orienting, Sensory and Motivational Processes* (pp. 97-135). New Jersey: Erlbaum.

Lange, J., Georg, K., & Lappe, M. (2006). Visual perception of biological motion by form: a template-matching analysis. *J Vis, 6*(8), 836-849.

Lange, J., & Lappe, M. (2006). A model of biological motion perception from configural form cues. *J Neurosci, 26*(11), 2894-2906.

Lehky, S. R. (1988). An astable multivibrator model of binocular rivalry. *Perception, 17*(2), 215-228.

Lehmkuhle, S. W., & Fox, R. (1975). Effect of binocular rivalry suppression on the motion aftereffect. *Vision Res, 15*(7), 855-859.

Levelt, W. J. M. (1968). *On binocular rivalry*. The Hague, Paris,: Mouton.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. *Nat Neurosci, 5*(9), 910-916.

Lu, H., Tjan, B. S., & Liu, Z. (2006). Shape recognition alters sensitivity in stereoscopic depth discrimination. *J Vis, 6*(1), 75-86.

Marcel, A. J. (1983). Conscious and unconscious perception: experiments on visual masking and word recognition. *Cognit Psychol, 15*(2), 197-237.

Mather, G., & Murdoch, L. (1994). Gender discrimination in biological motion displays based on dynamic cues. *Proc R Soc Lond B Biol Sci, 258*, 273-279.

Mather, G., Radford, K., & West, S. (1992). Low-level visual processing of biological motion. *Proc Biol Sci, 249*(1325), 149-155.

Mather, G., & West, S. (1993). Recognition of animal locomotion from dynamic point-light displays. *Perception, 22*(7), 759-766.

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends Cogn Sci, 7*(7), 293-299.

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *J Cognit Neurosci, 9*, 605-610.

Michels, L., Lappe, M., & Vaina, L. M. (2005). Visual areas involved in the perception of human movement from dynamic form analysis. *Neuroreport, 16*(10), 1037-1041.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford; New York: Oxford University Press.

Miyoshi, M., Katayama, J., & Morotomi, T. (2004). Face-specific N170 component is modulated by facial expressional change. *Neuroreport, 15*(5), 911-914.

Moradi, F., Koch, C., & Shimojo, S. (2005). Face adaptation depends on seeing the face. *Neuron, 45*(1), 169-175.

Morris, J. S., DeGelder, B., Weiskrantz, L., & Dolan, R. J. (2001). Differential extrageniculostriate and amygdala responses to presentation of emotional faces in a cortically blind field. *Brain, 124*(Pt 6), 1241-1252.

Morris, J. S., Friston, K. J., Buchel, C., Frith, C. D., Young, A. W., Calder, A. J., et al. (1998). A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain, 121 (Pt 1)*, 47-57.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., et al. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature, 383*(6603), 812-815.

Morris, J. S., Ohman, A., & Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature, 393*(6684), 467-470.

Morris, J. S., Ohman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proc Natl Acad Sci U S A, 96*(4), 1680-1685.

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychol Rev, 98*(2), 164-181.

Moutoussis, K., & Zeki, S. (2002). The relationship between cortical activation and perception investigated with invisible stimuli. *Proc Natl Acad Sci U S A, 99*(14), 9527-9532.

Mueller, T. J. (1990). A physiological model of binocular rivalry. *Vis Neurosci, 4*(1), 63-73.

Narumoto, J., Okada, T., Sadato, N., Fukui, K., & Yonekura, Y. (2001). Attention to emotion modulates fMRI activity in human right superior temporal sulcus. *Brain Res Cogn Brain Res, 12*(2), 225-231.

Nguyen, V. A., Freeman, A. W., & Alais, D. (2003). Increasing depth of binocular rivalry suppression along two visual pathways. *Vision Res, 43*(19), 2003-2008.

Norman, J. F., Payton, S. M., Long, J. R., & Hawkes, L. M. (2004). Aging and the perception of biological motion. *Psychol Aging, 19*(1), 219-225.

Oram, M. W., & Perrett, D. I. (1994). Responses of anterior superior temporal polysensory (STPa) neurons to "biological motion" stimuli. *J Cogn Neurosci, 6*, 99-116.

Oram, M. W., & Perrett, D. I. (1996). Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *J Neurophysiol, 76*(1), 109-129.

Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., Coleman, M., et al. (2005). Plasticity of face processing in infancy. *Proc Natl Acad Sci U S A, 102*(14), 5297-5300.

Pasley, B. N., Mayes, L. C., & Schultz, R. T. (2004). Subcortical discrimination of unperceived objects during binocular rivalry. *Neuron, 42*(1), 163-172.

Pavlova, M., & Sokolov, A. (2000). Orientation specificity in biological motion perception. *Percept Psychophys, 62*(5), 889-899.

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *J Neurophysiol, 93*(1), 603-608.

Peelen, M. V., Wiggett, A. J., & Downing, P. E. (2006). Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. *Neuron, 49*(6), 815-822.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis, 10*(4), 437-442.

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: distinguishing feature integration from detection. *J Vis, 4*(12), 1136-1169.

Pelphrey, K. A., Mitchell, T. V., McKeown, M. J., Goldstein, J., Allison, T., & McCarthy, G. (2003). Brain activity evoked by the perception of human walking: controlling for meaningful coherent motion. *J Neurosci, 23*(17), 6819-6825.

Perrett, D. I., Smith, P. A., Mistlin, A. J., Chitty, A. J., Head, A. S., Potter, D. D., et al. (1985). Visual analysis of body movements by neurones in the temporal cortex of the macaque monkey: a preliminary report. *Behav Brain Res, 16*(2-3), 153-170.

Pessoa, L., Kastner, S., & Ungerleider, L. G. (2002). Attentional control of the processing of neural and emotional stimuli. *Brain Res Cogn Brain Res, 15*(1), 31-45.

Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proc Natl Acad Sci U S A, 99*(17), 11458-11463.

Peuskens, H., Vanrie, J., Verfaillie, K., & Orban, G. A. (2005). Specificity of regions processing biological motion. *Eur J Neurosci, 21*(10), 2864-2875.

Pinto, J., & Shiffrar, M. (1999). Subconfigurations of the human form in the perception of biological motion displays. *Acta Psychol (Amst), 102*(2-3), 293-318.

Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition, 82*(2), B51-61.

Posamentier, M. T., & Abdi, H. (2003). Processing faces and facial expressions. *Neuropsychol Rev, 13*(3), 113-143.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In B. H. & B. D. G. (Eds.), *Attention and performance* (pp. 531-556). Hillsdale, New Jersey: Erlbaum.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *J Exp Psychol, 109*(2), 160-174.

Price, C. J., & Devlin, J. T. (2004). The pro and cons of labelling a left occipitotemporal region: "the visual word form area". *Neuroimage, 22*(1), 477-479.

Proffitt, D. R., & Bertenthal, B. I. (1990). Converging operations revisited: assessing what infants perceive using discrimination measures. *Percept Psychophys, 47*(1), 1-11.

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *J Neurosci, 18*(6), 2188-2199.

Reed, C. L., Stone, V. E., Bozova, S., & Tanaka, J. (2003). The body-inversion effect. *Psychol Sci, 14*(4), 302-308.

Sabatinelli, D., Bradley, M. M., Fitzsimmons, J. R., & Lang, P. J. (2005). Parallel amygdala and inferotemporal activation reflect emotional intensity and fear relevance. *Neuroimage, 24*(4), 1265-1270.

Sagiv, N., & Bentin, S. (2001). Structural encoding of human and schematic faces: holistic and part-based processes. *J Cogn Neurosci, 13*(7), 937-951.

Sams, M., Hietanen, J. K., Hari, R., Ilmoniemi, R. J., & Lounasmaa, O. V. (1997). Face-specific responses from the human inferior occipito-temporal cortex. *Neuroscience, 77*(1), 49-55.

Santi, A., Servos, P., Vatikiotis-Bateson, E., Kuratate, T., & Munhall, K. (2003). Perceiving biological motion: dissociating visible speech from walking. *J Cogn Neurosci, 15*(6), 800-809.

Saunders, D. R., Suchan, J., & Troje, N. F. (in press). Off on the wrong foot: local features in biological motion. *Perception*.

Saygin, A. P., Wilson, S. M., Hagler, D. J., Jr., Bates, E., & Sereno, M. I. (2004). Point-light biological motion perception activates human premotor cortex. *J Neurosci, 24*(27), 6181-6188.

Schenk, T., & Zihl, J. (1997). Visual motion perception after brain damage: II. Deficits in form-from-motion perception. *Neuropsychologia, 35*(9), 1299-1310.

Schwarzlose, R. F., Baker, C. I., & Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. *J Neurosci, 25*(47), 11055-11059.

Sereno, M. I., McDonald, C. T., & Allman, J. M. (1994). Analysis of retinotopic maps in extrastriate cortex. *Cereb Cortex, 4*(6), 601-620.

Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain, 115 Pt 1*, 15-36.

Sheinberg, D. L., & Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proc Natl Acad Sci U S A, 94*(7), 3408-3413.

Shiffrar, M., Lichtey, L., & Heptulla Chatterjee, S. (1997). The perception of biological motion across apertures. *Percept Psychophys, 59*(1), 51-59.

Shipley, T. F. (2003). The effect of object and event orientation on perception of biological motion. *Psychol Sci, 14*(4), 377-380.

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proc Natl Acad Sci U S A, 105*(2), 809-813.

Stoerig, P., & Cowey, A. (1997). Blindsight in man and monkey. *Brain, 120 (Pt 3)*, 535-559.

Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature, 400*(6747), 869-873.

Sugie, N. (1982). Neural models of brightness perception and retinal rivalry in binocular vision. *Biol Cybern, 43*(1), 13-21.

Sugita, Y. (2008). Face perception in monkeys reared with no exposure to faces. *Proc Natl Acad Sci U S A, 105*(1), 394-398.

Sumi, S. (1984). Upside-down presentation of the Johansson moving light-spot pattern. *Perception, 13*(3), 283-286.

Sumi, S. (2000). Perception of point-light walker produced by eight lights attached to the back of the walker. *Swiss J Psychol, 59*, 126-132.

127

Thompson, J. C., Clarke, M., Stewart, T., & Puce, A. (2005). Configural processing of biological motion in human superior temporal sulcus. *J Neurosci, 25*(39), 9059-9066.

Tong, F., & Engel, S. A. (2001). Interocular rivalry revealed in the human cortical blind-spot representation. *Nature, 411*(6834), 195-199.

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron, 21*(4), 753-759.

Tootell, R. B., Hadjikhani, N. K., Vanduffel, W., Liu, A. K., Mendola, J. D., Sereno, M. I., et al. (1998). Functional analysis of primary visual cortex (V1) in humans. *Proc Natl Acad Sci U S A, 95*(3), 811-817.

Tootell, R. B., Tsao, D., & Vanduffel, W. (2003). Neuroimaging weighs in: humans meet macaques in "primate" visual cortex. *J Neurosci, 23*(10), 3981-3989.

Tovee, M. J. (1994). Neuronal processing. How fast is the speed of thought? *Curr Biol, 4*(12), 1125-1127.

Townsend, J. T., Taylor, S. G., & Brown, D. R. (1971). masking for letters with unlimited viewing time. *Percept Psychophys, 10*(5), 375-378.

Tranel, D., Damasio, A. R., & Damasio, H. (1988). Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity. *Neurology, 38*(5), 690-696.

Troje, N. F. (2002). Decomposing biological motion: a framework for analysis and synthesis of human gait patterns. *J Vis, 2*(5), 371-387.

Troje, N. F. (2003). Reference frames for orientation anisotropies in face recognition and biological-motion perception. *Perception, 32*(2), 201-210.

Troje, N. F., & Westhoff, C. (2006). The inversion effect in biological motion perception: evidence for a "life detector"? *Curr Biol, 16*(8), 821-824.

Troje, N. F., Westhoff, C., & Lavrov, M. (2005). Person identification from biological motion: effects of structural and kinematic cues. *Percept Psychophys, 67*(4), 667-675.

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nat Neurosci, 8*(8), 1096-1101.

Tsuchiya, N., Koch, C., Gilroy, L. A., & Blake, R. (2006). Depth of interocular suppression associated with continuous flash suppression, flash suppression, and binocular rivalry. *J Vis, 6*(10), 1068-1078.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale & R. J. W. Mansfield (Eds.), *Analysis of visual behavior*. Cambridge, MA: MIT Press.

Vaina, L. M., Lemay, M., Bienfang, D. C., Choi, A. Y., & Nakayama, K. (1990). Intact "biological motion" and "structure from motion" perception in a patient with impaired motion mechanisms: a case study. *Vis Neurosci, 5*(4), 353-369.

Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P., & Belliveau, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proc Natl Acad Sci U S A, 98*(20), 11656-11661.

Valentine, T. (1988). Upside-down faces: a review of the effect of inversion upon face recognition. *Br J Psychol, 79*(4), 471-491.

Vanrie, J., & Verfaillie, K. (2004). Perception of biological motion: a stimulus set of human point-light actions. *Behav Res Methods Instrum Comput, 36*(4), 625-629.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. *Neuron, 30*(3), 829-841.

Wade, N. J., & Wenderoth, P. (1978). The influence of colour and contour rivalry on the magnitude of the tilt after-effect. *Vision Res, 18*(7), 827-835.

Walk, R. D., & Homan, C. P. (1984). Emotion and dance in dynamic light displays. *Bull Psychon Soc, 22*, 437-440.

Weiskrantz, L. (1996). Blindsight revisited. *Curr Opin Neurobiol, 6*(2), 215-220.

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *J Neurosci, 18*(1), 411-418.

White, K. D., Petry, H. M., Riggs, L. A., & Miller, J. (1978). Binocular interactions during establishment of McCollough effects. *Vision Res, 18*(9), 1201-1215.

Williams, M. A., McGlone, F., Abbott, D. F., & Mattingley, J. B. (2005). Differential amygdala responses to happy and fearful facial expressions depend on selective attention. *Neuroimage, 24*(2), 417-425.

Williams, M. A., Morris, A. P., McGlone, F., Abbott, D. F., & Mattingley, J. B. (2004). Amygdala responses to fearful and happy facial expressions under conditions of binocular suppression. *J Neurosci, 24*(12), 2898-2904.

Wilson, H. R., Blake, R., & Lee, S. H. (2001). Dynamics of travelling waves in visual perception. *Nature, 412*(6850), 907-910.

Winston, J. S., Henson, R. N., Fine-Goulden, M. R., & Dolan, R. J. (2004). fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *J Neurophysiol, 92*(3), 1830-1839.

Wojciulik, E., & Kanwisher, N. (1999). The generality of parietal involvement in visual attention. *Neuron, 23*(4), 747-764.

Wolfe, J. M. (1983). Influence of spatial frequency, luminance, and duration on binocular rivalry and abnormal fusion of briefly presented dichoptic stimuli. *Perception, 12*(4), 447-456.

Wunderlich, K., Schneider, K. A., & Kastner, S. (2005). Neural correlates of binocular rivalry in the human lateral geniculate nucleus. *Nat Neurosci, 8*(11), 1595-1602.

Xu, Y., Liu, J., & Kanwisher, N. (2005). The M170 is selective for faces, not for expertise. *Neuropsychologia, 43*(4), 588-597.

Yang, E., Zald, D. H., & Blake, R. (2007). Fearful expressions gain preferential access to awareness during continuous flash suppression. *Emotion, 7*(4), 882-886.

Young, A. W., Aggleton, J. P., Hellawell, D. J., Johnson, M., Broks, P., & Hanley, J. R. (1995). Face processing impairments after amygdalotomy. *Brain, 118 (Pt 1)*, 15-24.

Young, A. W., McWeeny, K. H., Hay, D. C., & Ellis, A. W. (1986). Matching familiar and unfamiliar faces on identity and expression. *Psychol Res, 48*(2), 63-68.

Young, A. W., Newcombe, F., de Haan, E. H., Small, M., & Hay, D. C. (1993). Face perception after brain injury. Selective impairments affecting identity and expression. *Brain, 116 (Pt 4)*, 941-959.

Yovel, G., & Kanwisher, N. (2004). Face perception: domain specific, not process specific. *Neuron, 44*(5), 889-898.

Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Curr Biol, 15*(24), 2256-2262.

Yu, K., & Blake, R. (1992). Do recognizable figures enjoy an advantage in binocular rivalry? *J Exp Psychol Hum Percept Perform, 18*(4), 1158-1173.

Zimba, L. D., & Blake, R. (1983). Binocular rivalry and semantic processing: out of sight, out of mind. *J Exp Psychol Hum Percept Perform, 9*(5), 807-815.