The processing of bodily and facial expressions with and without visual awareness

Stienen, B.M.C.

Publication date:
2012

Citation for published version (APA):
Stienen, B. M. C. (2012). *The processing of bodily and facial expressions with and without visual awareness*. Ridderprint.

The processing of bodily and facial expressions with and without visual awareness

ISBN: 978-90-5335-505-3

Printed by: Ridderprint BV, Ridderkerk
Cover design by: Nikki Vermeulen

The processing of bodily and facial expressions with and without visual awareness

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan Tilburg University, op gezag van de rector magnificus, prof dr. Ph. Eijlander, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op woensdag 1 februari 2012 om 16.15 uur door

Bernardus Maria Christiaan Stienen,

geboren op 28 juni 1980 te Arnhem.

PROMOTOR

prof. dr. B. de Gelder

PROMOTIECOMMISSIE

prof. dr. S.M.A. Aglioti

dr. J. Jolij

prof. dr. V.A.F. Lamme

prof. dr. J.H.M. Vroomen

prof. dr. P. de Weerd

Table of contents

1. Introduction

Social interaction depends on the processing of a multitude of signals carrying information about the emotional state of others. This information ranges from facial and bodily expressions to prosody in the voice, an offensive touch, or even the scent of someone.

In previous decades a number of research reports focused on the processing of faces and their expressions in order to explore how we process emotions. More recently, researchers have started to investigate the issue of bodily expression recognition. By switching this stimulus category, we may be capable of extending the scope of face-based research and provide evidence that human emotion theories may be generalized to other affective signals as well (de Gelder, 2006, 2009). Results from a number of behavioral experiments using independent stimulus sets now persuade us to conclude that recognition of expressions is similar for face and body stimuli. Available literature has already firmly established that emotional bodily expressions clearly and rapidly convey the emotional, intentional and mental state of a person (Meeren, van Heijnsbergen, & de Gelder, 2005; Stekelenburg & de Gelder, 2004) and that full awareness of the visual stimulus is not essential (de Gelder, Vroomen, Pourtois, & Weiskrantz, 1999; Tamietto et al., 2009; Tamietto & de Gelder, 2010).

In this thesis several studies are described on the recognition of bodily expressions, the relation of this recognition with visual awareness and whether or not visual awareness is necessary for the integration of affective auditory and visual information to occur. In addition, studies are described which examine the contributions of several cortical regions to the perception of bodily expressions, with and without awareness.

In the next section I give a brief overview on what is known about the processing of affective visual stimuli without awareness especially in the context of facial expressions and explain the methods used in this thesis to explore this issue in the domain of bodily expressions. I then describe the neurological underpinnings of unconscious processing of emotional stimuli in general and review the literature involving bodily expressions in detail. I conclude by giving an outline of the remainder of this thesis.

1.1 How to measure processing without visual awareness in healthy participants?

In recent decades a number of research reports have concluded that emotional information can be processed without observers being aware of it (Barrett, Ochsner, & Gros, 2007; Kunst-Wilson & Zajonc, 1980). Many studies using facial expressions now provide both direct and indirect evidence for visual discriminations of affective stimuli in the absence of visual awareness of the stimulus (de Gelder, Vroomen et al., 1999; Dimberg, Thunberg, & Elmehed, 2000; Esteves, Dimberg, & Öhman, 1994; Jolij & Lamme, 2005; Tamietto et al., 2009). On forced choice tasks, it has been shown that clinically blind hemianopic patients can reliably guess the emotion of facial expressions presented in their blind field (de Gelder, Vroomen et al., 1999). This is referred to as affective blindsight and can be seen as an absolute dissociation between what can be detected and what is consciously being seen. This phenomenon however, has proven very difficult to replicate in healthy participants (Robichaud & Stelmach, 2003).

Backward masking is one of the most widely used techniques for exploring the processing of visual emotional information without awareness in neurologically intact observers. Esteves and Öhman (1993) found that short duration (e.g. 33 ms) presentations of facial expressions (happy and angry), replaced immediately by a neutral face (mask) with a longer duration (e.g. 50 ms), are below the participants' identification threshold. Esteves, Parra, Dimberg and Öhman (1994) reported that participants, prevented from conscious recognition of conditioned angry faces by backward masking, still showed elevated skin conductance response to these faces, while Esteves, Dimberg and Öhman (1994) found that this response could not be conditioned when happy faces were used. Dimberg, Thunberg and Elmehed (2000) used EMG to show that participants respond to happy and angry faces with corresponding specific muscles in the face while not being conscious of the presentation of the faces. However, Milders, Sahraie and Logan (2008) showed that the minimum presentation time and SOA required to present facial expressions that are below participants' identification threshold, is dependent on the mask and chosen criterion of the threshold, and found that shorter minimum presentation times rendered facial expressions invisible than are used by, for example, Esteves and Öhman (1993) (see also Wiens, 2006).

A critical issue in many backward masking experiments concerns the measure for visibility or visual awareness of the target. Most often this is assessed in a separate posttest session or after each block rather than on a trial by trial basis. This clearly complicates the interpretation of masking studies because visibility of the target co-varies with the performance on each target presentation. Yet it is possible to combine detection measurements with confidence ratings on a trial-by-trial basis. This provides insight in how the actual detection performance relates to the confidence of this detection and thus visibility of the targets. Lau and Passingham (2006) performed an elegant masking study based on this idea. They presented their participants with masked diamonds and squares and asked them on each trial to identify the target and, then to indicate whether they had seen the target. They parametrically varied the onset between target and mask. This method provided information about whether the participant was aware of the presence of a stimulus on a trial by trial basis and controls for the possibility that participants are likely to be more aware of the stimulus in the longer SOA trials. Lau and Passingham (2006) coined the term "relative blindsight" to refer to two SOA conditions where participants were performing equally in the identification task but differed in reporting whether they had seen the target or not. An advantage of the method is that it is not required to render a target completely invisible, hence overcoming the critical concerns identified by several authors (e.g. Milders et al., 2008; Pessoa, 2005; Wiens, 2006). I will report a study in which the method of Lau and Passingham (2006) is implemented in a parametric masking design in which bodily expressions are masked to explore the dependence of processing bodily expressions on visual awareness.

According to the definition of the 'objective' criterion, observers are perceptually unaware of a target when they perform at chance in a forced choice recognition task. Using the 'subjective' criterion, participants are unaware of the stimulus when they claim not to be able to discriminate perceptual information at better than chance level. This approach was developed by Cheesman and Merikle (1986) as a measure of the phenomenological experience of the participants' perception of the targets. Dissociation or a lack of covariance between the two measures could indicate that the processing of a stimulus is independent or less dependent on visual awareness.

Another method used in this thesis to explore the relation between the processing of bodily expressions and visual awareness is called binocular rivalry (BR). BR refers to the phenomenon of perceptual alternation when two incompatible stimuli are presented to the fovea of each eye separately at the same time. This perceptual alternation can be biased by factors such as differences in contrast, brightness, movement and density of contours (Blake & Logothetis, 2002). Given certain parameters the two stimuli compete with each other for perceptual dominance. This method seems to be perfectly suitable to contrast two classes of stimuli to test our processing sensitivity to either stimulus and to address the question of how emotion modulates this sensitivity.

This phenomenon is especially interesting in the domain of consciousness and the search for its neural correlates because the alternating conscious percept, while the physical stimulation remains stable, will most likely correlate with fluctuating neural activity (Blake & Logothetis, 2002). Blake and Logothetis (2002) also review evidence suggesting that, for example, brighter stimuli as opposed to dimmers ones alter the duration that these stimuli are suppressed rather than the durations of dominance. The "stronger" rivaling stimuli are suppressed for shorter durations, which results in dominating the visual percept as opposed to "weaker" rivaling stimuli.

Previous BR studies have shown that the meaningfulness of the stimulus influences the rivalry pattern as well (e.g. Yu & Blake, 1992). Subsequent studies have used BR to investigate dominance between faces expressing different emotions (Alpers & Gerdes, 2007; Bannerman, Milders, de Gelder, & Sahraie, 2008; Yoon, Hong, Joormann, & Kang, 2009) and found that emotional faces dominate over neutral faces. In an fMRI study Tong, Nakayama, Vaughan, & Kanwisher (1998) showed that the fusiform face area (FFA), a category specific brain area for processing faces (Haxby et al., 1994), is activated with the same strength as when the faces were presented in a nonrivalrous condition. However, to date no BR experiments or masking experiments were conducted using bodily expressions or have investigated the contribution of information from both the face and the body and how this influences our conscious percept.

1.2 Neurological processing of emotional stimuli without awareness

Full awareness of the visual stimulus or intact visual cortex seems not to be essential for processing facial and bodily expressions (de Gelder, Vroomen et al., 1999; Tamietto et al., 2009; Tamietto & de Gelder, 2010). In the case of binocular rivalry using fMRI, studies showed that suppressed images of fearful faces still activated the amygdalae (Pasley, Mayes, & Schultz, 2004; Williams, Morris, McGlone, Abbott, & Mattingley, 2004).

Theoretical models have been advanced arguing that partly separate and specialized pathways may sustain emotional perception, with or without awareness (LeDoux, 1996; Morris, Friston et al., 1998; Morris, Öhman, & Dolan, 1998; Panksepp, 2004; Tamietto et al., 2009; Tamietto & de Gelder, 2010). When visual signals are prevented from being processed by the cortical mechanisms via the striate cortex, the so-called colliculo-thalamo-amygdala pathway could still process the stimulus. These phylogenetically ancient brain structures are shown to be active in normal conscious perception as well as unconscious perception of emotional stimuli (Morris, Öhman et al., 1998; Tamietto et al., 2009; Tamietto & de Gelder, 2010; van de Riet, Grezes, & de Gelder, 2009; Whalen et al., 1998).

The visual and emotional parallel processing pathways have been recently reviewed extensively by Tamietto and de Gelder (2010) and are illustrated in Figure 1.1. The primary visual cortex (V1) receives projections from the lateral geniculate nucleus (LGN) of the thalamus which gets its input directly from the retina (Lamme & Roelfsema, 2000). The pathway splits roughly into two parallel pathways via cortico-cortical connections; the dorsal stream via occipitoparietal connections and the ventral stream via occipitotemporal connections (Ungerleider & Haxby, 1994). In addition, a parallel visual pathway is formed by a minority of connections that project to the superior colliculus (SC) and the pulvinar. The SC and the LGN are interconnected and project to the dorsal stream, and the SC and the pulvinar are interconnected and project to extrastriate cortex, both bypassing

V1 (Lamme & Roelfsema, 2000; Lyon, Nassi, & Callaway, 2010; Schiller & Malpeli, 1977). The key subcortical structures of the visual subcortical pathway consist of the SC, the pulvinar and the amygdala. The SC is connected with the amygdala via the pulvinar (Day-Brown, Wei, Chomsung, Petry, & Bickford, 2010) and the amygdala is connected directly with various areas in the cortex such as the orbitofrontal cortex (OFC), the anterior cingulated cortex (ACC), the posterior superior temporal cortex (pSTS), the premotor cortex, and the striate and extrastriate cortex. (Amaral, Behniea, & Kelly, 2003; Amaral & Price, 1984, Avendano, 1983 #14142; Krettek & Price, 1977; McDonald, 1991). It is important to note that evidence for the connections is mainly based on research on Macaque monkeys, rats and cats. However, numerous neuroimaging studies report co-activation among, for example, the superior colliculus, pulvinar and the amygdala (Tamietto & de Gelder, 2010).

The special status of fear stimuli is still a matter of debate, specifically in relation to the role of the amygdale (Duncan & Barrett, 2007; Pessoa, 2005) and the specific contribution of the different subcortical and cortical pathways (de Gelder, van Honk, & Tamietto, 2011; Pessoa & Adolphs, 2010). Pessoa and Adolphs (2010) propose an alternative model in which the main role of the amygdala is the coordination of the functions of cortical networks when processing affective visual stimuli. However, Shewmon, Holmes and Byrne (1999) showed that patients with congenital absence of the cerebral cortex still exhibit affective responses (de Gelder et al., 2011). The cortex is obviously involved in processing emotional stimuli, but seems not to be necessary for the processing of emotional stimuli without awareness. Although anatomical connections exist between, for example, the superior colliculus and the motion sensitive cortical area MT, this pathway is not shown to be involved in non-conscious emotion perception (de Gelder et al., 2011).

Neurological evidence is appearing that masking selectively disrupts re-entrant signals to V1. For example, Lamme, Zipser and Spekreijse (2002) showed that masking might selectively interrupt the recurrent interactions between higher visual areas and V1 in the macaque monkey brain. Lamme and Roelfsema (2000) propose two distinct modes of vision that consist of feedforward and recurrent processing. The authors propose that these two modes are crucial in the distinction between non-conscious and conscious vision respectively.

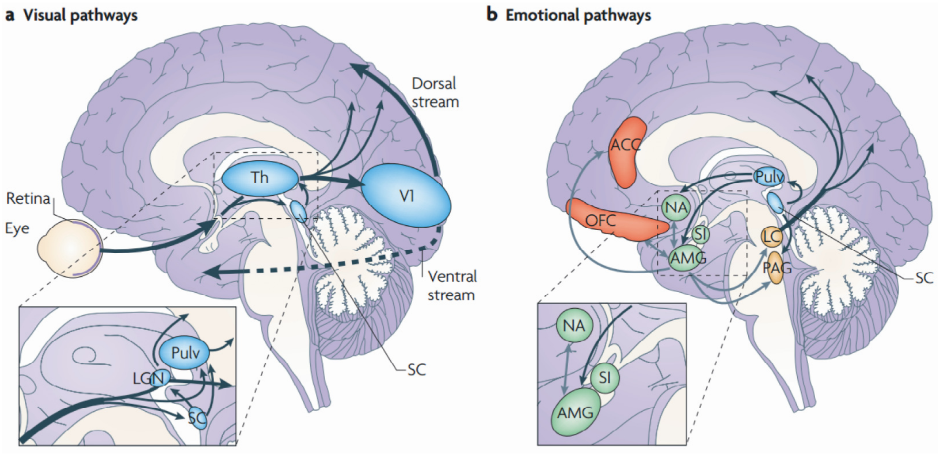


Figure 1.1 *(**a**) The (cortical) visual pathways. The primary visual cortex (V1) receives projections from the lateral geniculate nucleus (LGN) of the thalamus (Th) which gets its input directly from the retina. The pathway splits roughly into two parallel pathways; the dorsal stream via occipitoparietal connections and the ventral stream via occipitotemporal connections. In addition, a parallel visual pathway is formed by a minority of connections that project to the superior colliculus (SC) and the pulvinar (Pulv). The SC and the thalamus are interconnected and project to the dorsal stream; the SC and the pulvinar are interconnected and project to extrastriate cortex both bypassing V1. (**b**) The emotion perception pathways. The SC is connected with the amygdala via the pulvinar and the amygdala is connected directly and via brainstem nuclei (locus coeruleus (LC) and periaqueductal gray (PAG)) with various areas in the cortex like the orbitofrontal cortex (OFA), the anterior cingulate cortex (ACC), and the posterior superior temporal cortex (pSTS). The latter is not shown in this figure. Figure 1.1 is reproduced, with permission, from Tamietto and de Gelder (2010).*

1.3 Processing of bodily expressions

Available literature has already firmly established that bodily expressions clearly and rapidly convey the emotional, intentional and mental state of a person (Meeren et al., 2005; Stekelenburg & de Gelder, 2004) and that full awareness of the visual stimulus or intact visual cortex are not essential (de Gelder, Vroomen et al., 1999; Tamietto et al., 2009; Tamietto & de Gelder, 2010). Tamietto et al. (2009) showed that hemianopic patients could still categorize bodily expressions reliably better than chance when pictures of bodily expressions were presented in their blind field. In addition, they showed that the patients contracted the muscles in their face used for expressing certain emotions that mimic the expression of the body that was presented in their blind field. However, the independence of detection and processing of bodily expressions on visual awareness is only demonstrated in hemianopic patients so far.

Some emotions are better conveyed by the body than by the face and vice versa. For example, although one can show anger by frowning or wrinkling the brows, the tension in the body muscles would indicate better the strength of the anger or the intention of the angry person. Disgust is an example where the facial expression is very specific while the body posture associated with disgust is less so since it shares features with the display of fear. When we add this kind of emotion specificity to the overall picture it emerges that, depending on the emotion, we can consider the primary sensory channel to be the face, the whole body, or the voice.

Another major difference between facial and bodily expressions is that the latter can be recognized from far away while the former requires the viewer to be nearby. This is potentially an important difference between how facial and bodily expressions play their communicative roles and it should have consequences for how the information is processed (de Gelder, 2006, 2009).

Facial expressions are often recognized less than perfectly, but bodily expressions that are emotionally congruent with the facial expression shown at the same time, improve accuracy of facial recognition while incongruent bodily information significantly hampers it (Meeren et al., 2005; Van den Stock, Righart, & de Gelder, 2007). However, the relative importance of facial and bodily signals and their relation to visual awareness is still poorly understood.

Recent studies have shown that in addition to facial expressions, perception of bodily expressions is influenced by concurrent auditory information and that affective information in sounds modifies the viewers' appreciation of the affective body image. For example, recognition of dynamic whole-body expressions of emotion are influenced not only by both human and animal vocalizations (Van den Stock, Grèzes, & de Gelder, 2008), but also by instrumental music (Van den Stock, Peretz, Grèzes, & de Gelder, 2009), suggesting that the brain is efficient at extracting affective information from different sources and combining it across different sensory channels.

Category specific brain areas have been identified for faces and bodies in the visual cortex as well as in the fusiform gyrus, part of the temporal cortex. In the visual cortex the occipital face area (OFA) and the extrastriate body area (EBA) showed a selective response to faces and bodies respectively (Allison, Puce, & McCarthy, 2000; Downing, Jiang, Shuman, & Kanwisher, 2001; Grossman & Blake, 2002; Puce, Allison, Asgari, Gore, & McCarthy, 1996; Spiridon, Fischl, & Kanwisher, 2006; van de Riet et al., 2009). High resolution imaging revealed spatial segregation of face and body sensitive areas in the fusiform gyrus (FG), specifically the fusiform face area (FFA) and the fusiform body area (FBA) respectively (Haxby et al., 1994; Kanwisher, McDermott, & Chun, 1997; Peelen & Downing, 2005; Schwarzlose, Baker, & Kanwisher, 2005; Taylor, Wiggett, & Downing, 2007; Tong et al., 1998). One suggestion from this line of work is that a substantial part of the processing of faces and bodies is separate and devoted to visual stimulus categorization using the ventral processing stream.

Modulation by bodily expressions of the fusiform gyrus was found by Hadjikhani and de Gelder (2003) in addition to modulation by facial expressions (Dolan, Morris, & de Gelder, 2001; Morris, Öhman et al., 1998; Rotshtein, Malach, Hadar, Graif, & Hendler, 2001). The posterior superior temporal sulcus (pSTS) is activated by both facial and bodily expressions (Allison et al., 2000; LaBar, Crupain, Voyvodic, & McCarthy, 2003; Pichon, de Gelder, & Grezes, 2009; van de Riet et al., 2009). The elevated activation levels of pSTS and FG might be explained by the connections of these areas with the amygdala (Morris, Friston et al., 1998; Rotshtein et al., 2001; Vuilleumier, Richardson, Armony, Driver, & Dolan, 2004).

Consistent with this, in studies with macaque monkeys using anterograde tracers, connections are found between amygdala and all levels of visual cortex, as well as to the temporal lobe (Amaral et al., 2003; Amaral & Price, 1984). However, what is not clear yet is whether the pSTS is playing a differential role in processing the facial and bodily expressions.

Oram and Perrett (1996) made the interesting observation that STPa, an area in the macaque brain within the rostral superior temporal cortex, deals with the integration of form and motion. This role is also hinted at by the cortico-cortical connections between STS with ventral and dorsal areas (Ungerleider & Haxby, 1994). Monkey studies show that there are extensive reciprocal connections between STS and inferotemporal cortex (IT) (see also Allison et al., 2000 for a review on the role of STS in visual perception; Sugase, Yamane, Ueno, & Kawano, 1999; Suzuki & Amaral, 1994).

Human fMRI and neurophysiologic studies revealed that neurons in the posterior part of the STS respond to a wide variety of socially relevant stimuli such as gaze and mouth movements (Puce, Allison, Bentin, Gore, & McCarthy, 1998), facial expressions (Haxby, Hoffman, & Gobbini, 2000), actions (Decety & Grezes, 1999), biological motion (Puce & Perrett, 2003) and bodily expressions and movements (Candidi, Stienen, Aglioti, & de Gelder, 2011; de Gelder, 2006; de Gelder & Partan, 2009; Grèzes, Pichon, & de Gelder, 2007; Kret, Pichon, Grèzes, & de Gelder, 2011; Pichon, de Gelder, & Grezes, 2008). Therefore, it has been suggested that pSTS is mainly involved in processing the intentions of the observed agent (Allison et al., 2000; Ghazanfar & Santos, 2004). The differential adaptive function of perceiving and responding to facial and bodily emotional expressions may be reflected in a dissociation of the links between pSTS and emotion and action-related networks. Thus, inhibition of temporal cortical regions may be reflected in different perceptual effects according to whether the emotion was conveyed via the face or the body.

The left ventral premotor (vPM) cortex is thought to be crucial for action representation, and understanding and is generally considered to be a key node of the "mirror neuron" system initially described in monkeys (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992) and consisting of bimodal visuo-motor cells that fire during real action execution as well as during the passive observation of the same movement

(Kilner, Neal, Weiskopf, Friston, & Frith, 2009). The activation of premotor regions is facilitated by the emotional valence of the observed posture (de Gelder, Snyder, Greve, Gerard, & Hadjikhani, 2004; Grèzes et al., 2007; Pichon et al., 2008), strengthening the notion that emotion perception is tightly linked to action programming (Darwin, 1872/1965). Consistent with this, it has been shown that monkeys' premotor cortex receives neural projections from the amygdaloid complex (Avendano, Price, & Amaral, 1983) and that electrical stimulation of this area induces defensive motor behaviors (Cooke & Graziano, 2004; Graziano, Taylor, & Moore, 2002).

1.4 Outline of this thesis

This thesis studies the processing and detection of bodily expressions without awareness in healthy humans. Chapters 2 – 4 consist of behavioral studies, chapter 5 reviews a behavioral experiment in which performance of human participants on a masking experiment is compared with the performance of a computational neural model, and chapter 6 and 7 report two studies using Transcranial Magnetic Stimulation (TMS). This multidisciplinary approach is chosen in order to explore the processing of bodily and facial expressions and its relation to visual awareness from as many different perspectives as possible.

Chapter 2 reports on a study investigating the dependence of detection of bodily expressions on visual awareness. In this study a parametric masking technique is used in combination with confidence ratings on a trial by trial basis. Participants were asked to detect, in three separate experiments, masked fearful, angry and happy bodily expressions among masked neutral bodily actions as distractors and subsequently to indicate their confidence.

To better understand the relative contribution of affective signals from the face only or from the rest of the body we used binocular rivalry experiments which are described in chapter 3. The role of emotional modulation on the rivalry pattern is central here.

Chapter 4 describes two studies in which the masking technique is combined with the presentation of auditory emotional information to explore multisensory integration of body-voice pairs and whether visual awareness is necessary for the integration to occur.

In chapter 5 the performance of a computational neural model, which exclusively modeled feed-forward processing and was engineered to fulfill the computational requirements of recognition, is compared with the performance of participants on a parametric masking task in which bodily expressions were being masked. This was done to gain insight as to whether cortical feed-forward mechanisms only can account for the recognition of masked bodily expressions.

Chapter 6 presents a TMS study in which participants had to detect small postural changes in human bodies and animals which could be either threatening or neutral. Several brain areas were stimulated (EBA, vPM, and pSTS) in order to answer the question of whether these brain areas are critical to detect postural changes and how this is related to other processes in which those brain areas are involved.

In Chapter 7 a study is described in which binocular rivalry was combined with offline rTMS in order to induce a transient lesion in the right pSTS. The binocular rivalry experiment contrasted bodily and facial expressions with houses. As pSTS is involved in the processing of social stimuli, the difference between the processing of facial and bodily expressions after pSTS stimulation was of interest.

In the final chapter I summarize and attempt to compare or combine all findings and discuss their implications in the context of the parallel cortical and subcortical visual pathways.

2. Fear detection and visual awareness in perceiving bodily expressions

Abstract

Many research reports have concluded that emotional information can be processed without observers being aware of it. The case for perception without awareness has almost always been made with the use of facial expressions. In view of the similarities between facial and bodily expressions for rapid perception and communication of emotional signals, we conjectured that perception of bodily expressions may also not necessarily require visual awareness. Our study investigates the role of visual awareness in the perception of bodily expressions using a masking technique in combination with confidence ratings on a trial by trial basis. Participants had to detect in three separate experiments masked fearful, angry and happy bodily expressions among masked neutral bodily actions as distractors and subsequently the participants had to indicate their confidence. The onset between target and mask (Stimulus Onset Asynchrony, SOA) varied from -50 to +133 ms. Sensitivity measurements (d-prime) as well as the confidence of the participants showed that the bodies could be detected reliably in all SOA conditions. A lack of covariance was observed between the objective and subjective measurements when the participants had to detect fearful bodily expressions while this was not the case when participants had to detect happy or angry bodily expressions.

Stienen, B. M. C., & de Gelder, B. (2011). Fear detection and visual awareness in perceiving bodily expressions. *Emotion, 11*(5), 1182-1189

2.1 Introduction

In recent decades a number of research reports have concluded that emotional information can be processed without observers being aware of it (Barrett et al., 2007; Kunst-Wilson & Zajonc, 1980). Many studies using facial expressions now provide both direct and indirect evidence for visual discriminations of affective stimuli in the absence of visual awareness of the stimulus (de Gelder, Vroomen et al., 1999; Dimberg et al., 2000; Esteves, Dimberg et al., 1994; Jolij & Lamme, 2005; Tamietto et al., 2009). On forced choice tasks, it has been shown that not only neurologically intact but also clinically blind hemianope patients can reliably guess the emotion not only of facial but also of bodily expressions presented in their blind field (de Gelder, Vroomen et al., 1999; Tamietto et al., 2009). This finding in patients can be seen as an absolute dissociation between what can be detected and what is consciously being seen. However, this phenomenon has proven difficult to replicate in healthy participants (Robichaud & Stelmach, 2003). The present study investigates the role of visual awareness in the perception of bodily expressions using a masking technique combined with confidence ratings.

Backward masking is one of the most widely used techniques for exploring the processing of visual emotional information without awareness in neurologically intact observers. Esteves and Öhman (1993) found that short duration (e.g. 33 ms) presentations of facial expressions (happy and angry), replaced immediately by a neutral face (mask) with a longer duration (e.g. 50 ms), are below the participants' identification threshold. Esteves, Parra, Dimberg and Öhman (1994) reported that participants, prevented from conscious recognition of conditioned angry faces by backward masking still showed elevated skin conductance response to these faces, while Esteves, Dimberg and Öhman (1994) found that this response could not be conditioned when happy faces were used. Dimberg, Thunberg and Elmehed (2000) used EMG to show that participants respond to happy and angry faces with corresponding specific muscles in the face while not being conscious of the presentation of the faces.

A critical issue in many backward masking experiments concerns the measure for visibility or visual awareness of the target. Most often this is assessed in a separate posttest session or after each block rather than on a trial by trial basis.

This clearly complicates the interpretation of masking studies because visibility of the target co-varies with the performance on each target presentation. Yet it is possible to combine detection measurements with confidence ratings on a trial-by-trial basis. This provides insight in how the actual detection performance relates to the confidence of this detection and thus visibility of the targets. Lau and Passingham (2006) performed an elegant masking study based on this idea. They presented their participants with masked diamonds and squares and asked them on each trial to identify the target and, then to indicate whether they had seen the target. The onset between target and mask (Stimulus Onset Asynchrony, SOA) varied from minus 50 to 133 ms. This method provided information about whether the participant was aware of the presence of a stimulus on a trial by trial basis and controls for the possibility that participants are likely to be more aware of the stimulus in the longer SOA trials. Lau and Passingham (2006) coined the term "relative blindsight" to refer to two SOA conditions where participants were performing equally in the identification task but differed in reporting whether they had seen the target or not.

We adopted this approach to investigate the relation between detection performance and confidence. In three experiments participants had to detect masked and unmasked emotional expressions (fear, angry or happy) among masked and unmasked distractors (a neutral action; combing). The pictures and the mask were controlled for several factors such as lighting, size of the postures on the retina, contrast, and the actors were uniformly dressed in black clothing. A mask was presented at 12 different SOAs varying from minus 50 to 133 ms. The participants were instructed to detect the emotional expression and subsequently to indicate whether they were sure or whether they were guessing. The different emotional expressions were not mixed within one design to prevent that dominant or more salient emotional expressions would influence the percept of the other emotions.

According to the definition of the 'objective' criterion, observers are perceptually unaware of a target when they perform at chance in a forced choice recognition task. Using the 'subjective' criterion, participants are unaware of the stimulus when they claim not to be able to discriminate perceptual information at better than chance level (Cheesman & Merikle, 1984). In this experiment the detection rates are used as the objective measurements, while the confidence ratings are used as the subjective measurements. In line with Lau and Passingham (2006) we expected to find relative dissociations between the two measurements. Because we used a pattern mask it is expected that the lowest detection performance and confidence will be around the SOA of 0 ms and will be U-shaped (Enns & Di Lollo, 2000). Following Lau and Passingham (2006) we conjectured that this U-shape implies that we can find SOA conditions where the detection performance is the same. We are specifically interested how this detection performance relates to the confidence of the participants. Based on reports in the literature (Esteves, Dimberg et al., 1994; Morris et al., 1996; Pessoa, Japee, & Ungerleider, 2005; Vuilleumier et al., 2004) we also conjectured that the resulting patterns may be dependent on the specific emotional category.

EXPERIMENT 1

Our goal is to investigate the relation between objective (detection performance) and subjective (confidence) measurements of the perception of fearful bodily expressions.

2.2 Method

Participants

Twenty-three undergraduate students of Tilburg University participated in exchange of course credits or a monetary reward (16 women, 7 men, M = 19.8 years, SD = 2.3). All participants had normal or corrected-to-normal vision and gave informed consent according to the declaration of Helsinki.

Stimuli

Photos of actors expressing fear and combing their hair were selected from a photoset. During the photo shoot pictures were projected on the wall facing the actor meant to trigger the fear response as natural as possible. Moreover, a short emotion inducing story related to the image projected was told by the experimenter. For the combing pictures the actors were asked to pretend that they had a comb in their hands and that they were straightening their hairs.

The faces of the selected photographs were covered with an opaque oval patch to prevent that the facial expression would influence the identification of the emotional body expression. The color of the patch was the average grey value of the neutral and emotional face within the same actor. In addition, the colors were saturated to white and black with the color of the mask as anchor point. In this way, the participants were forced to base their judgments on the contours of the body because by isolating only two colors the color differences within the clothing disappeared. A total of 16 pictures (2 fear/combing x 2 gender x 4 actors) were selected for use in the present study. Average height of the bodies was 7.78 degrees, the average maximum width (distance between the hands) was 2.83 degrees and the average waist was 1.39 degrees of visual angle.

Using Adobe Photoshop 7.0 © a pattern mask was constructed by cutting the target bodies in asymmetric forms which were scrambled and replaced in the area occupied by the bodies. The parts were grouped with the restriction that parts containing white had to be grouped within the area occupied by the hands (which were saturated to white) and parts containing black had to be grouped within the area occupied by the bodies (which were black). Finally, the resulting picture was duplicated, rotated 180 degrees and pasted at the background to induce symmetry and extra noise to avoid the percept of a body. The result is the mask in Figure 2.1. The height of the mask was 9.85 degrees and the width was 6.48 degrees of visual angle. The mask covered the area of the stimuli completely.

The stimuli were presented on a 17” PC screen with the refresh rate set to 60 Hz. We used Presentation 11.0 to run the experiment. A white cross of 1.22 x 1.22 degrees was used as a fixation mark in the center of the screen. Finally, all stimuli were pasted on a gray background.

The SOA latencies were -50, -33, -17, 0, 17, 33, 50, 67, 83, 100, 117 and 133 ms. The actual presentation time was calibrated with the use of a photodiode and an oscilloscope measuring the latency between onset of the target and the mask. Negative SOA latencies represent forward masking and positive SOA latencies backward masking. When the SOA was -33, -17, 0 and 17 ms the target overlaps with the mask. The target was always presented at the foreground. Moreover a target-only condition and a mask-only condition were included. One complete run consisted of 224 trials (8 identities x 2 actions (fear/combing) x 14 timing conditions (including target-only and mask-only) which were randomly presented.

Procedure

Participants were comfortably seated in a chair in a soundproof experimental booth approximately 90 cm from the screen. A trial started with a white fixation cross on a gray background. The disappearance of this cross signaled the beginning of a trial. After 500 ms the target stimulus appeared for 33 ms. Next, a mask was presented for 50 ms after a variable interval (sometimes the mask was presented first). The participants were instructed to push a predefined button using the index finger of their left hand as soon as they thought a fearful bodily expression was presented (GO) and to withhold their response when they thought the neutral action was presented (NO-GO). Two thousand ms after the target a screen was presented with the text “Sure or Guessed?”. They had to respond with the other hand with two different buttons on the same response box labeled with “Sure” and “Guessed”. The latter two buttons were counterbalanced across participants. It was stressed that they had to respond as accurate and fast as possible and that they could use their “gut feeling” if they did not have seen the body. Finally a gray screen was presented with a random duration between 17 ms and 767 ms. This jitter was added to prevent that the participants would be caught in a mechanical rhythm. In total the trials were on average 4025 ms.

Prior to the experimental sessions the participants performed two practice sessions consisting of 33 trials each (16 target-trials, 16 distracter-trials, and 1 mask-only trial). Other identities than the ones used in the main experiment served as targets. When the participants had more than 12 hits and gave notice of a full understanding of the procedures the main experiment was started. A total of four runs were presented adding up to a total of 896 trials. Every 112 trials there was a 3 minute break. After the main experiment all targets were presented for 33 ms to validate the stimuli used as targets. The instruction remained the same for this session.

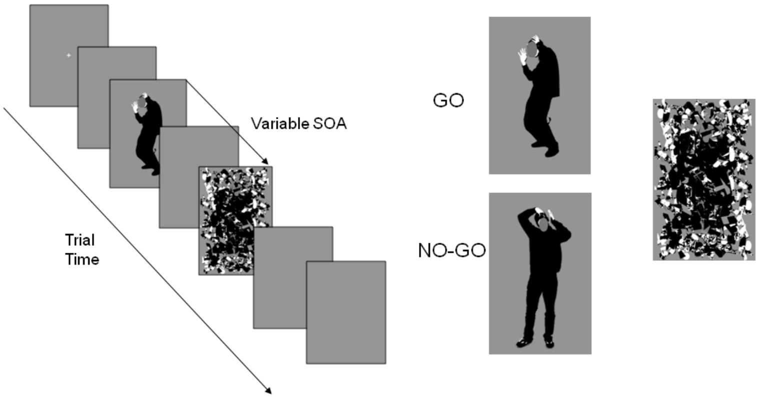


Figure 2.1 *A visual representation of a trial (**left**), example of a fearful bodily expression and a neutral bodily posture (**middle**), and the mask (**right**).*

Analysis

Trials where participants failed to indicate their confidence within the duration of the trial were discarded.

The sensitivity to the signal (detection of expressions) was estimated by calculating the d-prime (d’). The d’ is a measure for the distance between the signal and noise distribution means in standard deviation units (Green & Swets, 1966). A d’ of zero means that the participants are not able to discriminate the fearful bodily expressions from the neutral bodily actions. The d’ was calculated as:

$$d' = \Phi^{-1}(H') - \Phi^{-1}(FA')$$

Where H’ is the corrected hit rate and FA’ is the corrected false alarm rate. The function Φ^{-1} converts the rates into z-scores. The correction of the hit- and false alarm rates was performed to protect against ceiling effects as proposed by Snodgrass and Corwin (1988):

$H' = (h + 0.5) / (h + m + 1)$

$FA' = (f + 0.5) / (f + cr + 1)$.

Where *h* is the number of hits, *m* is the number of misses, *f* is the number of false alarms and *cr* is the number of correct rejections. See also Tamietto, Geminiani, Genero, and de Gelder (2007).

To assess whether participants could differentiate between the correct and incorrect answers confidence ratings were calculated. The number of sure responses when the detection of the emotional expression or the rejection of the neutral action was incorrect was subtracted from the number of sure responses when the response was correct. This was divided by the total number of correct and incorrect answers. A resulting value of zero would mean that the participants indicate subjectively that they are not more confident of their correct answers than their incorrect answers which is taken as a measure of subjective visual awareness. A similar approach was chosen by Cheesman and Merikle (1986) and Esteves and Öhman (1993) as a measure of the phenomenological experience of the participants' perception of the targets.

We used this calculation because of the analogue with the d-prime. This means that information from all four cells (hits, misses, false alarms and correct rejections) were used. In addition, this method automatically controls for how well the participants are engaged in the task. If, for example, a participant would just randomly hit the detection button, but always indicates to be sure, the confidence measure when calculated as overall percent sure would end up being 100 percent while the d' would not be higher than zero (for more details on the d' analysis see Macmillan and Creelman (1991). However, our measure of confidence would also result in a confidence rating of zero, because it automatically corrects for when the participants indicate to be sure when their answer is wrong.

Separate multivariate analyses of variance (MANOVA) were performed on d' values and confidence ratings with SOA (13 levels including target only trials) as a factor. Following Lau and Passingham (2006) the analysis is focused on the SOA conditions just lower and higher than the SOA condition with the lowest detection performance and confidence ratings. Because the fact that we use a pattern mask it is expected that the lowest detection performance and confidence ratings will be around the SOA of 0 ms (Enns & Di Lollo, 2000). Two SOA points were selected just before the falling edge of the curve and two after the rising edge of the curve[1].

Finally, we computed the area under the ROC curve for the relevant conditions to obtain a measure that does not assume equal variance of the distributions of the signal and the noise. There are four possible responses per stimulus type: 1. detection and sure, 2. detection and guess, 3. no detection and guess, and 4. no detection and sure. These responses were tabulated per stimulus category (emotion or neutral) and divided by the total amount of trials in that category to estimate a conditional probability. Next, we calculated the cumulative probability for each confidence level ranging from detecting an emotion with high certainty to not detecting an emotion with high certainty. Given that the target was an emotion this yields the hit rate, when a neutral action was presented this gives the false alarm rate. For more details see Macmillan and Creelman (1991). The actual graphs are not plotted because in several cases the ROC curves exactly align with each other.

2.3 Results and Discussion

One participant was discarded from analysis because he never indicated to be sure of his responses. In the validation session the fearful bodily expressions were correctly detected 91 percent of the cases (SD = 12) and the neutral action was correctly rejected 99 percent (SD = 3) of the cases, see Figure 2.6.

As shown in Figure 2.2a, the d’ results show a classical pattern masking curve with the lowest point of the curve around SOA of 0 ms (Enns & Di Lollo, 2000). There was a main effect of SOA ($F(12,10) = 26.57$, $p < .001$). The d’ was above zero when the SOA was 0 ms ($t(21) = 9.26$, $p < .001$), indicating that the participants were capable of detecting the fearful bodily expressions. The confidence is plotted in Figure 2.2b. In addition, a main effect of SOA is found here ($F(12,10) = 78.12$, $p <$

---

[1] During the review of this thesis by the committee concerns were raised regarding the statistical procedure. An alternative statistical procedure is discussed in the general discussion on pages 129-130.

.001). Participants were still more confident about their correct than incorrect answers when the d' was at its lowest point. This is indicated by the confidence ratings being still significantly above zero when the SOA was 0 ms (t(21) = 5.97, *p* < .001).

SOA latencies of -50 and -33 ms were just before the falling edge of the curve while SOA latencies of +33 and +50 ms were just after the rising edge of the curve. Planned comparisons showed that detection performance is equal between SOA latencies of -50 and +33, -50 and +50, -33 and +33, and -33 and +50 ms; this was indeed confirmed with paired t-tests showing no significant differences (resp. t(21) = -.58, *p* = .566; t(21) = -.95, *p* = .355; t(21) = -.10, *p* = .924; t(21) = -.37, *p* = .716). However, when performing statistical comparisons between the same SOA latencies on the confidence ratings it appeared that the confidence ratings differed significantly for the SOA pair -50 & +33 ms (t(21) = 2.17, *p* = .042). This was also the case for the comparison of the SOA conditions of -33 & +33 ms; the d' did not differ, while the confidence ratings did (t(21) = 2.23, *p* = .037). Similarly, the area under the ROC curve (A') did not differ significantly for each of these conditions (all *p* > .05). Table 2.1 shows the d' and A' per SOA.

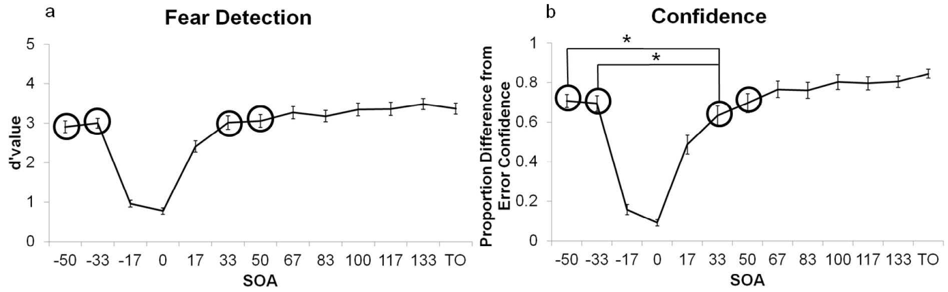


Figure 2.2 *The detection performance and confidence when detecting fearful bodily expressions as a function of SOA. Detecting fearful bodily expressions (**a**) seem to be equal at both sides of the U-shaped curve, while confidence ratings (**b**) seem to differ for SOA pairs -50 and 33 ms and -33 ms and +33 ms. Error bars indicate standard error mean. Asterisks = p < .05.*

While participants are equally capable of detecting the fearful bodily expressions in both SOA conditions, their confidence differed. The dissociation between the objective (what is detected) and subjective measures (the confidence about the detection) indicates that the mechanisms are independent. Lau and Passingham (2006) called this phenomenon relative blindsight.

The fact that we do not find a condition where the confidence ratings are not different from zero (indicating that the participants are guessing) while objective detection of expressions is above zero, does not force the conclusion that there is no processing of the stimulus without awareness. In fact, the relative difference indicates that different processes are at hand causing the subjective ratings to differ while the objective detection performance is on the same level. In experiment 2 and 3 the question is addressed whether this phenomenon generalizes to different emotions or whether it is specific for fearful expressions.

EXPERIMENT 2

In the second experiment we asked whether the observed effect is specific for fear or whether it is driven by negative emotions in general. For this purpose we used angry bodily expressions as targets.

2.4 Method

Participants

Twenty-one undergraduate students of Tilburg University participated in exchange of course credits or a monetary reward (11 women, 10 men, M = 21.8 years, SD = 3.4). All participants had normal or corrected-to-normal vision and gave informed consent according to the declaration of Helsinki.

Stimuli and procedure

The stimuli for this experiment were taken from the same photoset as in the first experiment, but this time we selected actors showing angry bodily expression. The average height of the bodies was 8.25 degrees, the average maximum width (distance between the hands) was 2.75 degrees and the average waist was 1.49 degrees of visual angle. See Figure 2.3 for an example.

The participants were instructed to push a predefined button as soon as they saw an angry bodily expression. The rest of the procedure was the same as in the first experiment.

Figure 2.3 *An example stimulus of an angry expression (**left**) and a happy expression (**right**).*

2.5 Results and Discussion

One participant was discarded from analysis because she failed to answer within time limits of the trials. The angry bodily expressions were correctly detected 94 percent of the cases (SD = 10) and the neutral action was correctly rejected 99 percent (SD = 3) of the cases in the validation session, see Figure 2.6.

Figure 2.4 shows the detection performance and the confidence per SOA. Again, there was a main effect of SOA on the d' and on the confidence ratings (resp. F(12,9) = 35.16, p < .001; F(12,9) = 203.07, p < .001). The lowest point was again when the SOA was 0 ms and also this time not only the detection performance but also the subjective confidence ratings were always above 0 (resp. t(20) = 6.28, p < .001; t(20) = 5.78, p < .001).

The d’ of the SOA latencies -50 and +50 ms and SOA latencies -33 and +33 ms did not differ (resp. t(20) = -.07, *p* = .944; t(20) = 1.71, *p* = .10) while in contrast with what we observed using fearful bodily expressions there was also no difference in the confidence ratings (resp. t(20) = .90, *p* = .381; t(20) = 1.80, *p* = .087). Moreover, the SOA latencies -50 and +33 ms and SOA latencies -33 and +50 ms differed significantly (resp. t(20) = 4.88, *p* < .001; t(20) = --2.88, *p* = .009), but so did the confidence ratings (resp. t(20) = 4.27, *p* < .001; t(20) = -2.68, *p* =.014). A’ values followed this pattern: SOA pair -50 and +50 ms and latency pair -33 and +33 ms did not differ (resp. t(20) = .68, *p* = .505; t(20) = 1.44, *p* = .166) while the SOA latencies -50 and +33 ms and latency pair -33 and +50 ms differed significantly (resp. t(20) = 4.48, *p* < .001; t(20) = -2.83, *p* = .010. Table 2.1 shows the d’ and A’ values per SOA.

In sum, angry bodily expressions can be detected objectively and subjectively better than chance even in the smallest absolute SOA latencies. However, when looking at the same SOA conditions as used with fearful body expressions the objective and subjective measures do not dissociate. This seems to indicate that the dissociation between measures is specific for fearful bodily expressions. The lack of covariance between subjective and objective measurements does not generalize to all negative emotions. The next question then is whether the phenomenon does extent to positive emotions such as happiness.

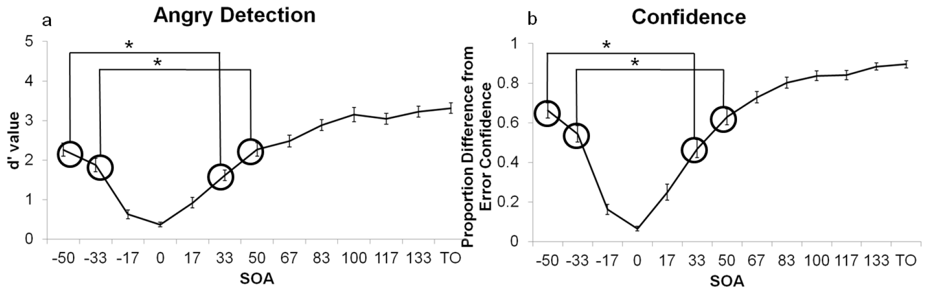


Figure 2.4 *The detection performance and confidence when detecting angry bodily expressions as a function of SOA. Detecting angry bodily expressions (**a**) and confidence ratings (**b**) seem not to dissociate when looking at the same SOA pairs as observed with fearful bodily expressions. Error bars indicate standard error mean. Asterisks = p < .05.*

EXPERIMENT 3

2.6 Method

Participants

Twenty undergraduate students of Tilburg University participated in exchange of course credits or a monetary reward (14 women, 6 men, M = 23.0 years, SD = 4.6). All participants had normal or corrected-to-normal vision and gave informed consent according to the declaration of Helsinki.

Stimuli and procedure

Stimuli consisted of bodily expressions of happiness selected from the same photoset as in the first and second experiment. The average height of the happy bodily expressions was 8.75 degrees, the average maximum width (distance between the hands) was 4.83 degrees and the average waist was 1.65 degrees of visual angle. See Figure 2.3 for an example.

The participants were instructed to push a predefined button as soon as they saw a happy bodily expression. The rest of the procedure was the same as in the first experiment.

2.7 Results and Discussion

In the validation session happy bodily expressions were correctly detected 99 percent of the cases (SD = 3) and the neutral action 100 percent (SD = 0) of the cases, see Figure 2.6.

Figure 2.5 shows the detection performance and the confidence per SOA. Again, there was a main effect of SOA on the d’ and on the confidence ratings (resp. F(12,8) = 27.62, $p < .001$; F(12,8) = 79.52, $p < .001$). The lowest point was again when the SOA was 0 ms. In line with fearful and angry detection not only the detection performance but also the confidence ratings were always above 0 (resp. t(19) = 8.58, $p < .001$; t(19) = 8.76, $p < .001$).

The d’ of the SOA pairs -50 and +50 ms, -33 and +33 ms, and -33 and +50 ms did not differ significantly (resp. t(19) = 1.51, $p = .146$; t(19) = 1.61, $p = .124$; t(19) = -1.07, $p = .298$), however comparisons of the confidence ratings showed the same pattern (resp. t(19) = 1.92, *p* = .072; t(19) = 1.80, *p* = .087; t(19) = -.73, *p* = .474). When comparing the SOA latencies of -50 with +33 ms a significant difference for the d' appeared (t(19) = 4.21, *p* < .001), but this was also found for the confidence ratings (t(19) = 3.90, *p* = .001). Here A' values followed the same pattern, the only SOA pair that significantly differed was -50 and +33 ms (t(19) = 2.48, *p* = .023). Table 2.1 shows the A' and d' values.

The objective and subjective detection of happy and angry bodily expressions show the same pattern, but this is not the case for fear detection. This indicates that the lack of covariance is specific to fearful bodily expressions. To rule out that the differences in the results between the experiments could be accounted for by how well the emotion is recognized a 3 (experiment) x 2 (target, distractor) between subjects ANOVA was done on the detection performance of emotional bodily expressions in the validation session. This showed that there was a main effect of emotion (F(2,57) = 4.44, *p* < .05). Bonferroni corrected post hoc tests showed that only the detection of happy bodily expressions was different from the detection of fearful expressions. However, the validation data did not show a difference between detection of angry and fearful expressions.

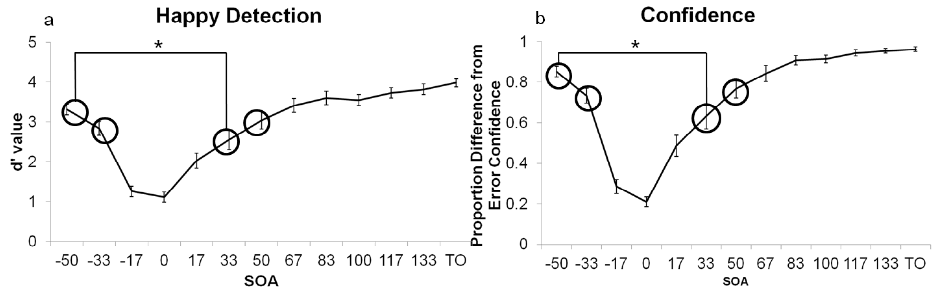


Figure 2.5 *The detection performance and confidence when detecting happy bodily expressions. Detecting happy bodily expressions (****a****) and confidence ratings (****b****) seem not to dissociate when looking at the same SOA pairs as observed with fearful bodily expressions. Error bars indicate standard error mean. Asterisks = p < .05.*

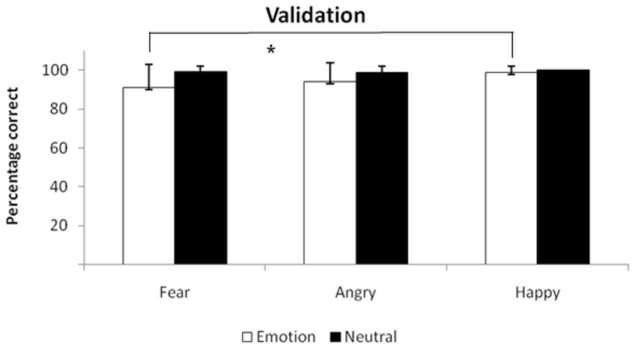


Figure 2.6 *Correct detection of emotional bodily expressions and correct rejection of neutral bodily actions in the validation session. Only the detection of happy expressions differed from the detection of fearful expressions. Error bars indicate standard error mean. Asterisk = p < .05.*

	Fear		Angry		Happy	
SOA	A'	d'	A'	d'	A'	d'
-50	0.94	2.91	0.91	2.26	0.96	3.31
-33	0.94	2.99	0.87	1.87	0.93	2.84
+33	0.93	3.01	0.85	1.62	0.89	2.55
+50	0.94	3.05	0.91	2.27	0.94	3.04

Table 2.1 *Area under the curve values and d' values for detecting fearful, angry and happy bodily expressions.*

2.8 General Discussion

We investigated the relation between the perception of bodily expressions with and without awareness. Our results show that the detection of bodily expressions of fear shows less covariance with how confident participants are about this detection than in the case of detecting angry and happy bodily expressions. This provides novel evidence for the processing of fear stimuli which apparently depends less on the visibility of the expression itself using a stimulus category that is as familiar as it is salient in daily life.

The question remains why the detection of fearful bodily expressions seems to covary less with the subjective confidence than the detection of angry and happy bodily expressions. Öhman (2002, 2005) suggests that fear stimuli automatically activate fear responses and captures the attention as shown in visual search tasks where participants had to detect spiders, snakes or faces among neutral distracters (Öhman, Flykt, & Esteves, 2001; Öhman, Lundqvist, & Esteves, 2001). The special status of fear stimuli is still a matter of debate, specifically in relation to the role of the amygdale (Duncan & Barrett, 2007; Pessoa, 2005). Theoretical models have been advanced arguing that partly separate and specialized pathways may sustain emotional perception, with or without awareness (LeDoux, 1996; Morris, Öhman et al., 1998; Panksepp, 2004; Tamietto et al., 2009; Tamietto & de Gelder, 2010). Our present findings are consistent with a recent study of Pichon, de Gelder, and Grèzes (in press) showing that threatening bodily actions evoked a constant activity in a network underlying reflexive defensive behavior (periaqueductal gray, hypothalamus and premotor cortex) that was independent of the level of attention and was not influenced by the task the subjects were fully engaged in.

When visual signals are prevented from being processed by the cortical mechanisms via the striate cortex, the colliculo-thalamo-amygdala pathway could still process the stimulus. This is in line with recent fMRI studies that have suggested differential amygdala responses to fear faces as compared to neutral faces when the participants were not aware (Morris, Öhman et al., 1998; Whalen et al., 1998).

This process may play an important role in everyday vision by providing us with information about crucial affective signals in our surroundings without being aware of it. Further research using neurological measures will give us insight whether these pathways are indeed mediating the independency of detecting fearful signals from visual awareness.

3. Fear modulates visual awareness similarly for facial and bodily expressions

Abstract

Social interaction depends on a multitude of signals carrying information about the emotional state of others. But the relative importance of facial and bodily signals is still poorly understood Past research has focused on the perception of facial expressions while perception of whole body signals has only been studied recently. To better understand the relative contribution of affective signals from the face only or from the whole body we performed two experiments using binocular rivalry. This method seems to be perfectly suitable to contrast two classes of stimuli to test our processing sensitivity to either stimulus and to address the question of how emotion modulates this sensitivity. In the first experiment we directly contrasted fearful, angry and neutral bodies and faces. We always presented bodies in one eye and faces in the other simultaneously for 60 seconds and asked participants to report what they perceived. In the second experiment we focused specifically on the role of fearful expressions of faces and bodies. Taken together the two experiments show that there is no clear bias towards either the face or body when the expression of the body and face are neutral or angry. However, the perceptual dominance in favor of either the face of the body is a function of the stimulus class expressing fear.

Stienen, B. M. C., & de Gelder, B. (2011). Fear modulates visual awareness similarly for facial and bodily expressions. *Frontiers in Human Neuroscience, 5:132,* doi:10.3389/fnhum.2011.00132.

3.1 Introduction

Social interaction relies on a multitude of signals carrying information about the emotional state of others. Facial and bodily expressions are among the most salient of these social signals. But the relative importance of facial and bodily signals is still poorly understood. Past research has focused on the perception of facial expressions while perception of whole body signals has only been studied recently. Many studies now provide both direct and indirect evidence for visual discriminations of facial expressions in the absence of visual awareness of the stimulus (e.g. de Gelder, Vroomen et al., 1999; Dimberg et al., 2000; Esteves, Dimberg et al., 1994; Jolij & Lamme, 2005; Tamietto et al., 2009). For bodily expressions this is shown in healthy participants (Stienen & de Gelder, 2011) and hemianopic patients (Tamietto et al., 2009). In addition, unattended bodily expressions can influence the judgment of the emotion of facial expressions (Meeren et al., 2005; Van den Stock et al., 2007) and the emotion of crowds is determined by a relative proportion expressing the emotion (McHugh, McDonnell, O'Sullivan, & Newell, 2011) and influences the recognition of the individual bodily expressions (Kret & de Gelder, 2010). However, the relative importance of facial and bodily signals and its relation to visual awareness is still poorly understood.

In this study we investigate directly the contribution of both signals in a binocular rivalry (BR) experiment. BR refers to the phenomenon of perceptual alternation when two incompatible stimuli are presented to the fovea of each eye separately at the same time. This perceptual alternation can be biased by factors such as differences in contrast, brightness, movement and density of contours (Blake & Logothetis, 2002). In addition visual attendance is necessary for rivalry to occur (Zhang, Jamison, Engel, He, & He, 2011). Given certain parameters the two stimuli compete with each other for perceptual dominance rather creating a percept that is a fusion of both. This method seems to be perfectly suitable to contrast two classes of stimuli to test our processing sensitivity to either stimulus and to address the question of how emotion modulates this sensitivity.

Previous BR studies have shown that the meaningfulness of the stimulus influences the rivalry pattern as well (e.g. Yu & Blake, 1992). Subsequent studies have used BR to investigate dominance between faces expressing different emotions (Alpers & Gerdes, 2007; Yoon et al., 2009) and found that emotional faces dominate over neutral faces. In an fMRI study Tong, Nakayama, Vaughan, & Kanwisher (1998) showed that the fusiform face area (FFA), a category specific brain area for processing faces (Haxby et al., 1994), is activated with the same strength as when the faces were presented in a nonrivalrous condition.

fMRI studies using BR in which emotional faces were contrasted showed that suppressed images of fearful faces still activated the amygdalae (Pasley et al., 2004; Williams et al., 2004). When visual signals are prevented from being processed by the cortical mechanisms via the striate cortex, the colliculo-thalamo-amygdala pathway could still process the stimulus (de Gelder, Vroomen et al., 1999; Van den Stock, Tamietto, Sorger, Pichon, & de Gelder, in press). This is in line with recent functional magnetic resonance imaging studies that have suggested differential amygdala responses to fear faces as compared to neutral faces when the participants were not aware (Morris, Öhman et al., 1998; Whalen et al., 1998). However, to date no BR experiments or masking experiments were conducted using bodily expressions or have investigated the contribution of information from both the face and the body and how this influences our conscious percept.

We performed two behavioral experiments addressing relative processing sensitivity to facial and bodily expressions and investigated how specific emotions modulate this sensitivity. First, we performed an experiment involving the rivaling of bodies and faces with fearful, angry and neutral expressions. We always presented bodies in one eye and faces in the other and asked participants to report what they perceived while stimuli were presented simultaneously for 60 seconds. In line with BR studies using facial expressions (Alpers & Gerdes, 2007; Pasley et al., 2004; Williams et al., 2004; Yoon et al., 2009) we expected that emotional bodily expressions would dominate over neutral expressions. The first experiment showed a special role of fearful expressions and therefore we isolated this condition in a second, more sensitive, experiment. In this second experiment we used the rivalry pattern resulting from the contrasting of neutral facial and bodily expressions as baseline performance and created two conditions in which fearful bodily expressions were contrasted with neutral facial expressions and fearful facial expressions with neutral bodily expressions. We expected that the perceptual dominance of the stimulus would be a function of the stimulus expressing fear.

EXPERIMENT 1

In this first experiment we contrasted bodily and facial expressions directly in a binocular rivalry design in which the emotion of the faces and bodies were fearful, angry or neutral.

3.2 Material and Methods

Participants

Twenty-two undergraduate students of Tilburg University participated in exchange of course credits or a monetary reward (19 women, 3 men, *M* age = 19.8 years, *SD* = 1.2). All participants had normal or corrected-to-normal vision and gave informed consent according to the declaration of Helsinki. The protocol was approved by the local Ethics Committee Faculteit Sociale Wetenschappen of Tilburg University.

Stimuli and procedure

Photos of two male actors expressing fear and anger the same actors performing a neutral action (hair combing) were selected from a well validated photoset as body stimuli (for details see Stienen & de Gelder, 2011). All body pictures had the face covered with an opaque oval patch to prevent that the facial expression would influence the rivalry process. The color of the patch was the average grey value of the neutral and emotional faces within the same actor. The face stimuli of two actors expressing fear and anger and the same actors showing a neutral expression were taken from the McArthur set (http://www.macbrain.org/resources.htm). A total of six pictures of bodily expressions and six pictures of facial expressions were selected for use in the present study.

All stimuli were fitted into an area with a white background of 3.00 degrees * 4.83 degrees enclosed by a black frame of with a border thickness of .29 degrees.

The function of the black frame was to enhance a stable fusion. A white fixation dot was pasted on each of the stimuli. Because we used a method which is comparable with the mirror stereoscope the faces and bodies were pasted 11.89 degrees left and right from the center. Pairing the face and body stimuli resulted in 18 unique displays (3 bodily expressions x 3 facial expressions x 2 identities).

One experimental run consisted of 36 trials because the displays were counterbalanced to control for eye dominance. The trials were randomly presented. The stimuli were presented on a 19" PC screen with the refresh rate set to 60 Hz. We used Presentation 11.0 to run the experiment.

The heads of the participants were stabilized using a chin and head rest. The fMRI compatible binocular rivalry method we used is described in detail by Schurger (2009) but was here adapted for use outside of the scanner. A black 70 cm wooden divider was placed between the screen and the middle of the eyes. The total distance between the screen and eyes was 77 cm. Participants wore glasses in which two wedge-shaped prism lenses of 6 DVA were fitted using gum. The prisms adjusted the viewing angle from which light from the screen enters each eye ensuring that the laterally presented stimuli would fall close to the participants' fovea. The wooden divider was placed between the eyes to keep the visual signals separated. Besides the fact that this is a low-cost method and it can be used in- and outside the MRI scanner there is no crosstalk between the eyes (Schurger, 2009) as is the case with for example red-green filter glasses. See Figure 3.1 for a picture of the experimental setup.

Before each trial two empty frames were shown with a black fixation dot in the middle. The participants were instructed to push and hold a button labeled "M" (Dutch for *mixture* = *mengsel*) on a response box with the middle finger to initiate a trial, but only if they saw one dot and one frame. This ensured that the participants fused the two black frames throughout the experiment. Subsequently, a facial expression and a bodily expression were presented for 60 seconds. For an example display see Figure 3.1. Whenever they saw a face or a body in isolation they were instructed to release the "M" button and push and hold the button corresponding to their percept; the "G" (Dutch for *face* = *gezicht*) if they saw a face or the "L" (Dutch for *body* = *lichaam*) if they saw a body with either their index or ring finger.

The “G” and “L” button was counterbalanced across participants and they always used their right hand. When seeing both stimuli they were told to push and hold the button labeled “M” again. The program registered the time the button was pressed and released. The participants were naïve regarding the presentation techniques and during the experiment no reference to the emotions was made.

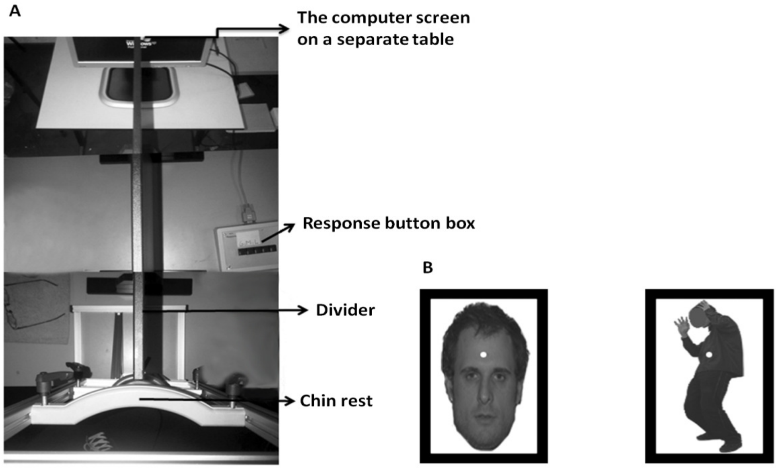


Figure 3.1 *Experimental setup (**A**). Example of a stimulus display with a neutral face on the left and a fearful body on the right. We always presented bodies in one eye and faces in the other (**B**).*

Prior to the experimental sessions the participants performed one practice session consisting of two trials. This session used different male identities taken from the same stimulus sets than the ones used in the main experiment. When the participants reported full understanding of the procedures the main experiment started. A total of two runs were presented adding up to a total of 72 trials. After each 10 trials there was a short break. Finally a short validation was performed in a separate session after a 5 minutes break. All stimuli were presented two times for two seconds adding up to a total of 24 trials (2 identities * 3 expressions * 2 face/body*2 runs). Participants were instructed to categorize the bodies and faces in fearful, angry or neutral bodily or facial expressions using three buttons labeled "A" for fearful (Dutch = *angst*), "B" for angry (Dutch = *boos*) and "N" for neutral (Dutch = *neutraal*).

3.3 Results and Discussion

Cumulative viewing time for faces, bodies, and mixed perceptions were calculated per participant irrespective of experimental condition. Two participants indicating having seen mixed percepts more often than two standard deviations below the group average (group mean = 104 s, SD = 34 s) were identified as outliers and excluded from analysis. See Figure 3.2 for the individual data.

Wilcoxon Signed Ranks Tests revealed that the cumulative viewing time of faces (*M* = 51 s, *SD* = 24 s) and bodies (*M* = 52 s, *SD* = 17 s) was equal (Z = -.075, *p* = .940) while the cumulative viewing time was longer for mixed perceptions (*M* = 111 s, *SD* = 34 s) in comparison to bodies and faces (resp. Z = -3.696, *p* < .001 and Z = -3.696, *p* < .001).

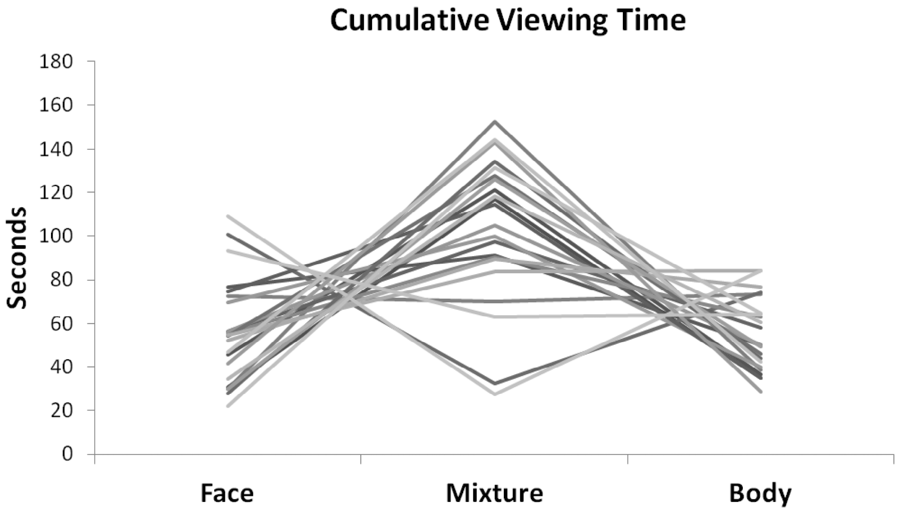


Figure 3.2 *Cumulative viewing time of the perception of faces, bodies and mixtures. The two subjects with the lowest cumulative viewing time of mixtures were removed from analysis.*

Following Levelt (1965) predominance ratios were calculated. The total time participants indicated seeing the face was subtracted from the total time participants indicated seeing the body. This value was divided by the total amount of time the body and the face was seen. If this predominance ratio has a value of zero it would mean they equally perceived the body and the face in time. A positive value means that the conscious percept of the body predominated over face while a negative value means that the conscious percept of the face dominated over body.

A 3 (bodily expressions) x 3 (facial expressions) GLM repeated measurements revealed a significant interaction between the bodily expressions and the facial expressions on the predominance ratios (F(4,76) = 3.877, p = .006) as well as a main effect of facial expressions (F(2, 38) = 24.718, p < .001). Figure 3.3 shows the predominance ratios when the bodily or the facial expression was emotional and the other was neutral (Figure 3.3a), when the facial and bodily expressions were the same (Figure 3.3b), and when the facial and bodily expressions both differed (Figure 3.3c). A difference was deemed significant when the p-value was lower than .005 (Bonferroni correction: α level divided by 10 comparisons).

Figure 3.3a shows that when the body expressed fear and the face was neutral the participants reported more often seeing the body than when the face was fearful and the body was neutral (t(19) = 2.903, p = .009), but this effect did not survive the Bonferroni correction. The predominance ratios were equal when the bodily or facial expression was angry. Figure 3.3b shows that when both stimulus classes express fear the face dominates over the body compared when they are both neutral (t(19) = 3.471, p = .003). Figure 3.3c shows that when the expressions were both emotional but different (fearful and angry) the fearful body triggered a stronger conscious percept of the body when the rivaling face was angry compared to when the face was fearful and the rivaling bodily expression was angry in which case the conscious percept of the face predominated t(19) = 4.586, p < .001).

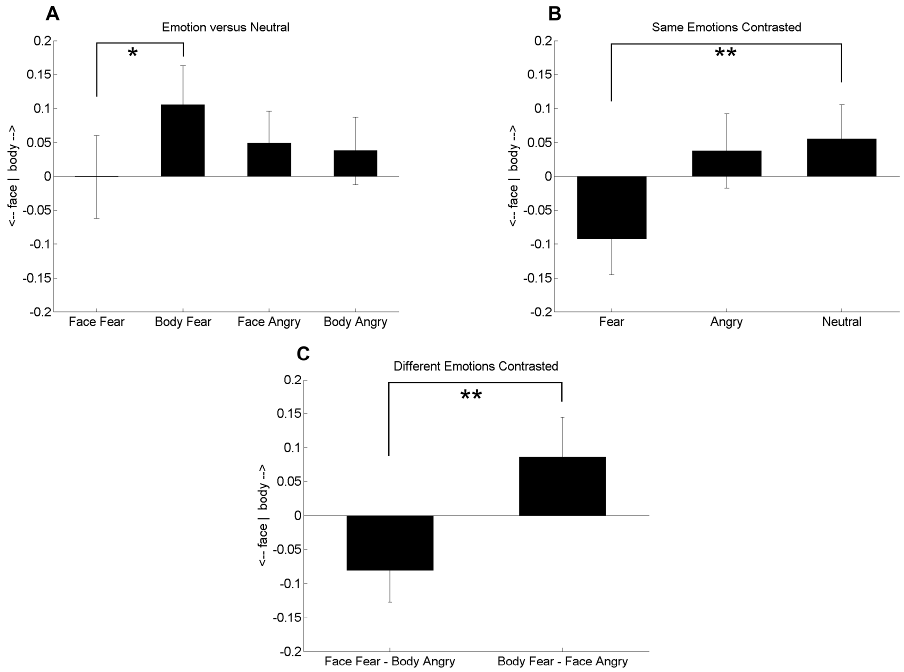


Figure 3.3 *A positive value means that the body predominates over the face and a negative value that the face predominated over the body. Predominance ratios when the bodily or facial expression was emotional and the other was neutral (**A**). Predominance ratios when the facial and bodily expressions were the same (**B**). Predominance ratios when the facial and bodily expressions both differed (**C**). Error bars represent standard error of the mean. One asterisk = p < .01, double asterisks = p < .005.*

To test the main effect of facial expressions pairwise Bonferroni corrected comparisons were performed between the predominance ratios irrespective of bodily expressions. When the facial expression was fearful the face dominated over the body more than when the facial expression was angry or neutral ($p < .001$).

A 2 (face/body) x 3 (fear/angry/neutral) GLM repeated measurements on the correct categorizations in the validation task revealed a main effect of stimulus class ($F(1,17) = 14.806$, $p = .001$). It appeared that the facial expressions were categorized better in general regardless of expression. Because the results in the main experiment are specific for fearful expressions a general effect on the recognition of faces alone cannot explain the specific effect. See Figure 3.4a for the validation results.

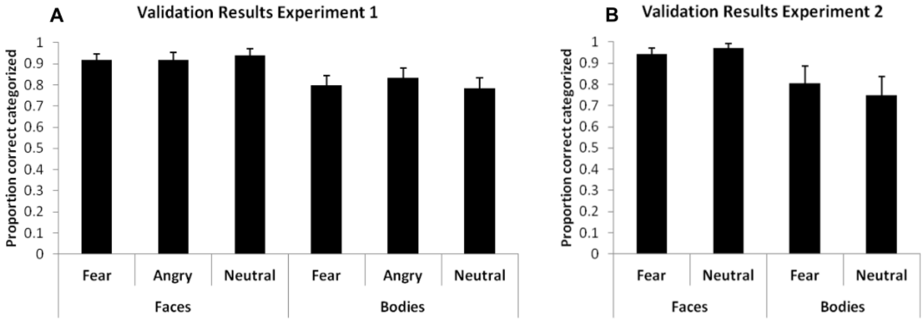


Figure 3.4 *Proportion correct categorizations in the validation session of experiment 1 (**A**) and experiment 2 (**B**). Error bars represent standard error of the mean.*

In line with previous reports on the special role of fearful expressions (Öhman, 2002; Stienen & de Gelder, 2011) the main finding of this first experiment is that the stimulus class carrying the fearful expression suppresses the percept of the competing stimulus more than angry and neutral expressions do. In addition, participants seemed to be equally sensitive in perceiving the face and the body when the emotional expression was neutral or angry.

Past research has focused on for example the perception of facial or bodily expressions in isolation, but never compared these two important social signals together in one display. Although, Meeren et al. (2005) and Van den Stock et al. (2007) showed the influence of unattended bodily expressions on the task relevant facial expressions, this study revealed how the two stimuli compete for visual awareness when they are both task relevant as it the case in natural situations.

There was no indication in this experiment that neutral or angry expressions modulated the rivalry pattern but there were clues indicating that fearful expressions modulated the resulting dominant percept. However, none of the conditions explicitly deviated from the value zero. The value zero meant an equal ratio between reporting the face or the body. To create a more sensitive design we repeated the first experiment but this time with only three conditions; one baseline condition in which neutral facial and bodily expressions were contrasted and two experimental conditions in which either the face or the body was expressing fear. By lowering the amount of conditions we could increase the number of trials.

EXPERIMENT 2

In this experiment a baseline was created by contrasting a neutral facial expression with a neutral bodily expression. The resulting perceptual alternation was compared when either the bodily or the facial expression was fearful while the other was neutral. Although these conditions were present in the first experiment as well we wanted to test these conditions in isolation. We hypothesized that based on our first experiment either the body or the face will dominate depending on which is expressing fear.

3.4 Material and Methods

Participants

Nineteen new undergraduate students of Tilburg University who had not taken part in the first experiment participated in exchange of course credits or a monetary reward (15 women, 4 men, *M* age = 19.9 years, *SD* = 1.6). All participants had normal or corrected-to-normal vision and gave informed consent according to the declaration of Helsinki. The protocol was approved by the local Ethics Committee Faculteit Sociale Wetenschappen of Tilburg University.

Stimuli and procedure

The stimuli were the same as in the first experiment, but this time only the bodily and facial neutral and fearful expressions were used. There were three conditions: a neutral body and face (baseline), a fearful body and a neutral face (fearful body), and a neutral body and a fearful face (fearful face). In total there were 12 different displays (2 body/face x 3 baseline/fearful body/fearful face x 2 identities). One complete run consisted of 24 trials because the displays were counterbalanced to control for eye dominance. A total of two runs were presented adding up to a total of 48 trials. The rest of the procedure remained the same as in experiment 1.

3.5 Results and Discussion

Wilcoxon Signed Ranks Tests revealed that the cumulative viewing time of faces (*M* = 11 s, *SD* = 6 s) was longer than for bodies (*M* = 7 s, *SD* = 3 s), Z = -3.622, $p < .001$. The cumulative viewing time was longer for mixed perceptions (*M* = 23 s, *SD* = 8 s) in comparison to bodies and faces (resp. Z = -3.702, $p < .001$ and Z = -2.696, $p = .007$).

Predominance ratios for all three conditions (baseline, fearful body, and fearful face) were calculated in the same manner as the predominance ratios in the first experiment were calculated. The ratio when the baseline trials were presented was subtracted from the predominance ratios of the fearful body condition and the fearful face conditions.

Figure 3.5a shows the baseline condition where neutral bodies were contrasted with neutral faces. A one sample t-test showed that the predominance ratio was not significantly different from zero which means that participants equally perceived the body or the face when the expressions were neutral (t(18) = .085, $p = .933$). Figure 3.5b shows the modulation of the fearful expression when either the neutral body or the neutral face was substituted by respectively a fearful body or a fearful face. As indicated by a paired t-test a fearful body triggered a more dominant body percept and a fearful face triggered a more dominant face percept (t(18) = -4.60, $p < .001$). When comparing directly to the baseline only fearful faces triggered a more dominant face percept (t(18) = 3.975, $p = .001$).

A different way of analyzing the results is by considering the participants' initial percept per condition (Berry, 1969; Long & Olszweski, 1999; Yoon et al., 2009). The frequency of reporting a face or a body as initial percept when a trial started was indexed. Subsequently the data was treated the same way as the predominance ratios.

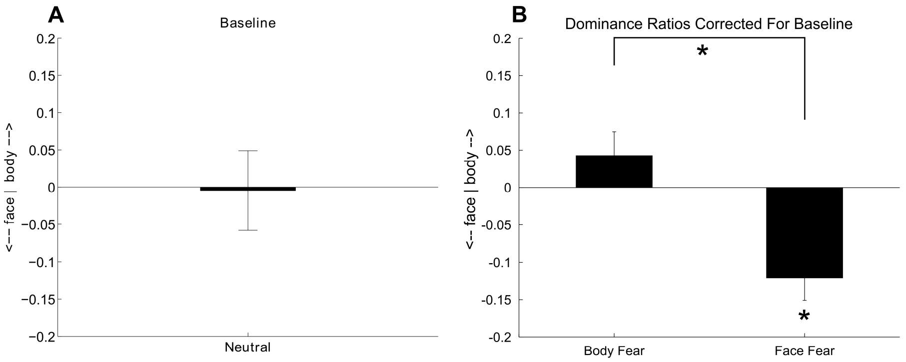


Figure 3.5 *A positive value indicates that the body predominated over the face and a negative value that the face predominated over the body. Predominance ratio when a neutral bodily expression is contrasted with a neutral facial expression (**A**). Predominance ratios when a fearful body is contrasted with a neutral face and when a fearful face is contrasted with a neutral body (**B**). Error bars represent standard error of the mean. Asterisk = p < .01.*

As Figure 3.6 shows these results follow approximately the same pattern. When both the bodily and facial expressions were neutral the reported initial percept was equally bodies and faces (t(18) = -,042, *p* = .967). Figure 3.5b shows that as an initial percept fearful body triggered more a body percept and a fearful face triggered more a face percept (t(18) = -4.60, *p* < .001). Neither a fearful body nor a fearful face triggered more initial percepts of their own stimulus class when directly compared to baseline performance.

See Figure 3.4b for the validation results. A 2 (face/body) x 2 (fear//neutral) GLM repeated measurements revealed a main effect of stimulus class on the validation scores (F(1,17) = 11.311, *p* = .004). It appeared that facial expression was categorized again better in general regardless of emotional expression.

This second experiment shows that indeed the stimulus class expressing fear leads to perceptual dominance of the stimulus class carrying this information, although the effect seems stronger for the fearful faces.

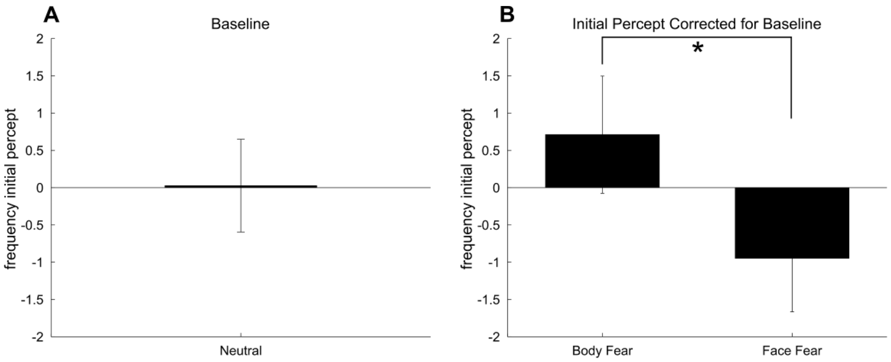


Figure 3.6 *A positive value means that the body is reported as the initial percept more often than the face, a negative value that the face is reported as the initial percept more often than the body. Initial percept ratio when a neutral bodily expression is contrasted with a neutral facial expression (****A****). Initial percept ratios when a fearful body is contrasted with a neutral face and when a fearful face is contrasted with a neutral body (****B****). Error bars represent standard error of the mean. Asterisk = p < .05.*

3.6 General Discussion

Taken together our experiments show that there is no clear bias towards either the face or body when both have a neutral or an angry expression. When both the face and the body were expressing fear participants perceived more the face compared to when both categories were neutral. As especially the results of the second experiment showed, the perceptual dominance in favor of either the face of the body is a function of the stimulus class expressing fear while the effect was stronger for fearful faces. In the second experiment the cumulative viewing time for faces were longer than for bodies. Finally, the validation results of both experiments show that facial expressions were recognized better.

When there is no emotion expressed, the reported conscious percept of the body and face was equal. This indicates that in this case there is equal processing sensitivity to either stimulus class. Only when signals of fear are transferred by the stimulus the perceptual alternation is influenced by suppressing non-fearful expressions. This is in line with Öhman (2002, 2005) suggesting that fear stimuli automatically activate fear responses and captures the attention as shown in visual search tasks where participants had to detect spiders, snakes or schematic faces among neutral distracters (Öhman, Flykt et al., 2001; Öhman, Lundqvist et al., 2001), or real faces as in our study (Hodsoll, Viding, & Lavie, 2011), although this is not always found in other studies (e.g. Calvo & Nummenmaa, 2008).

It is known that voluntary endogenous and involuntary exogenous attention can modulate the rivalry pattern (Blake & Logothetis, 2002; Tong, Meng, & Blake, 2006). However, The relative dominance of perceiving bodies when the body is fearful and the face is neutral in contrast when the face is fearful and the body is neutral is also consistent with a recent study of Pichon, de Gelder, and Grèzes (in press) showing that threatening bodily actions evoked a constant activity in a network underlying preparation of automatic reflexive defensive behavior (periaqueductal gray, hypothalamus and premotor cortex) that was independent of the level of attention and was not influenced by the task the subjects were fully engaged in.

The dominant perception of the faces and bodies expressing fear was mostly relative but there was one case, in the second experiment, in which the conscious percept of the fearful face dominated in absolute terms. Although the recognition of faces was better regardless of expression in both experiments; this alone cannot explain the specific effect of fearful faces on the rivalry pattern. The fearful face deviated from zero in the second experiment and not in the first probably because of two reasons. Firstly, there were fewer conditions and more trials which could have increased the signal-to-noise ratio. Secondly, the fearful expressions are likely to pop-out more when among neutral expressions without the angry expressions being present within the same experiment. Although, as already mentioned, this pop-out effect for fearful stimuli is not always found in visual search tasks using real faces.

Furthermore, it is possible that the relative proximity to the observer of the faces in contrast with bodies could explain why the face was more dominantly perceived than baseline and bodies were not. As suggested earlier (de Gelder, 2006, 2009; Van den Stock et al., 2007) the preferential processing of affective signals from the body and/or face may depend on a number of factors and one may be the distance at which the observer finds himself from the stimulus.

The special status of fear stimuli is still a matter of debate, specifically in relation to the role of the amygdalae (Duncan & Barrett, 2007; Pessoa, 2005). Theoretical models have been advanced arguing that partly separate and specialized pathways may sustain emotional perception, with or without awareness (LeDoux, 1996; Morris, Friston et al., 1998; Panksepp, 2004; Tamietto et al., 2009; Tamietto & de Gelder, 2010). Our results are in line with Pasley et al. (2004) and Williams et al. (2004) showing amygdala activity for suppressed emotional faces. This hints at the possibility that the suppressed fearful faces are being processed through the the colliculo-thalamo-amygdala pathway.

The underlying process may play an important role in everyday vision by providing us with information about important affective signals in our surroundings. Further research using neurological measures will give us insight whether the relevant pathways are indeed mediating detection of fearful signals independently of visual awareness. In addition, future studies using a different stimulus set or broadening the set to include other emotions would be of great value for the matter of validation and to investigate the generalization of the present findings to other emotions.

4. Emotional voice and emotional body postures influence each other independently of visual awareness

Abstract

Multisensory integration may occur independently of visual attention as previously shown with compound face-voice stimuli. We investigated in two experiments whether the perception of whole body expressions and the perception of voices influence each other when observers are not aware of seeing the bodily expression. In the first experiment participants categorized masked happy and angry bodily expressions while ignoring congruent or incongruent emotional voices. The onset between target and mask varied from -50 to +133 ms. Results show that the congruency between the emotion in the voice and the bodily expressions influences audiovisual perception independently of the visibility of the stimuli. In the second experiment participants categorized the emotional voices combined with masked bodily expressions as fearful or happy. This experiment showed that bodily expressions presented outside visual awareness still influence prosody perception. Our experiments show that audiovisual integration between bodily expressions and affective prosody can take place outside and independent of visual awareness.

Stienen, B. M. C., Tanaka, A, & de Gelder, B. (2011). Emotional voice and emotional body postures influence each other independently of visual awareness. *PLoS ONE, 6*(10), e25517.

4.1 Introduction

Our social interactions depend on receiving and combining affective signals from multiple sources such as faces, voices, body postures and other contextual information in our environment. Previous research has mainly investigated face-voice combinations (de Gelder & Bertelson, 2003; de Gelder, Böcker, Tuomainen, Hensen, & Vroomen, 1999; de Gelder & Vroomen, 2000; Massaro & Egan, 1996). For example, de Gelder & Vroomen (2000) presented facial expressions that were morphed on a continuum between happy and sad, while at the same time a short spoken sentence was presented. This sentence had a neutral meaning, but was spoken in either a happy or sad emotional tone of voice. Participants were instructed to attend to and categorize the face, and to ignore the voice, in a two-alternative forced-choice task. The results showed a clear influence of the task-irrelevant auditory modality on the target visual modality.

More recently body-voice combinations have also been studied (Van den Stock, de Jong, Hodiamont, & de Gelder, in press; Van den Stock, Grèzes et al., 2008) generalizing these multisensory effects to a broader domain. By switching to this affective stimulus category, we may be capable of extending the scope of face-based research and provide evidence that human emotion theories may generalize to other affective signals as well. Results from a number of behavioral experiments using independent stimulus sets now persuade us to conclude that recognition of expressions is similar for face and body stimuli. A major difference between facial and bodily expressions is that the latter can be recognized from far away while the former require the viewer to be nearby. This is potentially an important difference between how facial and bodily expressions play their communicative roles and it should have consequences how the specific information is conveyed (de Gelder, 2006, 2009).

Crossmodal emotion effects are shown whereby affective information in one sensory modality influences perception in the other while the signals are perceived both consciously (de Gelder & Vroomen, 2000; Tanaka et al., 2010). These crossmodal effects have again mainly been shown for faces. However, previous studies on the automaticity of audiovisual integration have mainly investigated the role of attention (Alsius, Navarra, Campbell, & Soto-Faraco, 2005; Vroomen, Driver,

& de Gelder, 2001). But attentional selection does not imply that one is consciously aware of the stimulus. Furthermore, the unattended stimulus could be consciously perceived (Tamietto & de Gelder, 2010). This uncontrolled role of consciousness could explain why multisensory integration occurs. For example, if consciousness is necessary for multisensory integration to occur then the process is not automatic. There is some evidence that visual awareness does not seem to be a prerequisite for audiovisual affect integration since crossmodal interactions are still observed when the face is not consciously perceived in hemianopic patients (de Gelder, Pourtois, & Weiskrantz, 2002), but, so far, whether this is the case in neurological intact observers remains unknown.

A number of research reports have concluded that emotional information can be processed without observers being aware of it. Many studies using facial expressions now provide both direct and indirect evidence for visual discriminations of affective stimuli in the absence of visual awareness of the stimulus. On forced choice tasks, it has been shown that clinically blind hemianopic patients can reliably guess the emotion of facial and bodily expressions presented in their blind field (de Gelder, Vroomen et al., 1999; Tamietto et al., 2009).

Masking is one of the most widely used techniques for exploring processing of visual emotional information without awareness in neurologically intact observers. For example, Esteves and Öhman (1993) found that short duration (e.g. 33 ms) presentations of facial expressions (happy and angry), replaced immediately by a neutral face (mask) with a longer duration (e.g. 50 ms), are below the participants' identification threshold. We have recently shown in a parametric masking study that the detection of fearful bodily expressions covaries less with visual awareness than the detection of other bodily expressions (Stienen & de Gelder, 2011).

Öhman (2002, 2005) suggests that fear stimuli automatically activate fear responses and captures the attention as shown in visual search tasks where participants had to detect spiders, snakes or faces among neutral distracters (Öhman, Flykt et al., 2001; Öhman, Lundqvist et al., 2001). The special status of fear stimuli is still a matter of debate, specifically in relation to the role of the amygdale (Duncan & Barrett, 2007; Pessoa, 2005).

Here our goal was to address whether affective information from voices influences the affective information from bodily expression independently of visual awareness. First, we investigated the influence of the perception of emotional voices on the recognition performance of emotional body expressions under conditions of visual uncertainty, and subsequently we investigated whether unseen bodily expressions affect the recognition of the prosody in the perceived voice.

In the first experiment voice fragments were presented simultaneously with pictures of bodily expressions. A mask was presented at 12 different latencies after or before the onset of the target (Stimulus Onset Asynchrony, SOA). The participants were instructed to categorize bodily expressions which were congruently or incongruently paired with emotional voices and subsequently to indicate whether they were sure of their answer or whether they were guessing. Instructions specified they had to ignore the voice. In the second experiment we masked emotional body pictures using only an SOA of 33 ms. This time the participants had to focus on the voice component of the stimulus. Our extensive semi-structured exit interview and our sensitive posttest assessed whether the participants had been aware of the emotional body pictures.

EXPERIMENT 1

This experiment investigated whether emotion in the voice influences the recognition of bodily expressions independently of visual awareness.

4.2 Material and Methods

Participants

Sixteen undergraduate students of The University of Tilburg participated in exchange for course credits or a monetary reward (9 women, 7 men, M = 20.0 years, SD = 2.2). All participants had normal or corrected-to-normal vision and gave informed consent according to the declaration of Helsinki. The protocol was approved by the local Ethics Committee Faculteit Sociale Wetenschappen of Tilburg University.

Stimuli and procedure

Frames from video clips were used as stills of bodies displaying angry and happy expressions. For full description of the set of video clips and information regarding their validation see Stienen and de Gelder (2011). In total 16 stimuli (2 emotions x 2 gender x 4 actors) were selected. These stimuli were frames from the video clips in which the actor seemed to be optimally expressing the emotion. The faces of the actors were covered to prevent that the facial expression would influence the identification of the emotional body posture. All actors wore black clothing and all images were converted to grey values.

Still images taken from neutral action video clips such as fixing one's hair or cloths were selected to construct the mask. A neutral bodily expression of a male with an average posture was chosen as the basis. The arms and legs were erased and twelve arms and legs from other identities expressing a neutral emotion were attached to the body at different positions and orientations creating the image of a body with more arms and legs than usual.

Average height of the bodies was 7.82 degrees (SD = .26 degrees), the average maximum width (distance between the hands) was 3.76 degrees (SD = .85 degrees) and the average waist was 1.55 degrees (SD = .14 degrees) of visual angle. The height of the mask was 8.12 degrees, the maximum width was 6.21 degrees and the waist was 1.64 degrees of visual angle. The mask covered the target stimuli completely. See Figure 4.1 for examples of the stimuli.

Twenty-four emotional meaningless human vocalizations (e.g., "ah" or "uh") expressing happy or angry emotions from 12 different speakers were recorded. Each recording was edited to create 8 different fragments of 8 different durations (25, 50, 75, 100, 150, 200, 250, and 400 ms), resulting in 192 stimuli in total. Loudness was equated in terms of the A-weighted sound pressure level. Sounds were gated with 5 ms raised-cosine onset and offset ramps to avoid clipping. In the pilot experiment, 10 participants categorized the emotion of all the 192 vocal expressions into happy or angry emotions. Based on the accuracy results, we decided to use the voice clip of 250 ms for which the overall accuracy was highly better than chance (89.8%), t(9) = 15.23, *p* < .001.. The accuracy results did not differ between emotions, t(3) = 0.77, p = .50. Angry and happy vocalizations from two male and two female speakers (e.g., “ah” or “uh”) were used and paired congruently and incongruently with the visual stimuli. The voice-body stimulus compound was always gender-congruent.

Participants were comfortably seated in a chair in a soundproof experimental booth approximately 90 cm from the screen. The disappearance of a fixation cross signaled the beginning of a trial. After 500 ms the target stimulus appeared for 33 ms accompanied with an angry or happy voice, which was congruent or incongruent (50 percent/50 percent) with the emotion of the bodily posture. After a variable interval the mask was presented for 50 ms (in case of forward masking the mask was presented first).

It is known that the largest masking magnitude associated with pattern masking is around an SOA of 0 milliseconds (Breitmeyer & Ganz, 1976; Enns & Di Lollo, 2000). Therefore the values for the SOA latencies included the SOA of 0 ms. The SOA latencies were -50, -33, -17, 0, 17, 33, 50, 67, 83, 100, 117 and 133 ms. Negative values represent forward masking and positive values backward masking. When the SOA was -33, -17, 0 and17 ms the target overlaps with the mask. The target was always presented at the foreground. Moreover a target-only condition and a no-target condition were included.

The participants were instructed to categorize the emotional expressions of the body and to ignore the emotional voice. They had to respond with the left hand using two response buttons situated in front of them with the labels “Happy” and “Angry” attached to it. Subsequently they had to indicate whether they were sure or guessing. They had to respond with the right hand with two different buttons on the same response box labeled with “Sure” and “Guessed”. They were instructed to use their “gut feeling” if they had not seen the body. Fingers, but not hands were counterbalanced across participants. See Figure 4.1 for a schematic representation of a trial.

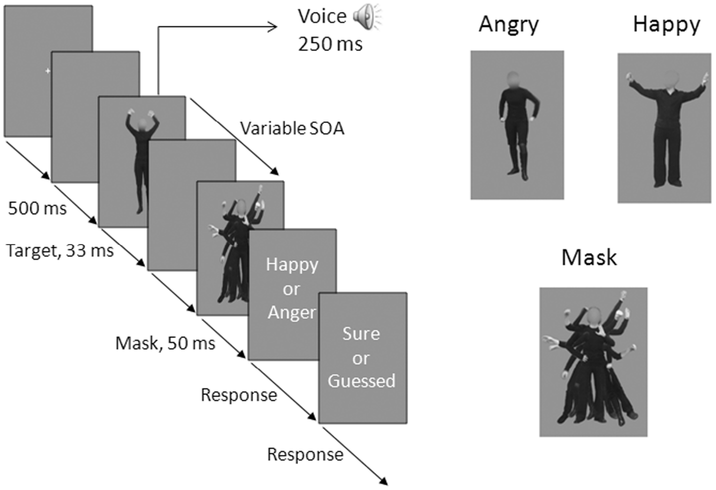


Figure 4.1 *An example trial (**left**), an example of an angry and happy bodily posture (**upper right**), the mask (**below right**).*

Prior to the experimental sessions the participants performed three practice sessions consisting of 27 trials each. Other identities than the ones used in the main experiment served as targets. When the participants did not miss trials and gave notice of a full understanding of the procedure the main experiment was started. A total of four runs were presented adding up to a total of 896 trials. Every 112 trials there was a 3 minute break. After the main experiment in a separate session all targets were presented for 33 ms without the pattern mask to validate the stimuli. The participants were instructed to categorize the targets in angry and happy expressions. The total duration of the experiment was 1 hour and 45 minutes.

Analysis

Percentage correct categorized bodily expressions were corrected for chance level which was 50 percent. To assess whether participants could differentiate between the correct and incorrect answers confidence ratings were calculated. The number of sure responses when the categorization the emotional expression was incorrect was subtracted from the number of sure responses when the response was correct.

This was divided by the total number of correct and incorrect answers. A resulting value of zero would mean that the participants indicate subjectively that they are not more confident of their correct answers then their incorrect answers which is taken as a measure of subjective visual awareness. A similar approach was chosen by Cheesman and Merikle (1986) and Esteves and Öhman (1993) as a measure of the phenomenological experience of the participants' perception of the targets.

This method automatically controls for how well the participants are engaged in the task. If, for example, a participant would just randomly categorize the emotion, but always indicates to be sure, the confidence measure would end up being 100 percent while the accuracy would be around zero after correction of chance level. However, our measure of confidence would also result in a confidence rating of zero, because it automatically corrects for when the participants indicate to be sure when their answer is wrong.

4.3 Results and Discussion

Two participants were discarded from analysis because they performed well below 50 percent in categorizing the angry and happy bodily expression in the validation study (37.5 and 25.0 percent), while the group average was 84.4 percent (SD = 13.9 percent). The correct identifications were on such a low level that there is a possibility that the two participants did not understand the instruction clearly, for example they confused the order of the response buttons. The validation study showed that the angry bodily emotion was correctly identified 84.8 percent of the cases (SD = 16.3) and happy bodily emotion 83.9 percent of the cases (SD = 19.3).

Two GLM repeated measures analyses with emotion (2 levels), congruency (2 levels) and SOA (13 levels) as factors were performed on the categorization performance and confidence ratings. There was a main effect of SOA and congruency on accuracy, resp. $F(12,156) = 14.50$, $p < .01$; $F(1,13) = 10.45$, $p < .01$. In addition, a main effect of SOA and congruency were observed on the confidence ratings, resp. $F(12,156) = 18.67$, $p < .01$; $F(1,13) = 10.96$, $p < .01$. Bonferroni corrected pairwise comparisons showed that the longer the SOA the higher the categorization performance and confidence ratings, e.g., when there was no mask in the case of the target-only trials the categorization performance was highest (mean = 76.7 %, SD = 4.02) and lowest when the SOA was -17 ms (mean = 50.5 %, SD = 1.20). For the confidence ratings this was also true. The participants were most confident when there was no mask (mean = .557, SD = .083), although they were the least confident when the SOA was 0 ms (mean = .019, SD = .019). In addition the comparisons between incongruent and congruent body-voice pairs showed that the categorization performance and confidence ratings were higher when the emotion was congruent. The specific emotion did not have a main effect on the accuracy or confidence ratings nor did it interact with the other factors. Figure 4.2 shows the accuracy and the confidence of the participants averaged over the two emotions.

Interestingly, there was no interaction between congruency and SOA on accuracy ($F(12, 156) = 1.09$, $p = .37$), while the factors interacted on confidence ratings ($F(12, 156) = 2.48$, $p < .01$). To investigate this interaction post hoc comparisons were done between congruent and incongruent trials on the confidence ratings per SOA. Results suggested that the difference between congruent and incongruent trials was absent in the confidence ratings when the SOA ranged from 0 to +50 ms ($p > .05$, Bonferroni corrected). Within this range it appeared that when the SOA ranged from 0 to +33 ms the confidence ratings when the emotion of voice and body were congruent or incongruent were never above zero (all $p > .0125$, Bonferroni corrected). Yet, when the emotion of the voice-body pairs was congruent the accuracy in the whole range (from 0 to +50 ms) was above zero (all $p < .0125$, Bonferroni corrected). This was not the case when the emotion of the voice-body pairs were not congruent (all $p > .0125$, Bonferroni corrected).

When emotional voices and body postures are congruent objective recognition of emotional body expressions is aided regardless of SOA. This same effect is not seen in subjective confidence ratings where there is no facilitation effect of congruent voice information for short SOA latencies. Conjointly, the confidence of the participants was not above zero in this range while the accuracy when the emotional voice-body pairs were congruent was better than chance. The subjective ratings can be taken as measure of the phenomenological experience of the participants' perception of the targets (Cheesman & Merikle, 1986; Esteves &

Öhman, 1993). The combination of these findings shows that the emotion of the voice exerts its influence independently of the visual awareness of the target.

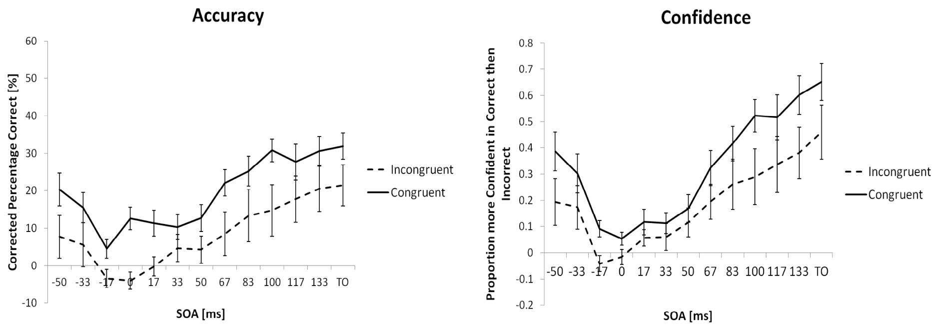


Figure 4.2 *Mean categorization performance plotted as function of SOA corrected for chance (50 %)(**left**).Mean confidence ratings plotted as function of SOA (**right**).The solid line represents the performance when the emotion of the voice is congruent with the emotion of the bodily expression and the broken line when the emotional information is incongruent. Error bars represent standard error of the mean. SOA = Stimulus Onset Asynchrony, TO = Target Only.*

The lack of the interaction between congruency and SOA in accuracy shows that these results do not reflect merely a decision or response bias (de Gelder & Bertelson, 2003). Such a bias would be stronger when visibility of the target is low and would thus result in an interaction of congruency and SOA on the categorization performance of the participants. In other words, this method shows to be a good control to check whether such a bias is present in the data set.

While this study shows that visual awareness is not necessary for the multisensory integration to occur the participants were in fact, capable of detecting the bodily expressions in the majority of the trials because this concerns a parametric masking study. In other words, they were aware that bodily expressions were presented while ignoring the human emotional vocalizations. In a second study we therefore isolated one SOA condition to ensure that the participants would not perceive bodily expressions throughout the whole experiment while judging the emotion of spoken sentences. If we would observe similar effects on the judgment of emotional prosody because of the influence of unseen bodily expressions this would strengthen the conclusion that bodily expressions and emotional voices influence each other independently of visual awareness.

EXPERIMENT 2

In the first experiment the influence of the emotion in the voice and its dependency on visual awareness was the focus of interest. In this second experiment we asked whether bodily expressions when presented outside visual awareness can influence the recognition of prosody in spoken words. While in the first experiment the visibility of the bodily expressions was parametrically varied we held the SOA constant (33 ms) in this experiment.

4.4 Method

Participants

Thirty-two undergraduate students of Tilburg University participated in exchange of course credits or a monetary reward (20 women, 11 men, M = 20.4 years, SD = 1.8). All participants had normal or corrected-to-normal vision and gave informed consent according to the declaration of Helsinki. The protocol was approved by the local Ethics Committee Faculteit Sociale Wetenschappen of Tilburg University.

Stimuli and procedure

Eight photos of four male actors expressing fear or combing their hair were selected from a well validated photoset described in (Stienen & de Gelder, 2011). The stimuli were from the same set as described in experiment 1 with the exception that the colors were saturated to white and black. This was done to remove extra line elements because of the wrinkling of the clothing of the actors making it easier to mask the bodily expressions. Average height of the bodies was

8.14 degrees (SD = .35 degrees), the average maximum width (distance between the hands) was 3.12 (SD = .25 degrees) degrees and the average waist was 1.57 degrees (SD = .07 degrees) of visual angle. See Figure 4.3 for examples of the stimuli.

The auditory stimuli consisted of a Dutch spoken sentence “met het vliegtuig” (which means “with the plane”), edited in order to express different levels of emotion on a 7-step continuum between fearful and happy. The editing consisted of adjusting the duration, pitch range and pitch register. The voice clips lasted on average 792 ms (SD = 51 ms). See for more details (de Gelder & Vroomen, 2000).

Thus, the emotional dimension was only matched for fear and not for happy. The main reason was that we conjectured that if the unseen bodily expressions were both emotional this could lead eventually to a mixed effect. If in one trial the emotional expression would be happy and in the other it would be fearful the effect on the participants would be unpredictable. When only using neutral and fearful bodily expressions one can be sure that if there would be an effect, it would be in the direction of fear induction.

A pattern mask was constructed by cutting the target bodies into asymmetric forms which were scrambled and replaced in the area occupied by the bodies (see Figure 4.3). The rationale behind creating a new mask for this study was to avoid inducing any percept of a body. The mask measured 9.85 by 6.48 degrees of visual angle and completely covered the area of the stimuli.

A trial started with a white fixation cross on a gray background. After 500 ms a voice clip was presented. On the onset of this voice clip the masked fearful bodily expression, the neutral bodily action or the no-body (mask) stimulus was presented for 33 ms and subsequently the mask for 50 ms. The no-body condition was added to create a baseline in which neither the neutral action or the fearful expression was presented, instead the mask was presented for 88 ms. In 22 percent of the trials the fixation cross turned 45 degrees clockwise and switched back to the original position after 133 ms. See Figure 4.3 for a schematic example of a trial.

The participants were instructed to categorize the emotion in the voice clip as fearful or happy. Whenever the cross turned clockwise they had to withhold their response. This functioned as a catch trial to make sure that the participants were looking at the screen when the displays of emotional body postures were presented. The participants were told that we were interested whether the recognition of emotion in the voice is influenced when the perceptual system is loaded with visual information. This was done to provide the participants with a reasonable explanation why they saw the mask during the experiment and why the catch trials were presented as well as it ensured that they were naive to the actual goal of the experiment.

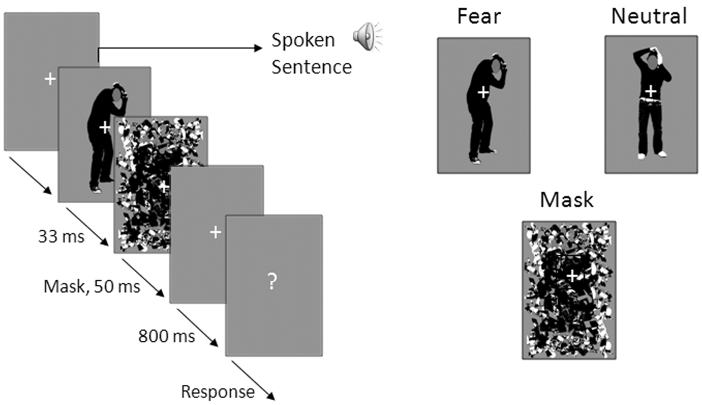


Figure 4.3 *An example of a trial of experiment 2 (**left**), an example of a fearful bodily expression and a neutral action (**upper right**) and the mask (**below right**).*

There were two experimental runs with a total of 216 trials (2 runs consisting of 108 trials: 4 identities x 3 masking conditions (fearful expression, neutral action, no-body) x 7 (emotional voice) + 24 catch trials). Every 54 trials there was a 2 minute break. The experiment was preceded with a practice session and was followed by a short validation session. The total duration of the experiment was 1 hour.

To check whether the participants had been unaware of the body stimuli we conducted an extensive semi structured exit interview and a sensitive posttest. In the exit interview we began by asking general questions such as “What do you think about the experiment?" and subsequently tuned in to find out whether the participants had been aware of the body stimuli. The questions ranged from "Have you noticed anything during the experiment?" to "Have you been distracted by anything?" to finally just asking them "Have you seen for example footballs, faces, bodies or shoes?". Only participants that never indicated having seen a body stimulus or even something like an object were included in the analysis.

Finally, in a posttest the 9 stimuli that were used in the main experiment (4 male fearful expressions, 4 male hair combing actions and the mask) and 40 new bodily expressions (4 female fearful expressions, 4 male and female angry bodily expressions, 4 male and female happy bodily expressions, 4 female hair combing actions, 4 male and female phoning actions and 4 male and female drinking actions) were presented. The participants were instructed to classify the stimuli as seen if they recollect that they have seen the bodily posture during the main experiment and as not seen when they could not recollect the bodily posture. The stimuli all were presented twice and the presentation duration was 33 ms which was enough to clearly see the body. Proportion classified as seen when it was a new stimulus was subtracted from proportion classified as seen when it was an old stimulus. Because the masks were included in the posttest and it was possible to detect the masks easily during the main experiment it was expected that the participants would at least identify the masks. This would result in having seen 2 out of totally 18 of the stimuli used in the experiment and 0 out of totally 80 of the new bodily expressions. The resulting value would then be .11 (2/18 – 0/80). Participants scoring above .11 were excluded from the analysis.

4.5 Results and Discussion

Posttest and Exit Interview

Seven out of thirty-two participants were excluded from analysis because their score was higher than .11 on the posttest. These participants also indicated in the exit interview having seen several body stimuli.

Main Results

One participant was discarded from analysis because he missed 26 percent of the catch trials (group mean = 2.0 %, SD = 4.7 %). In the validation session the fearful bodily expressions were correctly identified in 92.7 percent of the cases (SD = 15.6) and the neutral action was correctly identified in 95.8 percent (SD = 12.0) of the cases.

The no-body masked condition was used as baseline. The number of fear responses were corrected for this baseline performance per morphed emotional voice condition separately for masked fearful bodily expressions and neutral bodily actions, see Figure 4.4. A value of zero meant that the emotional sentence was not more or less categorized as fearful when a fearful bodily expression or a neutral action was shown in comparison to when no masked bodily stimulus was presented. A 2 (fearful bodily expression, neutral action) * 7 (sound) GLM repeated measures analysis indicated a significant interaction between the masked bodily expressions and the sound on the fear responses ($F(3,61) = 8.11$, $p < .001$, the Greenhouse–Geisser epsilon is reported because sphericity could not be assumed). This shows that the masked body stimuli influenced the categorization of the emotion in the voice and that this difference depended on which morphed sentence was presented to the participants. Bonferroni corrected paired t-testing (7 comparisons, thus $\alpha = .05/7 = .007$) were performed between fear responses to the voice when fearful or neutral bodily expressions were presented. This revealed that when the voice was slightly more fearful than happy and masked neutral pictures were presented participants categorized the voice more as being fearful (mean = .07, SD = .14) than when masked fearful pictures were presented (mean = -.021, SD = .16), t(23) = -3.252, $p = .004$. Interestingly, when the voice was a 50/50 morph between fearful and happy participants classified the voice more as being fearful when masked fearful bodily expressions were presented (mean = .04, SD = .25) in comparison to when masked neutral bodily actions were presented (mean = -.10, SD = .23), t(23) = 3.129, *p* = .005. See Figure 4.4.

We were primarily interested whether the bodily expression while unseen exerts its influence on the perceived emotion in the voice. This study revealed that when fearful bodily expressions and neutral actions are presented outside visual awareness they still influence the interpretation of the prosody in spoken words. Unseen fearful bodies triggered more fear responses when the emotion of the spoken sentence was a 50/50 morph of both emotions.

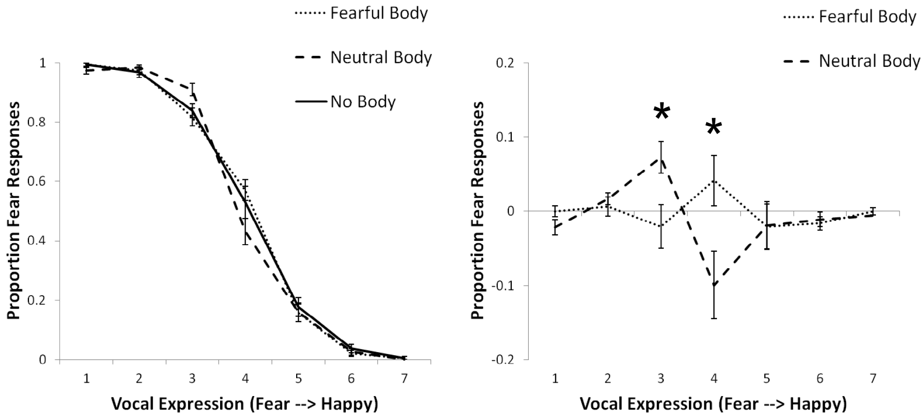


Figure 4.4 *Fear responses as a function of morphed emotional spoken sentences when masked neutral actions, fearful bodily expressions or no bodies were shown (**left**). Fear responses corrected for baseline performance (no-body trials) as a function of morphed emotional spoken sentences when masked neutral actions or masked fearful bodily expressions were shown (**right**). Error bars represent standard error of the mean. Asterisks indicate p < .001.*

The results leave us wondering why fear responses increased when the voice was slightly more fearful but the unseen bodily expression was neutral. It may be the case that this is caused by the mismatching of the emotional dimensions of the two sensory signals. The ambiguity that is introduced when the voice is fearful but the visual stimulus is neutral could have confused the participants. The unseen neutral bodily expressions did not, in fact, deliver extra information which could help processing the auditory signal. Alternatively, it might be that although the validation results were good, on an unreportable level the neutral bodily expression might be perceived as being fearful. This is a possibility which suggest further research on this intriguing question like developing a stimulus set which is not only validated explicitly but also with the use of autonomous responses such as pupil dilation or skin conductance.

The duration of the vocal stimulus was much longer than the duration of the masked visual stimulus. Although this study mainly focused on the influence of masked bodily expressions on the processing of overtly presented verbal sentences the large discrepancy might have attenuated the effect skewing the results towards the vocal stimuli. It might be that with shorter clips such as were used in experiment 1 lead to larger effects.

4.6 General Discussion

Our goal was to investigate whether the emotional voice influences the recognition of emotional bodily postures independently of visual awareness and whether unseen emotional bodily expressions influence the recognition of the emotion expressed in the voice. The results of the first experiment showed that dissociation occurred between objective and subjective measures. When SOA latencies were short the objective categorization performance was still facilitated by the congruent emotional voice while this facilitation effect was absent in the subjective confidence ratings. We conclude that the emotional voice influences the categorization of emotional body postures independently of visual awareness because participants seemed not to be aware while they were categorizing the emotional bodies better than chance. The second experiment showed that bodily postures presented outside visual awareness still influenced the interpretation of the emotion in the voice. When the bodily expression was fearful participants categorized the voice as being more fearful when the voice was a 50/50 morph between fearful and happy. Surprisingly masked neutral bodily actions triggered more fear responses to the voice than when the voice was already slightly more fearful.

In the second experiment a trial-by-trial measurement would have been possible except that this conflicts with the goal to present bodily expression outside the visual awareness of the participants. Therefore we combined an extensive semi-structured exit-interview with a sensitive posttest. During the exit-interview it was ensured to give the participants as much space as possible to express their experience they had during the experiment. If there was just the smallest hint towards reporting any bodily postures or even objects, it resulted in exclusion from analysis. In addition, we applied a strict criterion to the posttest which dictated that if any of the emotional postures was reported as seen, the participant was excluded from the analysis. Given the fact that only 7 out of 31 participants were excluded while the criteria were strict and the tests were sensitive it supports our assertion that the masking of the targets was effective.

Our findings are consistent with earlier studies showing the crossmodal influence of human emotional sounds on the recognition of emotional body postures (Van den Stock, Grèzes et al., 2008) and the influence of emotional body postures on the interpretation of voice prosody (Van den Stock et al., 2007). The study performed here adds the important notion that this crossmodal interaction is even taking place when the observer is not aware of the visual information. In addition, emotional information from one modality can influence the emotional information from another modality independently of visual awareness.

The influence of facial expressions of which there is no sensory awareness on the processing of emotional voices was already shown in hemianopic patients (de Gelder et al., 2002). Our study now generalizes these findings to healthy participants and to bodily expressions. When conscious processing of visual signals by the cortical mechanisms via the striate cortex is prevented, the colliculo-thalamo-amygdala pathway could still process the stimulus. This was already shown in recent fMRI studies that have suggested differential amygdala responses to fearful faces as compared to neutral faces when the participants were not aware of the faces (Morris, Öhman, & Dolan, 1999; Whalen et al., 1998). It would be interesting to evaluate these processing pathways in order to study the neurofunctional basis of how these signals interact in absence of visual awareness.

Future research should reveal how the results of the present study generalize to other emotions and different contexts to investigate the influence of environment on the affective multisensory integration. In addition, it would be interesting to see how the integration of other sensory modals is influenced such as haptics or smell. This field of research will give rise to insights in that affective signals often require a rapid reaction from the observer and intersensory redundancy, so it is assumed, contributes to speed by reducing uncertainty.

5. A computational feed-forward model predicts categorization of masked emotional body language for longer, but not for shorter latencies

Abstract

Given the presence of massive feedback loops in brain networks, it is difficult to disentangle the contribution of feedforward and feedback processing on the recognition of visual stimuli, in this case, of emotional body expressions. This study explores how well feed-forward processing explains rapid processing of this important class of stimuli. By means of parametric masking it may be possible to control the contribution of feedback activity in human participants. A close comparison could be made between human recognition performance and the performance of a computational neural model which exclusively modeled feed-forward processing and was engineered to fulfill the computational requirements of recognition. Results show that the longer the SOA (Stimulus Onset Asynchrony) latency the closer the performance of the human subjects was to the predicted values by the model with an optimum when the SOA was 100 ms. On short SOA latencies, however, the human performance deteriorated, but the categorization of the emotional expressions was still above baseline. The data suggest that although theoretically it is likely that feedback arising from infero-temporal cortex is blocked when the SOA is 100 ms, human participants seem to rely on more local visual feedback processing to equal the model's performance.

Stienen, B. M. C., Konrad, S., & de Gelder, B. (in press). A computational feed-forward model predicts categorization of masked emotional body language for longer, but not for shorter latencies. *Neural Computation.*

5.1 Introduction

Humans are capable of categorizing extremely quickly - and accurately - a wide variety of natural visual stimuli. Recent evidence suggests that this capability may be due to a fast feed-forward processing stream involving brain networks specialized in certain types of stimuli (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001). The aim of the present work is to shed some light on how well feed-forward processing explains rapid processing of an important class of stimuli represented by human body postures. To this end we compare a computational model of feed-forward categorization to a behavioral experiment in which the available processing time is carefully limited.

In previous decades a number of research reports focused on the processing of faces and their expressions in order to explore how we process emotions and many computational models have been offered. More recently researchers started to investigate the issue of bodily expression recognition. By switching to this category, we may be capable of extending the scope of face-based research and provide evidence that human emotion theories may generalize to other affective signals as well (de Gelder, 2006, 2009). Results from a number of behavioral experiments using independent stimulus sets now persuade us to conclude that recognition of expressions is similar for face and body stimuli. Available literature has already firmly established that emotional body expressions clearly and rapidly convey the emotional, intentional and mental state of a person (Meeren et al., 2005; Stekelenburg & de Gelder, 2004) and that full awareness of the visual stimulus or intact striate visual cortex are not essential (de Gelder, Vroomen et al., 1999; Stienen & de Gelder, 2011; Tamietto et al., 2009; Tamietto & de Gelder, 2010).

Schindler, van Gool and de Gelder (2008) have shown that a computational neural model which exclusively modeled feed-forward processing and was engineered to fulfill the computational requirements of recognition was capable of categorizing a set of 7 different emotional bodily expressions in much the same way as human observers did. However, there was no time limit on the presentation of the bodily expressions in the human categorization task. Given the presence of massive feedback loops in brain networks, it is unclear whether human performance was only based on feedforward brain processes without a significant contribution from feedback processes. By controlling the contribution of feedback activity in human participants a close comparison between the brain networks and the assumptions of the model is then possible.

Masking is one of the most widely used techniques for exploring unconscious processing of visual emotional information in neurologically intact observers and seems an excellent technique to control the contribution of feedback processes to visual processing. For example, Esteves and Öhman (1993) found that short duration (e.g. 33 ms) presentations of facial expressions (happy and angry) replaced immediately by a neutral face (mask) with a longer duration (e.g. 50 ms), are below the participants' identification threshold.

Lamme and Roelfsema (2000) and Lamme (2006) argue that a visual stimulus activates the visual cortex (striate and extrastriate) between 40 and 80 ms. Next, the infero-temporal cortex (IT) is feedforward-activated starting from 80 ms. Feedback signals arise from this area re-entering the visual cortex. Assuming 1 to 3 nodes that separate IT and visual cortex and a maximum firing rate of 100 Hz for cortical neurons (Rennie, Wright, & Robinson, 2000) the signal re-enters the visual cortex between 90-110 ms after the onset of the target. This means that a mask could interfere with re-entrant signals arising from IT when presented up to 110 ms after presentation of a target. In other words, it is increasingly more likely that feedback is possible from the infero-temporal cortex when the SOA (Stimulus Onset Asynchrony) latency and thus processing time of the target increases.

Neurological evidence indicates that masking selectively disrupts re-entrant signals to V1. For example, Lamme, Zipser and Spekreijse (2002) showed that masking seemed to selectively interrupt the recurrent interactions between V1 and higher visual areas in the macaque monkey brain. Fahrenfort, Scholte and Lamme (2007) found in a human EEG study that when a texture-defined square is being masked with an SOA of 16 ms ERP's typically associated with re-entrant processes turned out to be absent in the masked condition. No differences in bilateral occipito-temporal areas were found before 110 milliseconds while more posterior ERP's triggered by seen stimuli started to differ from ERP's triggered by unseen stimuli in later components.

However, the nature of the masking effect still remains a matter of discussion. The masking effect could be a consequence of imprecise temporal resolution, starting as early as in the retina. This is called ‘integration masking’. Alternatively, the masking effect could arise by interruption of the target processing on more higher level areas involved in object recognition or in this case, bodily expression recognition (see e.g. review by Enns & Di Lollo, 2000).

Here we presented participants with masked emotional bodily expressions using a parametric masking procedure to disentangle the contributions of feedback processing to the categorization of the bodily expression. Five emotional expressions (including neutral) were presented to the participants while the onset between target and mask (SOA, Stimulus Onset Asynchrony) was parametrically varied between 33 and 133 ms. The participants were instructed to categorize the emotion and use their intuition whenever they could not clearly see the target stimulus. The same set of stimuli was cross-validated using the neural model designed by Schindler et al (2008) and the outcomes were compared. In addition, the neural model was tested on linear combinations of bodily expressions and the mask to explore how the model performs on noisy images.

It is expected that up to an SOA of 100 ms feedback processes arising from IT would be blocked by the mask. Theoretically full feedback should be possible when the SOA is 133 ms or longer. If human participants can categorize bodily expressions in the absence of information carried by feedback processes, then the model should predict the human performance when SOA latencies are 100 ms or shorter.

5.2 Method

MASKING STUDY

Participants

Twenty-two undergraduates of The University of Tilburg participated in exchange of course credits or a monetary reward (12 women, 10 men, M = 21.6 years, SD = 3.2). All participants had normal or corrected-to-normal vision and gave informed consent according to the declaration of Helsinki.

Stimuli and procedure

The same photoset was used as in the study by Schindler et al. (2008). This time only angry, fearful, happy, sad and neutral bodily expressions were used and the faces were covered with an opaque gray mask. It was decided to use five categories instead of seven for pragmatic reasons. We did not want to make the button pressing too complicated and secondly, we wanted to keep the experiment within reasonable time limits. The principle reason for our selection of emotions was that "surprise" and especially "disgust" do not have an obvious bodily expression, as opposed to facial expressions, where they are quite clear. Neutral bodily postures of 6 actors were used to construct a mask. A male and a female with an average posture were chosen as the basis. Using Adobe Photoshop 7.0 © these actors were fused together. Arms and legs from the four other identities expressing a neutral emotion were attached to the body at different positions and orientations creating the image of two bodies with more arms and legs than usual (see Figure 5.1). Average height of the bodies was 8.83 degrees; the average width was 3.41 degrees (distance to the screen was 90 cm). The height of the mask was 10.40 degrees; the width was 6.27 degrees covering the area where the target stimuli were presented completely.

The stimuli were presented on a 17" PC screen with the refresh rate set to 60 Hz. We used Presentation 11.0 to run the experiment. A white cross of 1.22 x 1.22 degrees was used as a fixation mark in the center of the screen. Finally, all stimuli were pasted on a gray background.

Participants were comfortably seated in a chair in a soundproof experimental chamber. A trial started with the white fixation cross on a gray background. The disappearance of this cross signaled the beginning of a trial. After 500 milliseconds the target stimulus appeared for 33 milliseconds. After a variable interval the mask was presented for 50 milliseconds. The SOA latencies were 33, 67, 100 and 133 milliseconds. The actual presentation time was calibrated with the use of a photodiode and an oscilloscope measuring the latency between onset of the target and the onset of the mask. Moreover a target-only condition and a mask-only condition were included. After the categorization response a fixation cross appeared until the trial time was 3000 milliseconds.

Participants were instructed to categorize the target body expressions as angry, fearful, happy, sad or neutral. They responded with two hands using the ring, middle and index finger of the left hand and the index and middle finger of the right hand. The response buttons were labeled with the letter corresponding to the category and a reminder with the full names was situated on a board in front of them underneath the monitor. There were 5 between subjects counterbalance schemes making sure that each label occurred on every position once. They were instructed to be as accurate as possible but that the time for responding is short so they had to respond fast and to use their "gut feeling" if they had not seen the body.

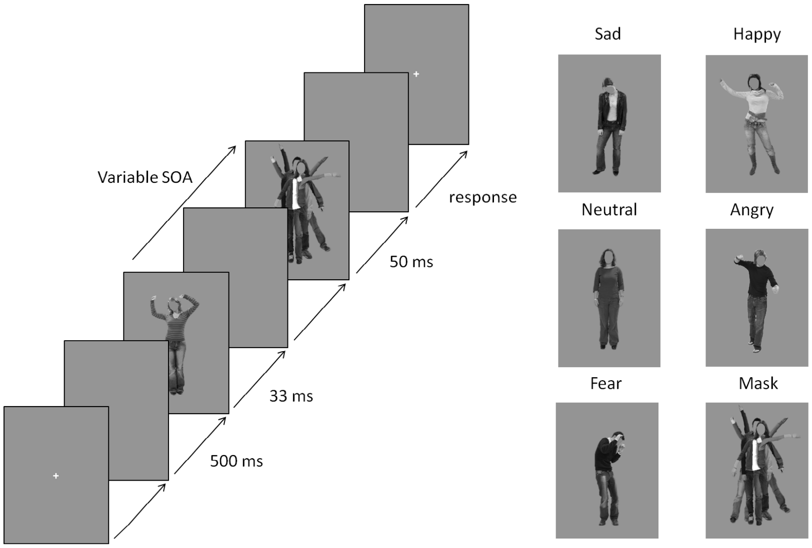


Figure 5.1 *An example trial (**left**). A typical example of each stimulus category (**right**).*

Prior to the experimental sessions the participants performed two practice sessions consisting of 60 trials each. Other identities than the ones used in the main experiment served as targets. When the participants did not miss trials and gave notice of a full understanding of the procedures the main experiment was started. One complete run summed up to a total of 1230 trials (41 identities x 5 postures (4 emotions + neutral) x 6 timing conditions (including target-only and mask-only) which were randomly presented. Every 160 trials there was a break. After the main experiment all targets were presented for 33 milliseconds to validate the stimuli used as targets. The instruction remained the same for this session. The experiment lasted 2 hours in total.

NEURAL MODEL

The computational model has been inspired by the ones of Riesenhuber and Poggio (1999) and Serre, Oliva, and Poggio (2007). It consists of a four-layer feed-forward hierarchy: each processing layer converts the inputs from the previous layer to a set of output features of higher complexity and/or larger receptive field. The input to the bottom layer is the raw image, whereas the output of the top layer is a score for each of the possible categories. A schematic illustration is given in Figure 5.2. For further details please refer to Schindler, van Gool and de Gelder (2008). The model was used without modification, thus the only difference to the original work is that in the present study the model was fed only five categories (four emotional bodily expressions and one neutral body pose) rather than seven.

To test for the possibility that the processing of the mask interfered with the early stages of processing the bodily expressions, which may the case when integration masking occurs, we tested the neural model with pixelwise linear combinations of the bodily expressions and the mask. We created three different stimulus sets by choosing three different weight ratios between the target and the mask: 1) 0.8*target + 0.2*mask (Mix_1), 2) 0.5*target + 0.5*mask (Mix_2), and 3) 0.2*target + 0.8*mask (Mix_3).

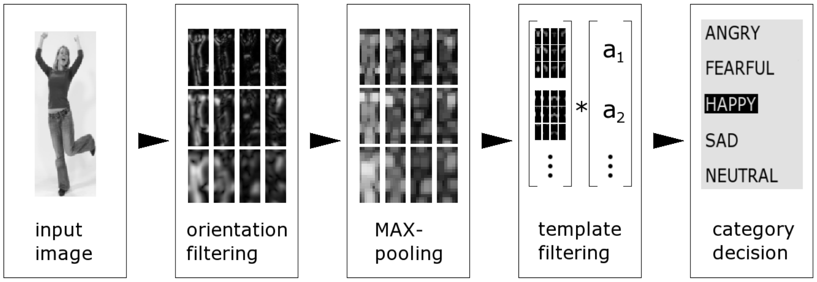


Figure 5.2 *The computational model. From the raw image, local orientations are extracted at multiple scales, pooled over spatial neighborhoods, and compared to learned complex feature templates. The similarities with all complex features are fed into a discriminative (forced-choice) classifier. Parameters were chosen for illustration purposes and are different from the actual implementation.*

5.3 Results

Trials where participants failed to categorize the bodily expression within the duration of the trial were discarded (0.4 percent of the trials, SD = 0.6). One participant was discarded as an outlier in the validation session. While the group was on average 91.3 percent (SD = 4.7) correct in categorizing the body postures, this participant was more than 3 standard deviations below this average. The validation scores for angry, fear, happy, sad and neutral expressions were respectively 81.8 (SD = 10.6), 94.5 (SD = 6.4), 97.5 (SD = 2.7), 84.6 (SD = 8.0) and 98.3 (SD = 2.3) percent correct.

To calculate Chi-square distances between the observed human performance and the performance of the model we used the basic definition $\chi^2 = \sum ((Fo\text{-}Fe)^2/Fe)$ where *Fo* is the observed correctly categorized stimuli per emotional category and *Fe* is the performance of the neural model per emotional category. The 5-D vector was fed into the Chi-square for each subject separately. Finally Chi-square distances were averaged, see Figure 5.3. When the SOA was 100, 133 milliseconds or when no mask was presented, the model predicted the human performance significantly well (resp. $\chi^2(4, N = 22) = 7.25, p > .05$; $\chi^2(4, N = 22) = 4.52, p > .05$; $\chi^2(4, N = 22) = 3.49, p > .05$).

As shown in Figure 5.4 while being below the predicted performance of the model, participants still performed above baseline for the expressions fear, happy and sad when the SOA was 33 milliseconds (all $p < .05$), and when the SOA was 67 milliseconds the participants categorized all expressions above baseline ($t(20) = 2.81$, $p < .05$).

To look more in depth which among the higher SOA conditions matched the model best we analyzed the common misclassification between model and human participants. We counted a stimulus as misclassified when the number of correct classifications was more than 1 standard deviation below average per SOA condition or, in the case of the model, below average performance. Because each unique stimulus was only shown once per SOA to the participants the number of correct classifications were indexed on the group level. Next, we indexed how many stimuli were misclassified by both the human participants and the model together.

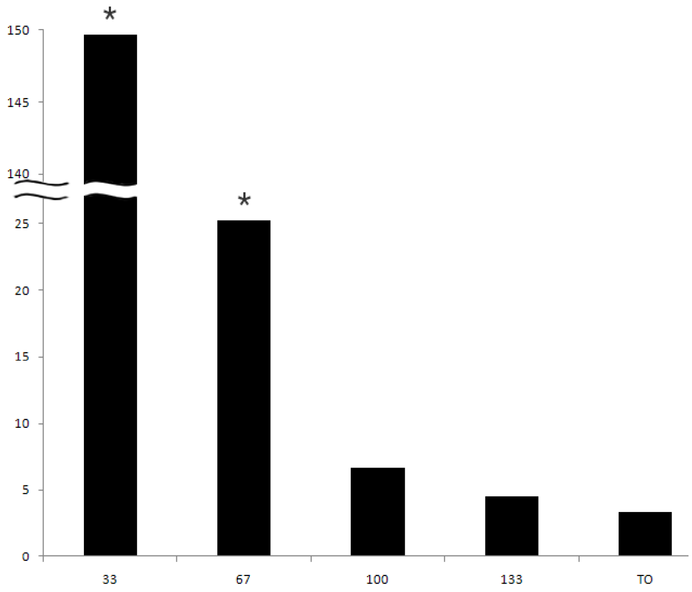


Figure 5.3 *Chi-square distances between neural model and the human performance per SOA condition. Asterisks indicate that the human's performance was different from the neural model. TO = Target-Only.*

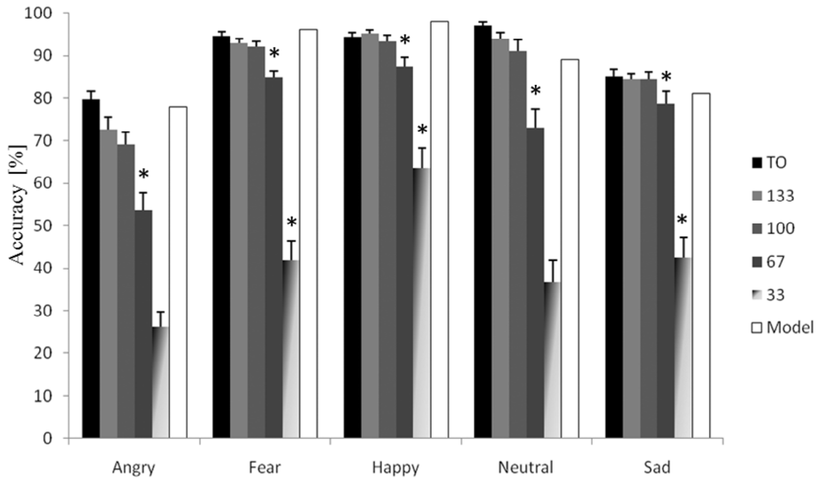


Figure 5.4 *Accuracy in percentages per emotion category per SOA condition. Asterisks indicate performance above baseline. Error bars indicate standard error mean. TO = target-only.*

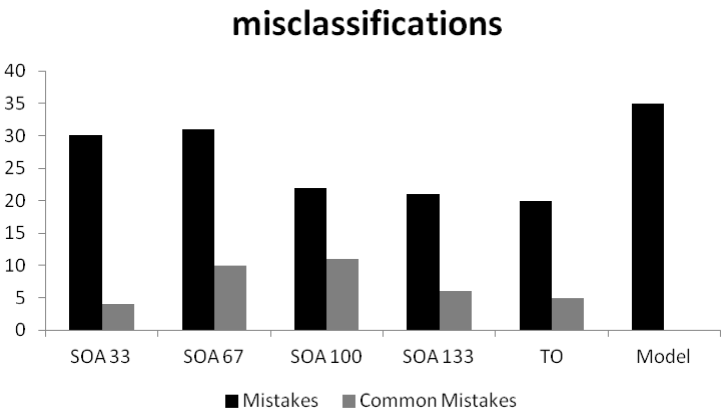
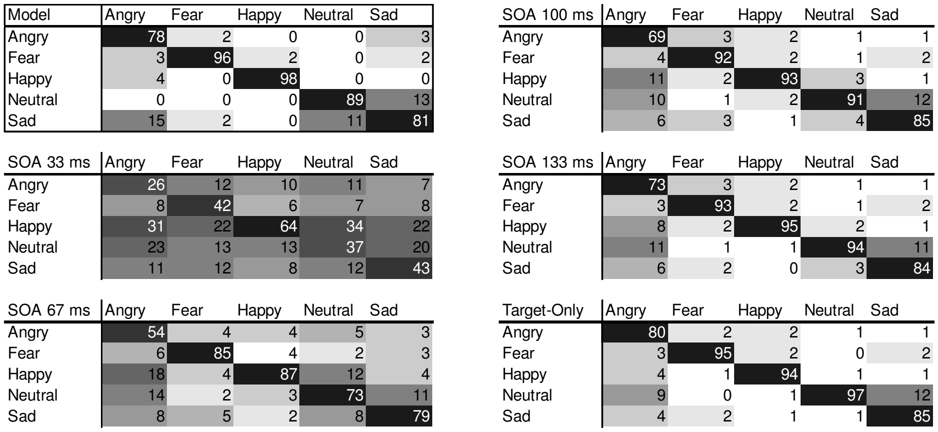


Figure 5.5 *The total misclassifications per SOA by the participants and by the model (Mistakes, shown in black) and the amount of stimuli that were misclassified by both the participants and the model (Common Mistakes, shown in grey).*

Figure 5.5 shows that the longer the SOA the smaller the number of misclassifications. Interestingly, the total common misclassifications by model and humans increases until the SOA is 100 ms and decreases again when the SOA is longer.

Confusion matrices were also taken into account. Table 5.1 shows an overview of the confusions that were observed in the model, respectively the human subjects. As can be seen by the grayscale color coding of the cells, the higher the SOA the more the model seems to predict the actual human behavior. In Table 5.2 the absolute differences between predicted and observed values are shown. Chi-square tests were not performed on these data because not all assumptions were met, e.g. not all cell values were larger than 5. The major differences between the confused categories of the model and the human participants was that the humans confused angry with neutral, while the model confused angry dominantly with sad. When no mask was presented the human participants, contrary to the model, did not confuse neutral with sad.



Model	Angry	Fear	Happy	Neutral	Sad
Angry	78	2	0	0	3
Fear	3	96	2	0	2
Happy	4	0	98	0	0
Neutral	0	0	0	89	13
Sad	15	2	0	11	81

SOA 100 ms	Angry	Fear	Happy	Neutral	Sad
Angry	69	3	2	1	1
Fear	4	92	2	1	2
Happy	11	2	93	3	1
Neutral	10	1	2	91	12
Sad	6	3	1	4	85

SOA 33 ms	Angry	Fear	Happy	Neutral	Sad
Angry	26	12	10	11	7
Fear	8	42	6	7	8
Happy	31	22	64	34	22
Neutral	23	13	13	37	20
Sad	11	12	8	12	43

SOA 133 ms	Angry	Fear	Happy	Neutral	Sad
Angry	73	3	2	1	1
Fear	3	93	2	1	2
Happy	8	2	95	2	1
Neutral	11	1	1	94	11
Sad	6	2	0	3	84

SOA 67 ms	Angry	Fear	Happy	Neutral	Sad
Angry	54	4	4	5	3
Fear	6	85	4	2	3
Happy	18	4	87	12	4
Neutral	14	2	3	73	11
Sad	8	5	2	8	79

Target-Only	Angry	Fear	Happy	Neutral	Sad
Angry	80	2	2	1	1
Fear	3	95	2	0	2
Happy	4	1	94	1	1
Neutral	9	0	1	97	12
Sad	4	2	1	1	85

Table 5.1 *The confusion matrices of the model (**upper left**) and the human participants. Columns represent true emotion; rows represent the percentage of the emotional category chosen. The cells are grayscale color coded using the logarithm of the percentage.*

	Angry	Fear	Happy	Neutral	Sad
SOA 33 ms					
Angry		9.53	9.54	10.82	4.11
Fear	5.50		3.92	6.87	5.71
Happy	27.50	21.62		33.74	22.32
Neutral	22.70	12.57	13.27		7.27
Sad	3.80	10.33	7.67	0.88	
Total	59.50	54.04	34.39	52.31	39.42
SOA 67 ms					
Angry		2.24	3.64	5.37	0.03
Fear	2.70		2.49	2.10	1.05
Happy	14.41	3.78		11.77	3.98
Neutral	13.78	2.00	2.83		1.77
Sad	6.58	3.19	1.66	3.29	
Total	37.47	11.22	10.62	22.54	6.83
SOA 100 ms					
Angry		0.58	2.01	1.05	2.41
Fear	1.00		0.22	0.94	0.34
Happy	6.57	1.53		3.18	0.82
Neutral	10.22	0.70	2.02		1.26
Sad	8.90	1.07	0.83	7.27	
Total	26.68	3.88	5.07	12.43	4.84

	Angry	Fear	Happy	Neutral	Sad
SOA 133 ms					
Angry		0.71	1.53	0.82	2.06
Fear	0.04		0.35	0.71	0.34
Happy	3.71	1.77		1.76	0.70
Neutral	10.91	0.58	1.19		2.42
Sad	9.25	0.11	0.47	8.31	
Total	23.91	3.17	3.55	11.59	5.52
Target-Only					
Angry		0.11	2.23	0.59	2.17
Fear	0.41		0.24	0.24	0.36
Happy	0.22	0.82		1.18	0.93
Neutral	9.43	0.24	1.18		1.43
Sad	10.98	0.47	0.58	10.06	
Total	21.04	1.65	4.24	12.07	4.89

Table 5.2 *Absolute differences between model and human performance per timing condition. Cells colored black indicate that the difference between expected and observed value was greater than 2 standard deviations from the average of all absolute differences per timing condition.*

Figure 5.6a shows the averaged Chi-square distances between the categorization performance of the model when categorizing the mix_1, mix_2 and mix_3 stimuli and the performance of the human participants. Figure 5.6b shows the actual human performance per emotion per SOA, the original model performance and the categorization performance of the model when categorizing the mix_2 and mix_3 images. The longer the latency, the less the categorization of the mix_3 images by the model is comparable to the human performance. Interestingly, the performance of the model in categorizing mix 3 angry and sad postures seems to match the human categorization when the SOA was short (33 ms) better, while this is not the case for fearful, happy and neutral images. However, all comparisons were significantly different between human and model performance when categorizing the mix_3 images (all p < .001).

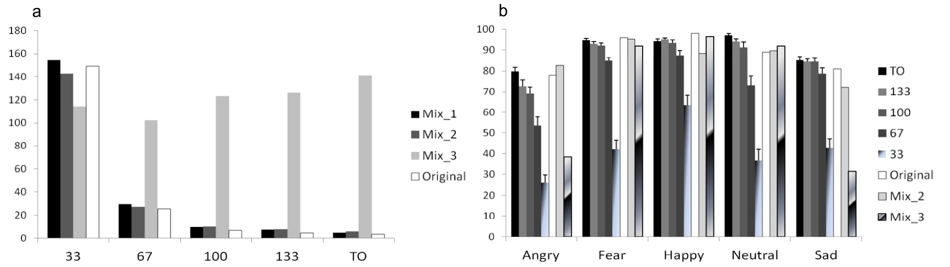


Figure 5.6 *Chi-square distances between neural model performance when cross-validated with the original, mix_1, mix_2 and mix_3 stimuli and the human performance per SOA condition (**a**). Accuracy rates in percentages per emotion category per SOA of the human participants (TO, 133, 100, 67, and 33) and model performance (Original, Mix_2, and Mix_3) (**b**). Error bars indicate standard error mean. TO = target-only.*

Finally, a 5 (emotions) x 5 (SOA) multivariate analyses of variance (MANOVA) showed that there was a main effect of emotion ($F(4,16) = 18.49$, $p < .001$) and SOA ($F(4,16) = 28,28$, $p < .001$) on the reaction times. Bonferroni corrected multiple comparisons show that angry bodily expressions are slower categorized in comparison to the other bodily expressions. All SOA conditions differed from each other significantly with the exception when the SOA was 100 and 133 ms. The general trend is that the longer the SOA, the shorter the reaction time.

5.4 Discussion

We have shown that a feed-forward computational model predicts the human categorization performance for emotional body language strikingly well. It appears that the longer the SOA the closer the performance of the human subjects matched the performance of the model with an optimum when the SOA was 100 ms. On short SOA latencies, however, the human categorization performance deteriorated, but was still above baseline. When testing the model with combinations of the mask and the bodily expressions also the performance of the neural model decreased, but its performance was still different from that of the human participants.

Based on the theoretic framework proposed by Lamme and Roelfsema (2000) one would expect that the performance of the feed-forward neural model would equal the performance of the human participants in the three shortest SOA conditions. Yet human participants are capable to perform the task better than chance when SOA is low but their performance is much worse than the neural model.

There are four explanations for this. Firstly, the model works in a context free environment and lacks the characteristic, as human participants do not, to be distracted by the environment for example by the processing of the mask itself. Alternatively, it would be interesting if, as proposed by Lamme (2006), one would be able to block re-entrant processing associated with bodily expressions with TMS as being done by Jolij and Lamme (2005) with schematic faces. This method loads the visual system less with distracting visual information and compare these results with the performance of a neural model as described here.

Secondly, it may be the case that the target and mask temporally overlap on the retinal level interfering with the processing of the bodily expressions on an early stage as could be the case in masking by integration (Enns & Di Lollo, 2000). We showed that although the neural model performed much worse when tested on the combinations of the mask and bodily expressions the performance still was different from the human performance. However, while the method used here gives some insights there are multiple ways to represent integration between two images on a retinal or cortical level. This multiple solution problem limits the interpretation of our current results. In addition, biases may be present when using the computational model, because contrary to the human visual system, the model learns only from the stimuli that are tested while it lacks exposure to the amount of images human participants are exposed throughout their lives. In addition, in natural context the human eye categorizes bodily expressions that appear in complex contexts.

Thirdly, when the SOA is 67 ms it happens to be close to the average required time of the feed-forward mechanism, and thus we observe a mixture of successful categorizations and random answers.

Fourthly, when the SOA is 100 ms it might be that local feedback processing of the target occurs. But when the SOA latencies are shorter, these local feedback processes are impaired. For example, it may be possible to make a distinction between recurrent activation originating from e.g. V3 or recurrent activation originating from e.g. IT. If we assume that V1 is activated 40 ms after target onset then V3 (via V2 or not) is activated 50-60 ms after target onset. If we assume that V3 feeds back directly to V1 then the re-entrant signal arrives here 60-70 ms after target onset. At these latencies the mask is already activating V1 when the SOA is 33 and 67 ms. In conclusion, when the SOA is 100 ms, feedback processes arising from IT are most likely to be disrupted, while when the SOA is shorter the more local feedback processing could be disrupted arising from extrastriate areas. This has important implications. For example, could it be that the conscious visual percept is disrupted when the SOA is 33 ms, while the human participants are conscious about the visual percept of the bodies but semantically categorize the bodies automatically when the SOA is 100 ms?

Pascual-Leone and Walsh (2001) showed that applying TMS to V1 after stimulating V5 in a time window of 5-45 ms led to a decrease in reporting that the TMS induced posphemes moved. In addition, a study of Koivisto, Railo, Revonsuo, Vanni, and Salminen-Vaparanta (2011) showed that recurrent interactions between ventral stream areas and V1/V2 are necessary for categorizing and perception of natural scenes. They found longer response times and degraded quality of subjective perception when applying single pulse TMS in the time window 90-210 ms to V1/V2 and longer response times when applying single pulse TMS to LO after 150 ms and longer. Jolij and Lamme (2005) found that when stimulating V1 110 ms after onset of a display with four smileys, participants had difficulties reporting the location, but not the emotion. It seems that feedback to V1 is necessary for visual awareness. These studies suggest that the processing of a given visual stimulus around 100 ms in V1 is crucial for conscious perception and to make perceptual decisions possibly because recurrent activation is necessary for this to occur.

Possibly there is another less accurate mechanism aiding the participants to classify the emotions. It is well known that subcortical structures play a role in visual perception. When the SOA was 33 ms three out of the four emotional body expressions (happy, fearful and sad) were recognized above baseline, while neutral and angry expressions were not. This result could be hinting at a subcortico-cortical pathway. When visual signals are prevented from being processed by the cortical mechanisms via the striate cortex, the colliculo-thalamo-amygdala pathway could still process the stimulus. This is in line with recent fMRI studies that have suggested differential amygdala responses to fear faces as compared to neutral faces when the participants were not aware (Morris et al., 1999; Whalen et al., 1998). However, this study lacks the additional measurement of e.g. subjective awareness to be conclusive on this topic (see e.g. Cheesman & Merikle, 1986).

Caution must be exercised before generalizing these findings to unconscious processing. While Esteves and Öhman (1993) found that an SOA of 33 ms rendered an emotional face invisible this is not found in this study. Stimulus specific properties in masking studies are known to modulate the sensitivity of the masking effect. See for a thorough review Wiens (2006). It could be that the arms formed a higher contrast against the background when there was no overlapping with the arms of the mask causing the above baseline performance. Further research is needed on this issue.

Our data indicate that the model and the human subjects confused more or less to the same degree sad bodily expressions with neutral ones. The major difference between the model and the human performance in terms of confusion is the fact that the model categorizes angry as sad, whereas the human subjects show a stronger bias to interpret angry poses as neutral. Some of the actors in the stimulus set expressed anger by a “controlled anger” pose, crossing their arms and tilting the head. The model tends to interpret these deviating poses as being sad, while the human subjects interpreted them as being neutral, possibly because they were attentionally biased towards the body and not the head (see Schindler et al. (2008) for more stimuli examples).

The fact that performance does not change a lot when the SOA latencies are 100 ms or longer deserves special attention. Assuming that the perceptual decision is made in V1 feedback from IT might be blocked by the mask when the SOA is 100 ms. The fact that there are no major performance changes when the processing time of the target increases and thus feedback processing arising from parietal-frontal areas are possible, suggests that in these kind of tasks participants do not rely on feedback coming from these higher areas. The only change was that there were fewer common mistakes between model and humans and that the confusion pattern changed slightly when no mask was presented.

To summarize, the feed-forward neural model predicts human behavior strikingly well although the model slightly outperforms the human participants. This study shows that it is likely that emotional bodily expressions can be recognized even when feedback from higher level areas are blocked, although they still might relay on some more local feedback processing (while the model does not).

6. Event-related repetitive TMS of posterior Superior Temporal Sulcus improves the detection of threatening human body postural changes.

Abstract

Perceiving others' emotions through their body movements and postures is crucial for successful social interaction. While imaging studies indicate that perceiving body emotions relies upon a wide network of subcortico-cortical neural regions, little is known on the causative role of different nodes of this network. We applied event-related rTMS over non-facial, body- and action-related extrastriate (EBA), temporal (pSTS) and premotor (vPM) cortices to test their active contribution in perceiving changes between two successive images of either threatening or neutral human body or animal postures. While stimulation of EBA and vPM showed no selective effect on threatening stimuli with respect to neutral ones, rTMS over pSTS selectively impaired neutral posture detection and increased the accuracy in detecting changes of threatening human postures with respect to all other experimental conditions. No such effect was found for animal stimuli. These results support the notion that pSTS is crucially devoted to the detection of socially relevant information concerning others' actions fostering the notion that amygdalo-temporo-cortical modulatory connections mediate perception of emotionally salient body postures.

Candidi, M., Stienen, B. M. C., Aglioti, S. M., & de Gelder, B. (2011). Event related repetitive transcranial magnetic stimulation of posterior superior temporal sulcus improves the detection of threatening postural changes in human bodies. Journal of Neuroscience, 31(48), 17547-17555.

6.1 Introduction

Humans have a refined ability to use their body language to interact with others as well as with animals of other species. A specific aspect of this skill is to express one's own emotional state by means of body movements (Darwin, 1872/1965; James, 1890). Correspondingly, this ability is paralleled by the well-developed adaptive capacity to understand others' emotional state through the decoding of their body movements and postures (de Gelder, 2006, 2009) which seems to occur in subcortical brain regions that are active when directly experiencing the same emotional state (de Gelder & Hadjikhani, 2006; de Gelder et al., 2004; Hadjikhani & de Gelder, 2003) as well as through cortical sensorimotor simulative mechanisms (Rizzolatti & Craighero, 2004).

Among other nodes of the sensorimotor cortical network, the posterior Superior Temporal Sulcus (pSTS) shows stronger activation for emotional and socially relevant body movement perception (Allison et al., 2000; de Gelder & Partan, 2009; Grèzes et al., 2007; Kret et al., 2011; Pichon et al., 2008; Puce & Perrett, 2003) probably due to its anatomo-functional connections with the amygdalae (Amaral & Price, 1984; Morris, Friston et al., 1998; Rotshtein et al., 2001; Sah, Faber, Lopez De Armentia, & Power, 2003). Body movements, however, are not always fully visible and are often only implied in body postures.

In recent decades researchers provided evidence for an occipito-temporal brain region specifically dedicated to the visual processing of neutral body images (Exstrastriate Body Area, Downing et al., 2001). In particular the right EBA is essential in discriminating morphological body details (Moro et al., 2008; Urgesi, Berlucchi, & Aglioti, 2004; Urgesi, Candidi, Ionta, & Aglioti, 2007) and its activity is influenced by the emotion expressed by the body (Kret et al., 2011; Peelen, Atkinson, Andersson, & Vuilleumier, 2007) suggesting a large-scale representation of the emotion expressed through the body.

The left vPM cortex is thought to be crucial for action representation and understanding and is generally considered to be a key-node of the "mirror neuron" system initially described in monkeys (di Pellegrino et al., 1992) consisting of bimodal visuo-motor cells which fire during real action execution as well as during the passive observation of the same movement (Kilner et al., 2009). The activation of premotor regions is facilitated by the emotional valence of the observed posture (de Gelder et al., 2004; Grèzes et al., 2007; Pichon et al., 2008) strengthening the notion that emotion perception is tightly linked to action programming (Darwin, 1872/1965). Consistent with this, it has been shown that monkeys' premotor cortex receives neural projections from the amygdaloid complex (Avendano et al., 1983) and that electrical stimulation of this area induces defensive motor behaviours (Cooke & Graziano, 2004; Graziano et al., 2002).

However, no study has thus far tested the causal contribution of right EBA, right pSTS and left vPM cortices in detecting threatening and neutral human body postural changes. Here we used event-related rTMS to transiently alter physiological neural activity of these regions and to investigate their respective role in perceiving threatening and neutral body and animal postural changes (Pinto & Shiffrar, 2009).

6.2 Methods

Participants

Sixteen participants (ten female) took part in the rTMS experiment (mean age 22 ± 1.6 year ± SD). Thirteen participants were right handed according to the Briggs and Nebes handedness inventory (Briggs & Nebes, 1975). All participants had normal or corrected to normal sight. The experimental procedures were approved by the ethics committee of the Fondazione Santa Lucia (Rome, Italy) and were in accordance with the ethical standards of the 1964 Declaration of Helsinki. None of the participants had neurological, psychiatric, or other medical problems or any contraindication for rTMS (Wassermann, 1998). All subjects gave their written informed consent prior to the beginning of the experimental procedure, were naïve to the aim of the study and were informed about its purpose only after all experimental procedures were completed. Subjects were paid for their participation in this study.

Twelve subjects (eight female) participated in a control no-rTMS experiment (mean age 21 ± 1,7 year ± SD). Ten participants were right handed according to the

Briggs and Nebes handedness inventory (Briggs & Nebes, 1975). All participants had normal or corrected to normal sight.

Stimuli

Body images were selected from a set of dynamic video clips showing the frontal view of 4 male actors in black clothing jumping with their fists toward the observer (threatening body) or neutrally swinging their arms along the side of their trunk (neutral body). Two different frames were selected per actor for each condition. To test for the specificity of EBA, pSTS and vPM cortex in the processing of human body postures we included threatening (snakes) and harmless (fishes) animal control images. By twirling the images plus and minus 30 degrees using the twirl tool in Adobe Photoshop software (Adobe Systems Incorporated) and maintaining the face of the animal in the original position, two different versions of each animal were created. Both humans and animals had the face covered by a grey mask. The presentation of all stimuli was managed to keep the mask in the same position to preclude the possibility for the subjects to base their judgment of postural change on any change in mask position (Figure 6.1).

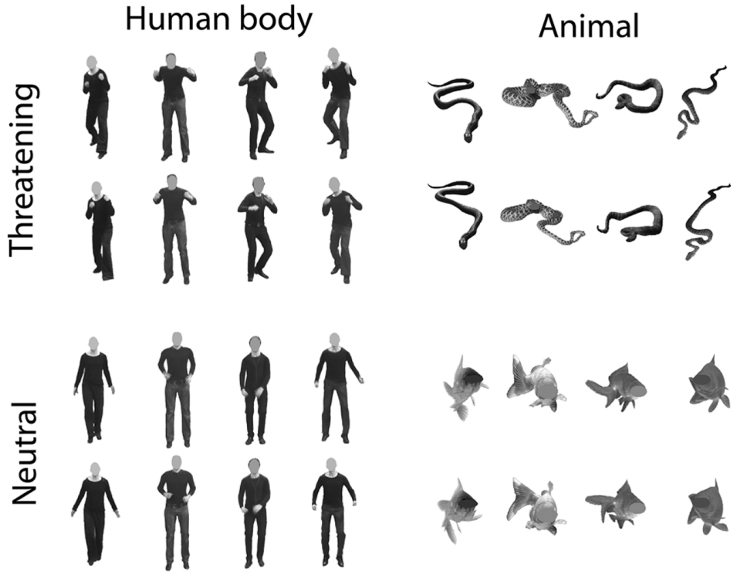


Figure 6.1 *Complete stimuli set of threatening and neutral animal and human body images.*

To measure the perceived intensity of motion induced by each pair of stimuli used in the main experiment, a group of 10 independent subjects were asked to rate their subjective perception of motion during the presentation of two successive images, which showed a postural change in 50 % of the trials. The timing of presentation of the stimuli and mask was the same in the behavioral and rTMS experiments. Different trials evoked higher sensation of motion with respect to same trials both in the threatening and neutral body ($p = 0.002$ and $p < 0.001$ respectively) and animal ($p = 0.006$ and $p < 0.001$ respectively) conditions. Moreover, as revealed by Bonferroni corrected t-tests against zero (the value zero represents absence of postural change perception) the perception of motion was only evoked during different trials of both threatening and neutral body and animal trials (all $p < 0.001$ uncorrected, all ps > 0.01 uncorrected during same trials n.s.) (Figure 6.2). Thus only the different trials were regarded as showing the presence of the critical signal (postural change detection) for the present purposes.

On debriefing participants were asked the following questions: 1) “What were the actors doing?”; 2) “Would you define some of these postures as emotional?”; 3) “If yes, which emotion do they express?”. Moreover, the participants performed a categorization task in which they had to label the seen postures as threatening or neutral. Average performance was 88 % correct for recognizing the threatening postures as fighting or threatening.

Transcranial Magnetic Stimulation

Participants wore a tightly fitting bathing cap on which scalp stimulation points were marked. Motor evoked potentials (MEPs) were recorded from first dorsal interosseous (FDI) muscle of the right hand. Surface Ag/AgCl electrodes were placed in a belly-tendon montage with the active electrode placed over the motor point and the reference over the interfalangeal joint. Electromyographic (EMG) signal was amplified at a gain of 1000x by a Digitimer D360 amplifier (Digitimer), band-pass filtered (20 Hz-2.5 kHz) and digitized (sampling rate: 10 kHz) by means of a CED Power 1401 controlled with Spike 2 software (Cambridge Electronic Design). The resting motor threshold (rMT), defined as the lowest intensity able to evoke five out of ten MEPs with an amplitude of at least 50 μV, was determined by holding the stimulation coil over the optimal scalp position (OSP). The OSP for inducing MEPs in the right FDI muscle was found by moving the coil in steps of 1 cm over the left primary motor cortex until the largest MEPs were found and then marked with a pen on a bathing cap worn by participants.

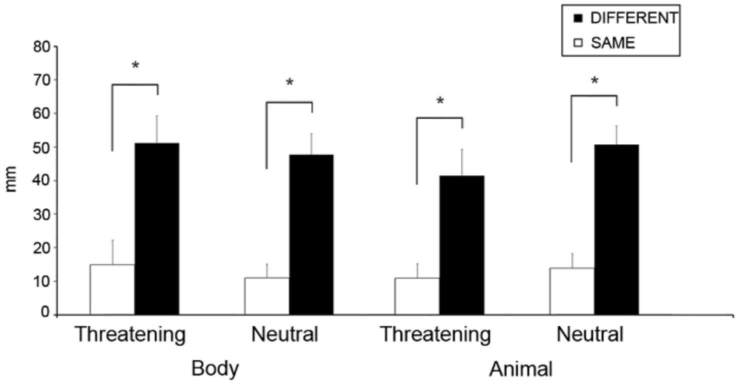


Figure 6.2 *Subjective ratings concerning the perceived intensity of postural changes in threatening/neutral bodies and threatening/harmless animals (millimeters ± s.e.m.).*

Stimulation sites were identified on each participant's scalp with SofTaxic Navigator system (EMS). Skull landmarks (nasion, inion, and two preauricular points) and about 60 points providing a uniform representation of the scalp were digitized by means of a Polaris Vicra Optical Tracking System (NDI, Canada). Coordinates in Talairach space (Talairach & Tournoux, 1988) were automatically estimated by the SofTaxic Navigator from an MRI-constructed stereotaxic template using an individualized probabilistic head model computation. This individualized head model preserves the anatomical scalp-brain correlates of a mean MR template, providing an accurate set of estimated MRI data, specific for the subject under examination. Furthermore, as the present design provided within-site controls no effect could be explained by errors in localizing the different sites within-individuals. The scalp locations that corresponded best to right EBA, pSTS and left vPM cortex coordinates as reported by a selected group of imaging and neurophysiologic studies (Bonda, Petrides, Ostry, & Evans, 1996; Candidi, Urgesi, Ionta, & Aglioti, 2008; Decety & Grezes, 1999; Grèzes et al., 2007; Grossman & Blake, 2002; Iacoboni et al., 2001; Michels, Lappe, & Vaina, 2005; Nishitani & Hari, 2002; Pelphrey, Viola, & McCarthy, 2004; Peuskens, Vanrie, Verfaillie, & Orban, 2005; Pichon et al., 2008; Urgesi et al., 2004; Urgesi, Calvo-Merino, Haggard, & Aglioti, 2007; Urgesi, Candidi et al., 2007; van de Riet et al., 2009) were identified and marked with a pen. Mean (± SD) coordinates corresponded to Brodmann area 37 in the posterior part of the middle temporal gyrus (right EBA), Brodmann area 22 in the posterior part of the superior temporal sulcus (right pSTS) and to Brodmann area 44 in the pars opercularis of the inferior frontal gyrus (left vPM) (Table 6.1).

	Right EBA	**Right pSTS**	**Left vPM**
x	51.7 ± 1.1	51.5 ± 1.2	-56.8 ± 0.9
y	–71.6 ± 1.5	–47.6 ± 0.9	10.3 ± 1.0
z	3.6 ± 0.7	9.0 ± 0.4	21.4 ± 0.6

Table 6.1 *Mean (± SD) Talairach coordinates of the three stimulation sites according to Talairach space.*

Repetitive TMS was performed by connecting two Magstim Model 200 stimulators with a Bistim module (The Magstim Company), producing a maximum output of 1.75 T at the coil surface (stimulus attenuation, 22%; duration, 1 ms; rise time, 110 μs). Two pulses were delivered with an interstimulus interval of 100 ms by means of a 70 mm figure eight stimulation coil (Magstim polyhurethane-coated coil). In keeping with two previous studies (Urgesi et al., 2004; Urgesi, Candidi et al., 2007), the first TMS pulse was delivered 150 ms after the onset of sample presentation. The same pulse delay was used for stimulation of pSTS and vPM cortex in keeping with magnetoencephalography and rTMS studies revealing activation of STS and ventral premotor areas within 150-200 ms after the visual presentation of moving body parts (Candidi et al., 2008; Nishitani & Hari, 2000, 2002; Urgesi, Calvo-Merino et al., 2007; Urgesi, Candidi et al., 2007). Stimulation intensity was 120% of the rMT for both pulses and ranged from 40% to 72% (mean = 53%) of the maximum stimulator output. During stimulation of EBA the coil was held by hand tangential to the scalp, with the handle pointing backward and medially at a 45° angle from the middle sagittal axis of the participants' head (Candidi et al., 2008; Urgesi et al., 2004; Urgesi, Calvo-Merino et al., 2007; Urgesi, Candidi et al., 2007). During stimulation of pSTS the coil was held tangential to the scalp, with the handle pointing backward and medially at a 45° angle from the middle sagittal axis of the participants' head. During stimulation of vPM cortex the coil was held tangential to the scalp, with the handle pointing backward and medially at a 45° angle from the middle sagittal axis of the participants' head (Candidi et al., 2008; Urgesi et al., 2004; Urgesi, Calvo-Merino et al., 2007; Urgesi, Candidi et al., 2007). The position of the coil with respect to the marks was checked continuously. During stimulation, participants wore commercial earplugs to protect their hearing. None of the participants reported phosphenes or hand muscular twitches after rTMS of EBA, pSTS and vPM cortex.

Procedure

Blocks in which subjects had to detect body or animal postural changes were presented separately in alternate order. The order of stimulation site was counterbalanced within each subject according to an ABCABCCBACBA pattern. The order of the stimulated site was counterbalanced across subjects. The starting block was alternated between animals and bodies and counterbalanced across participants. A short rest was allowed before proceeding to a different block. For each task, participants completed as much practice blocks as needed to reach accuracy level better than chance before proceeding to the experimental blocks. During the experimental session, two blocks of 32 trials (8 body neutral different, 8 body neutral same, 8 body threatening different, 8 body threatening same) were presented in each site stimulation condition, for a total of 16 trials per condition. We adopted a 2 x 3 (Valence x Site) factorial design.

Participants had their chin rested 57 cm away from a 17-inch monitor (resolution, 1151 x 964 pixels; refresh frequency, 60 Hz), where stimuli appeared on a white background and subtended a 18 x 9 centimeter region. Stimulus presentation timing, randomization and rTMS triggering were controlled by E-Prime v1.1 software (Psychology Software Tools Inc., Pittsburgh, PA).

A trial started with the presentation of a central fixation point lasting 1000 ms. The sample stimulus was presented for 150 ms. After the presentation of the sample a random-dot mask (18 x 9 centimeter in size) was presented for 200 ms. The masks were constructed by scrambling one image from the corresponding stimulus category by custom-made software created with Matlab (The MathWorks, Natick, MA). The custom-made software breaks down the initial image in a fixed number of dots maintaining their original grayscale. The obtained dots are randomly distributed in a space of fixed dimensions. After the disappearance of the mask, the probe stimulus appeared on the screen for 150 ms. A blank screen was then presented until a response was made (Figure 6.3). The first pulse was delivered in coincidence with Mask onset, thus 150 ms after sample presentation. Participants were asked to respond as quickly as possible by using their index or middle finger to press the left or the right key, respectively, on a custom-made response box. Crucially, the instruction was identical in the body and animal condition so that any differential modulation of rTMS on a given task was likely to occur at an entirely implicit level. Each key corresponded to the answer “same” or “different”. The responding hand and the finger used to press the “same” and “different” button were counterbalanced across participants. Each participant was tested in a single experimental session lasting approximately 2 hours.

A different group of participants who had not taken part in the first experiment performed the same experiment without rTMS. Besides leaving out all the TMS related procedures, the experimental procedure was exactly the same. The participants were presented with the same amount of trials and experimental blocks using the same counterbalance scheme as used during the rTMS experiment. Only the performance in two experimental blocks was sampled from the results controlling for possible learning processes and other processes involved in the repetition of trials such as attention.

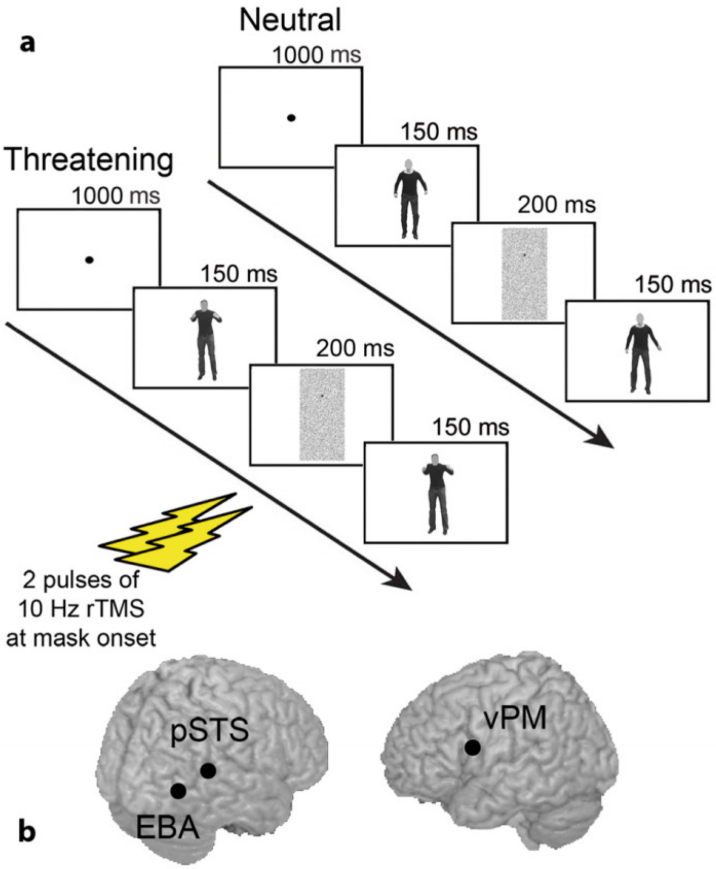


Figure 6.3 *Timeline of the experimental procedure (**a**) and mean coordinates of the stimulation sites (**b**).*

Data handling

Based on the results of the subjective ratings on the perceived postural change in the different and same trials, a trial where a postural change occurred was treated as the signal. The sensitivity to the signal (detection of postural change) was estimated by calculating the d-prime (*d'*). The *d'* is a measure of the distance between the signal and noise distribution means in standard deviation units (Green & Swets, 1966). A *d'* of 0 means that the participant is not able to detect the postural change between the sample and the probe image. D-prime scores were calculated using the formula:

$d' = \Phi^{-1}(H') - \Phi^{-1}(FA')$.

Hits and False Alarms rates were corrected for ceiling effects by applying the formulas:

$H' = (h + 0.5) / (h + m + 1)$,
$FA' = (f + 0.5) / (f + cr + 1)$,

where h is the number of hits, m is the number of misses, f is the number of false alarms and cr is the number of correct rejections (Snodgrass & Corwin, 1988; Tamietto et al., 2007).

Reaction times of the corresponding conditions were analyzed to control for behavioral trade-off effects. Only reaction times of correct responses were considered. Moreover, reaction times that fell below or above three standard deviations from each individual mean were identified for each condition and removed as outliers (0.7 % of the total).

A preliminary analysis was performed on raw rTMS data (reported in Table 6.2). Two separate two-way repeated measure ANOVAs with Valence (threatening / neutral) and Site (EBA / pSTS / vPM) as within-subject factors were performed on body and animal stimuli respectively.

Analysis on raw d' in the human body posture change detection task showed that neither Site (F(2,30) = 0.69, p = 0.51) nor Valence (F(1,15) = 1.11, p = 0.31) reached statistical significance as main effects. Crucially, the interaction between Valence and Site reached statistical significance on detection of human postural change (F(2,30) = 6.79, p = 0.003). The same analysis performed on RTs showed a non significant effect of Valence (F(1,15) = 2.72, p = 0.12), Site (F(2,30) = 0.03, p = 0.97) and their interaction (F(2,30) = 2.24, p = 0.12). The ANOVA performed on d' during the animal body postural change task revealed that detection was more difficult for threatening than neutral posture change (main effect of Valence

(F(1,15) = 5.53, p = 0.03). No effect of Site of stimulation (F(2,30) = 0.68, p = 0.51) or the interaction between Valence and Site (F(2,30) = 1.93, p = 0.16) was found. The ANOVA on RTs of the animal data showed a non significant effect of Valence (F(1,15) = 1.80, p = 0.20), a significant main effect of Site (F(2,30) = 5.47, p = 0.009) which was accounted for by slower reaction times during vPM stimulation with respect to pSTS (p = 0.004) but not EBA (p = 0.06), and a non significant interaction between Valence and Site (F(2,30) = 0.73, p = 0.49).

Human Postures		**EBA**	**pSTS**	**vPM**	**no-rTMS**
d'	**Threatening**	2.19 ± 0.17	2.48 ± 0.17	2.09 ± 0.13	2.31 ± 0.17
	Neutral	2.31 ± 0.13	2.32 ± 0.16	2.63 ± 0.15	2.90 ± 0.24
RTs (ms)	**Threatening**	477 ± 32	493 ± 33	485 ± 34	554 ± 47
	Neutral	482 ± 30	466 ± 29	467 ± 29	515 ± 34

Animal Postures		**EBA**	**pSTS**	**vPM**	**no-rTMS**
d'	**Threatening**	2.52 ± 0.16	2.62 ± 0.19	2.57 ± 0.14	2.81 ± 0.20
	Neutral	2.96 ± 0.13	2.75 ± 0.23	3.07 ± 0.16	3.21 ± 0.17
RTs (ms)	**Threatening**	463 ± 29	436 ± 27	474 ± 30	512 ± 34
	Neutral	442 ± 28	433 ± 25	455 ± 26	494 ± 32

Table 6.2 *Raw d' and RTs (milliseconds) means ± s.e.m. for body (**up**) and animal (**below**) postural change detection in all experimental conditions.*

To eliminate from the analysis any task-specific difficulty difference and to be able to directly compare changes in performance during threatening and neutral detections due to rTMS, the reaction times and *d*' data from the rTMS experiment were transformed to z-scores using the control no-rTMS means and standard deviations values. The same two separated ANOVAs were performed on z-scores. Post-hoc multiple comparisons were carried out using the Newman-Keuls test.

6.3 Results

Human bodies

The 2 Valence (threatening / neutral) * 3 Site (EBA / pSTS / vPM) ANOVA on *d'* values for human body postural change detection revealed a significant interaction between Valence and Site (F(2,30) = 6.70, $p = 0.004$). Crucially, post-hoc tests revealed that pSTS stimulation produced a higher accuracy level during threatening (*d'* z-score = 0.32 ± 0.29 s.e.m.) compared to neutral (*d'* z-score = -0.71 ± 0.19 s.e.m.) body postural change detection ($p < 0.001$) (Figure 6.4). The dissociation between threatening and neutral stimuli was not present during stimulation of either EBA ($p = 0.12$) or vPM ($p = 0.81$). Furthermore, the stimulation of pSTS during threatening body postural change detection induced higher accuracy with respect to EBA (*d'* z-score = -0.19 ± 0.29, $p = 0.02$) and vPM (*d'* z-score = -0.39 ± 0.22, $p = 0.01$) stimulation. By contrast, stimulation of the three sites did not result in any modulation of the accuracy in detecting neutral body postural changes (all *p*s > 0.20). In sum, the sensitivity of the participants to detect changes in threatening body postures resulted to be higher when stimulating pSTS with respect to all other sites and conditions (all *p*s < 0.02). No other comparison was significant (all *p*s > 0.09). Furthermore, when compared to no-rTMS condition, only EBA and pSTS stimulation impaired performance during detection of neutral posture change (corrected t-tests against zero t(15) = -3.76, $p = 0.01$ for pSTS and t(15) = -4.63, $p = 0.002$ for EBA, all other $p > 0.09$).

Although the Valence of the body posture did not significantly affect the d' per se (F(1,15) = 4.27, $p = 0.06$), detection of neutral body postural changes tended to be more difficult with respect to detection of threatening body postural changes independently from the site of stimulation. The factor Site of stimulation did not reach statistical significance as main factor (F(2,30) = 0.91, $p = 0.41$).

The ANOVA on the RTs showed no significant main effect of Valence (F(1.15) = 0.84, $p = 0.37$), Site (F(2,30) = 0.06, $p = 0.94$) or their interaction (F(2,30) = 2.26, $p = 0.12$). No condition resulted different from no-rTMS baseline performance (all corrected *p*s > 0.19).

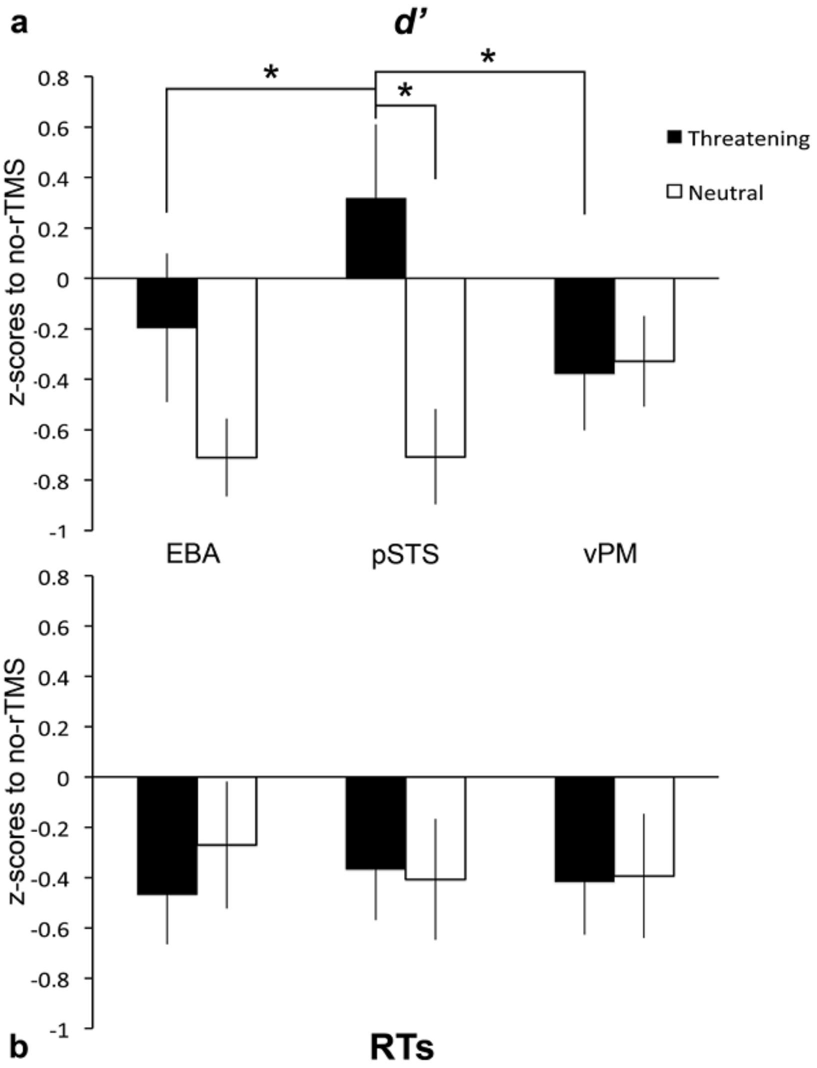


Figure 6.4 *Accuracy (d') in detecting threatening and neutral human body postural changes was modulated in opposite directions only during pSTS stimulation (**a**). No effect on RTs (**b**). Error bars represent s.e.m.*

Animal bodies

The 2 Valence (threatening / neutral) x 3 Site (EBA / pSTS / vPM) ANOVA on *d*' values for animal postural changes showed no significant main effect of Valence (F(1,15) = 0.35, *p* = 0.57), Site (F(2,30) = 0.83, *p* = 0.45) or their interaction (F(2,30) = 2.06, *p* = 0.15) thus indicating that the accuracy of the performance was not affected by rTMS over the different sites of stimulation or by the valence of animal body stimuli (Table 6.3). When compared to baseline performance no condition resulted to be modulated (all *p*s > 0.137).

The ANOVA on RTs showed a main effect of Site of stimulation (F(2,30) = 5.31, *p* = 0.01). Post-hoc tests show that performance during pSTS stimulation was slower only with respect to vPM (*p* = 0.008) regardless of the Valence of the stimuli. Neither the main effect of Valence of stimuli (F(1.15) = 0.01, *p* = 0.91) nor the interaction between Valence and Site of stimulation reached significance (F(2,30) = 0.63, *p* = 0.54) (Table 6.3). When compared to baseline performance no condition resulted to be modulated (all *p*s > 0.09).

Animal Postures (z-scores)		**EBA**	**pSTS**	**vPM**
***d*'**	**Threatening**	-0.41 ± 0.23	-0.27 ± 0.27	-0.34 ± 0.20
	Neutral	-0.42 ± 0.23	-0.78 ± 0.39	-0.24 ± 0.27
RTs	**Threatening**	-0.41 ± 0.24	-0.63 ± 0.23	-0.31 ± 0.25
	Neutral	-0.45 ± 0.25	-0.53 ± 0.22	-0.34 ± 0.23

Table 6.3 *Z-scores of d' and RTs of all experimental conditions during animal postural change detection. Group means (± s.e.m) of d' and RTs of the detection of threatening and neutral animal postural changes during stimulation of all cortical sites.*

6.4 Discussion

The main finding of the present research is that the posterior part of the right Superior Temporal Sulcus (pSTS) plays an opposite causal role in detecting changes in threatening and neutral human body postures. In particular, the ability to detect threatening human postural changes after pSTS stimulation was selectively modulated with respect to stimulation of body sensitive occipito-temporal (EBA) and action sensitive premotor (vPM) regions. These results expand previous functional imaging evidence showing the predominant involvement of pSTS in processing socially relevant bodily movements (Allison et al., 2000; de Gelder, 2006; de Gelder & Partan, 2009; Pinto & Shiffrar, 2009; Puce & Perrett, 2003) by demonstrating that this area has a crucial differential role in detecting emotional and neutral postural changes. Such an effect is probably due to the pSTS direct anatomo-functional connections with the amygdalae (Amaral & Price, 1984; Morris, Friston et al., 1998; Vuilleumier et al., 2004).

The role of STS in threat perception

Neurons in the posterior part of the STS respond to a variety of socially relevant stimuli such as gaze and mouth movements (Puce et al., 1998), facial expressions (Haxby et al., 2000), actions (Decety & Grezes, 1999), biological motion (Puce & Perrett, 2003) and emotional body postures and movements (de Gelder, 2006; de Gelder & Partan, 2009; Grèzes et al., 2007; Kret et al., 2011; Pichon et al., 2008). This area contains cells whose activity is reduced when presented with pairs of successive similar body postures (Perrett, Xiao, Barraclough, Keysers, & Oram, 2009). The quick succession of two similar images seems to induce the perception of apparent motion thus making likely our tasks engaged brain areas involved in real motion perception. However, since in our paradigm no explicit motion was provided by the stimuli, a cautious interpretation of the results is indicated. Here we report that stimulating the pSTS has a differential effect in detecting changes between similar body postures when these convey an emotional or a neutral content. Improved accuracy in detecting threatening postural changes after right pSTS magnetic stimulation with respect to right EBA and left vPM likely occurred because of the selective sensitivity of this area to emotional (socially relevant) body postures. A variety of functional and behavioral studies support the notion that the amygdalae may strengthen the visual processing of emotional stimuli which is thought to be carried out in extrastriate areas (Chouchourelou, Matsuka, Harber, & Shiffrar, 2006; Morris, Friston et al., 1998; Vuilleumier et al., 2004).

MEG and EEG studies in humans reported activation at occipito-temporal sites occurring 178-190 ms after perception of communicative mouth postures (Nishitani & Hari, 2002), body images (Taylor, Roberts, Downing, & Thierry, 2010) and showed that occipito-temporal activation is delayed for implied motion with respect to real motion perception (Lorteije et al., 2006) and for inverted compared to up-right human body presentation (Stekelenburg & de Gelder, 2004). Previous rTMS and patient studies reported the crucial role of pSTS in biological motion perception (Grossman, Battelli, & Pascual-Leone, 2005; Saygin, 2007), in the representation of observed actions on the basis of low-level visual features rather high-order conceptual properties (Cattaneo, Sandrini, & Schwarzbach, 2010) and in the integration between visual (mouth) and auditory (voice) information (McGurk effect) (Beauchamp, Nath, & Pasalar). Moreover, delayed RTs in a gaze-shift task after STS magnetic stimulation has also been reported (Pourtois et al., 2004). Here we demonstrated, for the first time, that pSTS plays an active, crucial role in detecting emotional postural changes early in time after stimulus onset. An alternative explanation for the present results that cannot be excluded is that interference with pSTS facilitates amygdala's reaction to threatening postures, thus facilitating the individual's ability to detect threatening postural changes rather than neutral ones.

Besides being impaired with respect to detecting threatening human postural changes after right pSTS stimulation, detection of neutral human body postural change was also impaired with respect to baseline performance proving that our detection task causally relied on the activity of this area. This evidence is in line with studies reporting superior temporal activations during observation of static body images implying an action (Giese & Poggio, 2003; Peuskens et al., 2005) or when observing rapid succession of images in different postures (Perrett et al., 2009).

EBA and vPM cortex are not selectively involved in detecting changes of emotional body postures

Previous studies showed that the activity of EBA is crucial for local postural change detection 150-250 milliseconds after stimulus presentation (Urgesi, Calvo-Merino et al., 2007). Although stimulation of EBA impaired detection of neutral postural change detection, it did not dissociate between threatening and neutral body postural changes. The involvement of EBA in responding to neutral postural changes has been shown by a previous imaging study (Downing, Peelen, Wiggett, & Tew, 2006). Fusiform cortex is also important for body processing (fusiform body area, Peelen & Downing, 2005) and was first reported to play a role in processing emotional bodies (Hadjikhani & de Gelder, 2003). The reported EBA activity during movement execution showed in an fMRI study (Astafiev et al., 2004) may result from late feedback signals sent from anterior motor brain regions into high-level visual cortices.

The present study shows that vPM cortex is not involved in the visual discrimination between two consecutive body postures, possibly because perception of subtle postural changes depends on higher-order visual areas (Cattaneo et al., 2010; Taylor et al., 2007) rather than the sensorimotor system (Urgesi, Calvo-Merino et al., 2007). A recent TMS study provided compelling functional evidence that, while the activity of STS is involved in the visual description of observed actions (i.e. an action is linked to the body-part that performs it), the activity of the left frontal gyrus is related to the representation of actions' meaning and not to the specific body part used to perform them (Cattaneo et al., 2010) supporting the notion that higher-order action-related representation are supported by premotor regions (Rizzolatti & Craighero, 2004).

6.6 Conclusions

To summarize, the present study significantly expands our understanding of the role of occipital, temporal and premotor cortical regions in the perception of emotional and neutral body motion as seen in postural changes. The main finding is that the right pSTS plays a differential role in the processing of threatening and neutral body postural changes. This dissociation occurs at early stages of stimuli processing and is coherent with studies showing fast subcortical-cortical processing of emotional stimuli. The stimuli used in the present experiment were meant to study the role of the three stimulated areas in detecting threatening and neutral body postural changes. Thus, further research is needed to explore whether the reported effect is specific for the detection of threatening postural changes or if it applies to emotional postural changes in general. The activity of the pSTS is, thus, not only related to socially relevant body-related stimuli processing but it is crucial to effective visual detection of changing socially relevant body postures.

7. Dissociating conscious perception of fearful faces and bodies by transient inhibition of right pSTS

Abstract

Smooth processing of the affective information conveyed by the face and the body is essential for fluent social communication, but the relative importance and the neurocorrelates of the perception of facial and bodily expressions are still poorly understood. We contrasted images of bodily and facial fearful and neutral expressions with houses using a binocular rivalry design and measured the perceptual sensitivity to either stimulus class after inhibition of neural activity in right posterior superior tempral sulcus (r-pSTS) and vertex with offline 1 Hz rTMS. The results showed a clear pattern in which the transient lesion of r-pSTS facilitates the conscious percept of fearful bodies and suppress that of fearful faces while leaving unaffected the perception of their neutral posture. Thus, r-pSTS plays a dissociated role in processing facial and bodily emotional expressions. We propose that the differential adaptive function of perceiving and responding to facial and bodily fearful expressions may be reflected in a dissociation of the link between pSTS and emotion and action related systems. Posterior STS seems to regulate action programming for socially relevant stimuli and possibly plays a major role in behavioral control.

Stienen, B.M.C., Candidi, M., de Gelder, B., Alioti, S.M.A. (in preparation). Dissociating conscious perception of fearful faces and bodies by transient inhibition of right pSTS.

7.1 Introduction

Social interaction depends on a multitude of signals carrying information about the emotional state of others, such as facial and bodily expressions. Category specific brain areas have been identified for faces and bodies in the visual cortex as well as in the fusiform gyrus, which is part of the temporal cortex. In the visual cortex the occipital face area (OFA) and the extrastriate body area (EBA) showed a selective response to faces and bodies respectively (Allison et al., 2000; Downing et al., 2001; Grossman & Blake, 2002; Puce et al., 1996; Spiridon et al., 2006; van de Riet et al., 2009). High resolution imaging revealed spatial segregation of face and body sensitive areas in the fusiform gyrus (FG) called the fusiform face are (FFA) and the fusiform body area (FBA) respectively (Haxby et al., 1994; Kanwisher et al., 1997; Peelen & Downing, 2005; Schwarzlose et al., 2005; Taylor et al., 2007; Tong et al., 1998). One suggestion from this line of work is that a substantial part of processing faces and bodies is separate and devoted to visual stimulus categorization using the ventral processing stream.

Modulation by bodily expressions of the fusiform gyrus was found by Hadjikhani and de Gelder (2003) in addition to modulation by facial expressions (Dolan et al., 2001; Morris, Öhman et al., 1998; Rotshtein et al., 2001). The posterior superior temporal sulcus (pSTS) is activated by both facial and bodily expressions (Allison et al., 2000; LaBar et al., 2003; Pichon et al., 2009; van de Riet et al., 2009). The elevated activation levels of pSTS and FG might be explained by the connections of these areas with the amygdala (Morris, Friston et al., 1998; Rotshtein et al., 2001; Vuilleumier et al., 2004). Consistent with this, in studies with macaque monkeys using anterograde tracers, connections are found between amygdala and all levels of visual cortex, as well as to the temporal lobe (Amaral et al., 2003; Amaral & Price, 1984). However, it remains unclear whether or not the pSTS plays a differential role in processing the facial and bodily expressions.

Oram and Perrett (1996) made the interesting observation that STPa, an area in the macaque brain within the rostral superior temporal cortex, deals with the integration of form and motion. This role is also hinted at by the cortico-cortical connections between STS with ventral and dorsal areas (Ungerleider & Haxby, 1994). Monkey studies show that there are extensive reciprocal connections between STS and inferotemporal cortex (IT) (Sugase et al., 1999; Suzuki & Amaral, 1994). A recent study by Qi et al (submitted) also underscores the central role of pSTS in processing dynamic facial expressions (see also Allison et al., 2000 for a review on the role of STS in visual perception).

Human fMRI and neurophysiologic studies revealed that neurons in the posterior part of the STS respond to a wide variety of socially relevant stimuli such as gaze and mouth movements (Puce et al., 1998), facial expressions (Haxby et al., 2000), actions (Decety & Grezes, 1999), biological motion (Puce & Perrett, 2003) and bodily expressions and movements (Candidi et al., 2011; de Gelder, 2006; de Gelder & Partan, 2009; Grèzes et al., 2007; Kret et al., 2011; Pichon et al., 2008). Therefore, it has been suggested that pSTS is mainly involved in processing the intentions of the observed agent (Allison et al., 2000; Ghazanfar & Santos, 2004). The differential adaptive function of perceiving and responding to facial and bodily emotional expressions may be reflected in a dissociation of the links between pSTS and emotion and action-related networks. Thus, inhibition of temporal cortical regions may be reflected in different perceptual effects according to whether the emotion was conveyed via the face or the body.

In this study we contrasted images of bodily and facial emotional expressions and neutral actions with images of houses in a binocular rivalry (BR) design. The resulting alternation is known to be biased by low-level visual factors such as differences in contrast, brightness, movement and density of contours (Blake & Logothetis, 2002). Given certain visual parameters the two stimuli compete with each other for perceptual dominance rather than resulting in a stable perception that is a fusion of both.

Previous BR studies have shown that high-order properties of the stimuli, such as their meaningfulness, influences the rivalry pattern as well (e.g. Yu & Blake, 1992). Subsequent studies have used BR to investigate dominance between faces expressing different emotions (Alpers & Gerdes, 2007; Yoon et al., 2009) and found that emotional faces dominate over neutral faces. In an fMRI study Tong, Nakayama, Vaughan, & Kanwisher (1998) showed that the fusiform face area (FFA), a category specific brain area for processing faces (Haxby et al., 1994), was activated with the same strength as when the faces were presented in a nonrivalrous condition. Other fMRI studies using BR in which emotional faces were contrasted, showed that suppressed images of fearful faces still activated the amygdalae (Pasley et al., 2004; Williams et al., 2004). When visual signals are prevented from being processed by the cortical mechanisms via the striate cortex, the colliculo-thalamo-amygdala pathway is still able to process the stimulus (de Gelder, Vroomen et al., 1999; Van den Stock, Tamietto et al., in press). This is in line with recent functional magnetic resonance imaging studies that have suggested differential amygdala responses to fear faces as compared to neutral faces when the participants were not aware (Morris, Friston et al., 1998; Whalen et al., 1998).

In order to study the relation between perceiving intentionality transferred by different agents (the whole body or the face) and the function of pSTS in the processing network we inhibited cortical activity in right pSTS through offline 1 Hz rTMS and subsequently measured the perceptual sensitivity to fearful and neutral facial and bodily expressions. If the functionality of pSTS is unrelated to the agent, but related to intentionality in the abstract sense, then one expects that the perceptual dominance of both stimuli classes would be altered to the same extent and in the same direction in respect to vertex stimulation. Alternatively, if pSTS is performing a function that is related to the specific stimulus class dissociation in the modulation of the perceptual dominance would be expected.

7.2 Method

Participants

Seventeen healthy volunteers (16 women, 1 man, *M* age = 23.8 years, *SD* = 2.1) participated in exchange of a monetary reward. All participants were right handed according to the Briggs and Nebes handedness inventory (Briggs & Nebes, 1975) and had normal or corrected-to-normal vision. The experimental procedures were approved by the ethics committee of the Fondazione Santa Lucia (Rome, Italy) and were in accordance with the ethical standards of the 1964 Declaration of Helsinki. None of the participants had neurological, psychiatric, or other medical problems or any contraindication for rTMS (Wassermann, 1998).

Stimuli

Two actors once expressing fear and once combing their hair were selected from a validated photoset as body stimuli (for details see Stienen & de Gelder, 2011). All body pictures had the face covered with an opaque oval patch to prevent that the facial expression would influence the rivalry process. The color of the patch was the average grey value of the emotional and neutral face. The face stimuli of two actors expressing fear and the same actors showing a neutral expression were taken from the McArthur set (http://www.macbrain.org/resources.htm). The house stimuli were taken from a set which is extensively explored in other studies and is known to activate specific brain areas (Aguirre, Zarahn, & D'Esposito, 1998; Epstein & Kanwisher, 1998; Levy, Hasson, Avidan, Hendler, & Malach, 2001; Van den Stock, van de Riet, Righart, & de Gelder, 2008) other than the brain areas activated by bodies and faces, such as EBA and FFA. This makes it a suitable control condition for our purposes. A total of four pictures of bodily expressions (fear and neutral), four pictures of facial expressions (fear and neutral) and two pictures of houses were selected for use in the present study. See Figure 7.1.

To control for contrast and brightness the mean grey value and standard deviation was calculated per image. Subsequently, the overall mean and standard deviation was calculated. Finally, the grey values were adjusted per image to fall within the grey value range of +/- 1 SD from the overall mean of the pictures.

All stimuli were fitted into an area with a white background of 3.00 degrees * 4.83 degrees enclosed by a black frame with a border thickness of .29 degrees. The function of the black frame was to enhance a stable fusion. A white fixation dot was pasted on each of the stimuli. Because we used a method that is comparable with the mirror stereoscope (Blake & Logothetis, 2002), the center of the rivaling stimuli was pasted 11.89 degrees left and right from the center of the screen. This procedure resulted in 8 body-house and 8 face-house displays (4 bodily/facial expressions x 2 houses). The displays were vertically left-right mirrored to control for eye dominance adding up to a total of 16 displays of body-house and face-house pairs. The stimuli were presented on a 19” PC screen with the refresh rate set to 60 Hz. We used Presentation 11.0 to run the experiment.

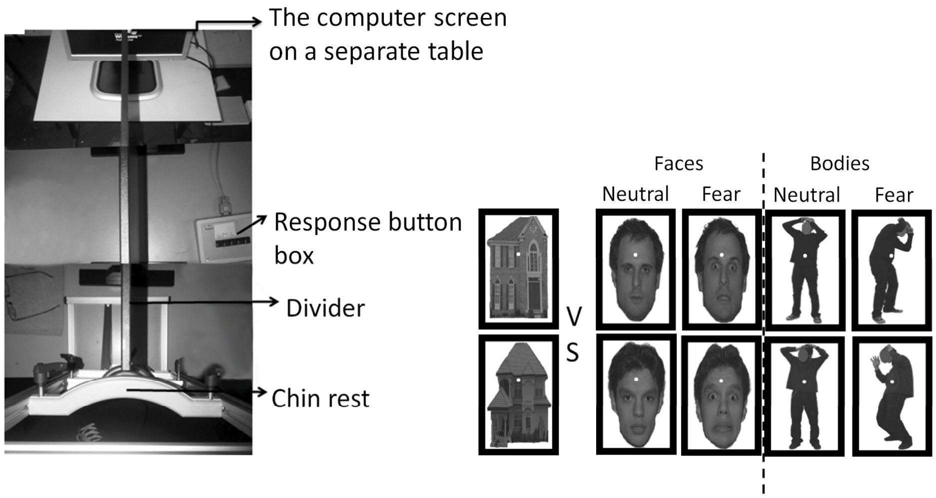


Figure 7.1 *Experimental setup (**left**). The stimulus displays. We always presented bodies or faces in one eye and houses in the other (**right**).*

rTMS and neuronavigation

The OSP for inducing MEPs in the right FDI muscle was found by moving the coil in steps of 1 cm over the left primary motor cortex until the largest MEPs were found. This was marked with a pen on a tightly fitted bathing cap worn by the participants. The rMT was defined as the lowest intensity able to evoke five out of ten MEPs with amplitude of at least 50 μV in the relaxed FDI. This was done by holding the stimulation coil over the optimal scalp position (OSP).

Subsequently, two different locations were identified: one cortical target site, right pSTS, and a control condition, the vertex. Area pSTS was identified on each participant's scalp with SofTaxic Navigator system (EMS). Skull landmarks (nasion, inion, and two preauricular points) and about 60 points providing a uniform representation of the scalp were digitized by means of a Polaris Vicra Optical Tracking System (NDI, Canada). Coordinates in Talairach space were automatically estimated by the SofTaxic Navigator from an MRI-constructed stereotaxic template. The scalp locations that corresponded best to area right pSTS was identified and marked with a pen. Mean (± SD) coordinates (x = 51.5 ± 1.0, y =

−48.2 ± 0.9, z = 8.3 ± 0.7) corresponded to Brodmann area 22. The vertex was identified by finding a point midway between the inion and the nasion and equidistant from the left and right intertragal notches (e.g., Pitcher, Walsh, Yovel, & Duchaine, 2007).

Repetitive TMS was performed via a figure-of-8 coil (Magstim polyhurethane-coated coil) connected to a Magstim Super Rapid Transcranial Magnetic Stimulator (Magstim, Whitland, Dyfed, U.K.). We applied a train of low frequency pulses (1 Hz) for 15 minutes to the vertex and pSTS in two separate experimental sessions separated by 45 minutes. Previous research shows that this procedure disrupts functions related to the targeted area for at least the period of stimulation (Boroojerdi et al., 2000; Grossman et al., 2005; Merabet et al., 2004; Mottaghy, Gangitano, Sparing, Krause, & Pascual-Leone, 2002). Stimulation intensity was 90% of the rMT (mean = 48 % of the maximum stimulator output, SD = 9). During stimulation of pSTS the coil was held tangential to the scalp, with the handle pointing backward and medially at a 45° angle from the middle sagittal plane of the participants' head. The vertex was stimulated with the handle pointing backward horizontally in line with the transverse plane. The position of the coil with respect to the marks was checked continuously during rTMS. During stimulation, participants wore earplugs to protect their hearing and sat comfortably in a chair with their hand relaxed on a pillow in a dimly lit room. None of the participants reported phosphenes, hand muscular twitches during rTMS of pSTS and vertex or showed any other discomfort.

Binocular rivalry task

Because the duration of the rivalry sessions after off-line rTMS had to fall within the time window of 15 min after the stimulation (i.e. when the inhibitory effect of the stimulation is still at play) the complete set of possible combinations between faces/bodies and houses images could not be used within each subject. Thus, half of the participants performed the task based on a different pairing of body/face identity with one of the houses than the remaining participants. For example, the bodily fearful and neutral expressions of identity 1 were paired with house 2 and bodily fearful and neutral expressions of identity 2 were paired with house 1. For the other group this was done vice versa. This procedure was used in order to be sure that the rivalry sessions lasted maximally 15 minutes (in-window). The trials within one run added up to a total of 4 fearful body-house trials, 4 neutral action body-house trials, 4 fearful face-house trials and 4 neutral face-house trials. The bodies and faces were presented in a blocked design, meaning that the body-house and face-house pairs were not randomly mixed within the same run. Trials displaying neutral and fearful expressions were randomized.

Previous to the experimental sessions the participants performed one practice session consisting of two rivalry trials. Other identities then the ones used in the main experiment served as targets. When the participants gave notice of a full understanding of the procedures the main experiment was started. They always participated one complete run previous to any rTMS stimulation block. This session was used to validate using vertex as a control site. Subsequently, right pSTS or vertex was stimulated for 15 minutes. The starting site was counterbalanced across participants and the starting block was alternated between body-house and face-house parings. Nine participants started with body-house pairings and the remaining participants started with face-house pairings. After these two experimental blocks the participants rested for 45 minutes before stimulating the next site to ensure complete recovery from offline rTMS stimulation.

Before each trial two empty frames were shown with a black fixation dot in the middle. The participants were instructed to push and hold a button labeled “Y” with the middle finger to initiate a trial, but only if they saw one dot and one frame. This ensured that the subjects fused the two black frames throughout the experiment. Subsequently, a body- or face-house display was presented for 55 seconds. Whenever they saw a body or a face in isolation they were instructed to release the “Y” button and push and hold the button “X” with their index or ring finger. Whenever they saw a house they were instructed to release the button they were currently pressing and push and hold the button “Z” with their ring or index finger. The participants responded with their right hand and the buttons “X” (Face/Body) and “Z” (Houses) were counterbalanced across participants. When seeing both stimuli they were told to push and hold the button labeled “Y” again.

The software program registered the time each button was pressed and released. The participants were naïve regarding the presentation techniques and during the experiment no reference to the emotions was made.

When performing the rivalry sessions the head of the participants was stabilized using a chin and head rest. The fMRI compatible binocular rivalry method we used is described in detail by Schurger (2009) but was made suitable to use outside of the scanner. A black 70 centimeter wooden divider was placed between the screen and the middle of the eyes. Participants wore glasses in which two wedge formed prism lenses of 6 DVA were fitted using gum. The prisms adjusted the viewing angle from which light from the screen enters each eye ensuring that the laterally presented stimuli would fall close to the participants' fovea. The wooden divider was placed between the eyes to keep the visual signals separated. This method has two major advantages. First, as opposed to the stereoscope or mirror method you can use exactly the same setup in an fMRI experiment because no steel is involved. Secondly, as opposed to other fMRI compatible methods, such as the red-green filter glasses, there is no crosstalk between the eyes (Schurger, 2009). See Figure 7.1 for a picture of the experimental setup.

At the end of all other experimental procedures a short validation session was performed in which all stimuli were presented two times for two seconds to both eyes synchronously. Participants were instructed to categorize the bodies and faces in fearful, angry or neutral bodily or facial expressions using three buttons labeled "F' for fearful, "A" for angry and "N" for neutral.

Data handling

Cumulative viewing times were calculated by adding up the total time participants pressed the button indicating that they saw a face or a body, a house or a combination of the two (a mixed percept). This was done per experimental face or body block.

Two participants were excluded from the analysis. One participant reported having difficulties fusing the black frames. This is reflected in the participant's high report of mixtures (M = 192 s, SD = 16 s) as illustrated in Figure 7.2. The other participant did not report having seen any fearful faces. This is also reflected in the seconds this participant reported having seen a face or a body (M = 7 s, SD = 5 s), see Figure 7.2.

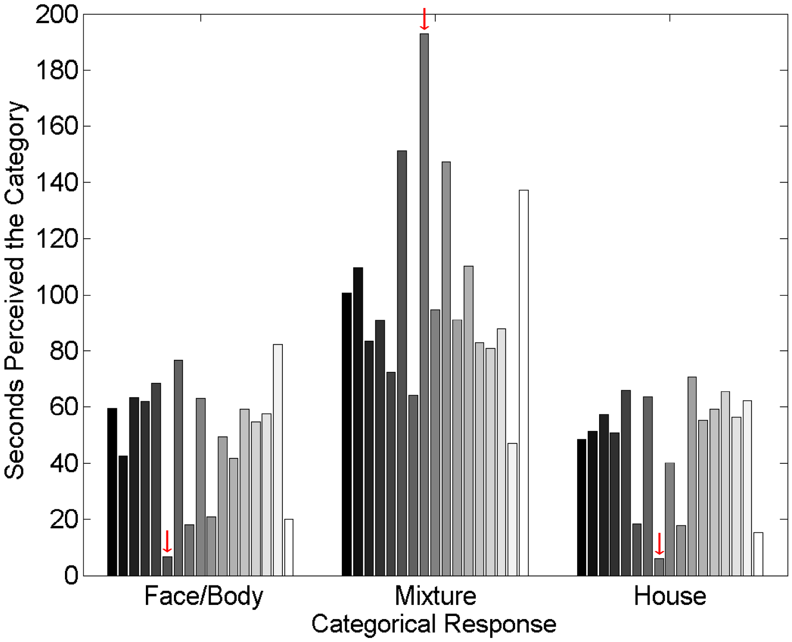


Figure 7.2 *Cumulative viewing times per response category (faces/bodies, response, or houses) per participant. The arrows indicate the participants who either reported difficulties with fusing or did not report seeing any faces when the expression was fearful.*

The main function of the no-TMS condition was to validate the choice for the stimulation of the vertex as a control site to create a reliable baseline. Because of general unspecific effects of rTMS and because the no-TMS condition was not counterbalanced between the two sites the data in the no-TMS condition was treated separately. A two-way repeated measurements ANOVA with stimulus type (face/body) and emotion (fearful/neutral) as within subject factors on the cumulative viewing times of faces/bodies, houses and the mixed percepts showed no main or interaction effects. In Table 7.1 the cumulative viewing times are presented. In addition, in Figure 7.3 the mean cumulative viewing times that the face or the body was reported are illustrated prior to TMS stimulation.

	Neutral Face			Fearful Face		
	Face	House	Mixed	Face	House	Mixed
Mean	59.19	50.24	89.11	60.68	49.27	86.60
SD	17.67	18.00	31.76	20.49	17.90	30.87

	Neutral Body			Fearful Body		
	Body	House	Mixed	Body	House	Mixed
Mean	55.28	57.52	89.00	55.38	52.20	91.21
SD	15.30	16.61	27.02	16.49	13.84	26.15

Table 7.1 *Average cumulative viewing time per visual percept that was reported prior to TMS stimulation when a neutral face (**top left**), a fearful face (**top right**), a neutral body (**bottom left**), and a fearful body (**bottom right**) were contrasted with a house. SD = Standard deviation.*

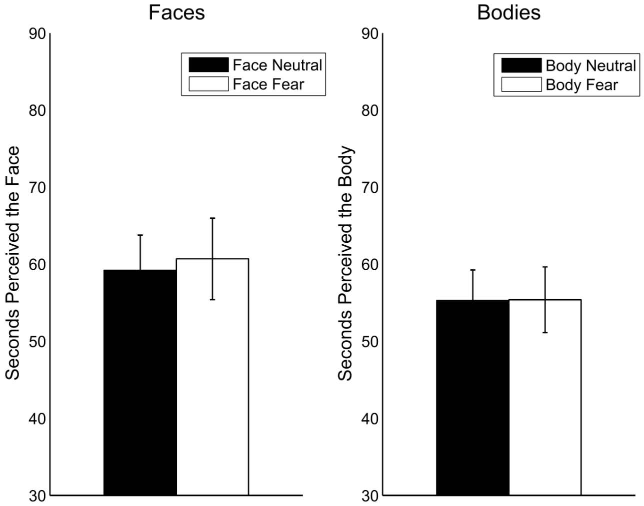


Figure 7.3 *Mean cumulative viewing time for the neutral and fearful face (**left**) and body (**right**) when participants performed the task prior to TMS stimulation. There were no statistical differences in reporting bodies or faces as a function of the emotion expressed. Error bars represent the standard error of the mean.*

7.3 Results

A three-way repeated measurements ANOVA with site (vertex/r-pSTS), stimulus type (face/body) and emotion (fearful/neutral) as within subject factors was performed on the cumulative viewing time of reporting faces/bodies, houses and the mixed percepts. Results showed a three-way interaction effect on the cumulative viewing time of faces and bodies of the three factors (F(1,14) = 10.21, *p* = .006) and a two-way interaction effect on reporting faces and bodies of stimulus type (face/body) and emotion (fearful/neutral) (F(1,14) = 9.78, *p* = .007). Simple main effects analysis showed that faces and bodies did not significantly differ from each other when compared within the emotion category. Although the cumulative viewing times for neutral faces (M = 60.14; SD = 20.31) was almost significantly longer than the cumulative viewing times for neutral bodies (M = 47.40; SD = 19.85), *p* = .050.

To understand the three-way interaction effect the repeated measures ANOVA was done separately on the cumulative viewing times of reporting faces, houses and mixed percepts in the face versus houses rivalry block and the cumulative viewing times of reporting bodies, houses and mixed percepts in the body versus houses rivalry block.

Faces

The repeated measurements ANOVA on the viewing time of reporting faces with site (vertex/r-pSTS) and emotion (fearful/neutral) as within subject factors revealed an interaction between emotion and site, F(1,14) = 5.38, *p* = .036. Duncan's post hoc tests revealed that the cumulative viewing time of reported fearful faces was shorter when pSTS was stimulated in comparison to neutral faces when pSTS was stimulated (*p* = .030), and shorter than the viewing time of neutral and fearful faces when vertex was stimulated (resp. *p* = .041, *p* = .014). See Figure 7.4.

Bodies

The repeated measurements ANOVA on the viewing time of reporting bodies with site (vertex/r-pSTS) and emotion (fearful/neutral) as within subject factors revealed an interaction between emotion and site as well, F(1,14) = 6.57, p = .023. Duncan's post hoc tests showed that the viewing time was longer for fearful bodies when pSTS was stimulated in comparison to neutral bodies (p = .003) when pSTS was stimulated, and longer than the viewing times of neutral and fearful bodies when vertex was stimulated (resp. p = .029, p = .040. See Figure 7.4. Also, a main effect of emotion was revealed by the ANOVA, F(1,14) = 6.86, p = .02. Pairwise comparisons show that the viewing times for fearful bodies were longer than for neutral bodies.

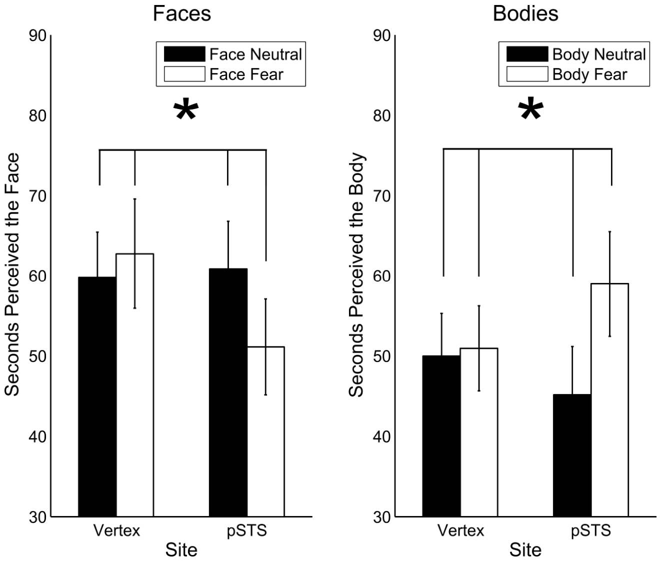


Figure 7.4 *Average viewing time for the neutral and fearful face (**left**) and body (**right**) per stimulation site. Stimulation of pSTS suppresses the percept of fearful faces while it facilitates the percept of fearful bodies. Error bars represent mean standard error, asterisks = p < .05.*

<u>*Houses and Mixtures*</u>

No significant results were revealed by the same repeated measurements ANOVA's on the cumulative viewing time of the houses or the face-house or the body-house mixtures. This indicates that the effects found in the analysis on the viewing time of faces and bodies are not due to inhibiting or facilitatory effects on the perception of houses or mixtures but that those effects are due to changes in the aware perception of the faces and bodies themselves.

To study the effect that 8 participants received different stimulus pairings than the remaining 7 an additional analysis including stimulus pairing as a covariate was performed. This revealed that the interaction effect found on viewing time of faces and bodies was still significant, resp. F(1,13) = 5.908, p = .030 and F(1,13) = 6.375, p = .025 and that stimulus pairing itself did not interact with this or any other term.

7.4 Discussion

Our goal was to test the function of right pSTS in aware perception of facial and bodily expressions by using offline 1 Hz rTMS. The results show a clear pattern in which the transient r-pSTS lesion facilitates the conscious perception of fearful bodies while suppressing that of fearful faces. Thus, r-pSTS seems to play a different role for the processing of facial and bodily expressions. This is in line with our previous study in which we showed that stimulating pSTS leads to better performance in detecting threatening postural changes (Candidi et al., 2011).

There are three possible explanations for these opposite perceptual effects. Fundamental to the three explanations is that the effect is emotion and stimulus specific. Firstly, the differential role of pSTS on the perception of facial and bodily expressions might be explained by its cortico-corical connections with action related areas. The left vPM cortex is thought to be crucial for action representation and understanding, and is generally considered to be a key node of the "mirror neuron" system initially described in monkeys (di Pellegrino et al., 1992), consisting of bimodal visuo-motor cells that fire during real action execution as well as during the passive observation of the same movement (Kilner et al., 2009). The activation of premotor regions is facilitated by the emotional valence of the observed posture (de Gelder et al., 2004; Grèzes et al., 2007; Pichon et al., 2008), strengthening the notion that emotion perception is tightly linked to action programming (Darwin, 1872). Particularly interesting are the cortico-cortical connections between STS with ventral and dorsal areas (Ungerleider & Haxby,

1994) and the connections between STS and subcortical structures as the amygdala as well as between the premotor (PM) cortex and the amygdala (Amaral et al., 2003; Amaral & Price, 1984). In addition, PM receives direct input from STS (Luppino, Calzavara, Rozzi, & Matelli, 2001). There are several studies that have demonstrated the anatomical correspondence between how macaque monkeys and humans process visual objects and faces (Tsao, Freiwald, Knutsen, Mandeville, & Tootell, 2003; Tsao, Moeller, & Freiwald, 2008). Granted that similar connections exist in the human brain, one possibility is that the role of pSTS is differentially activated by emotional faces and bodies because the neural networks involved are separate and are differently specialized.

Secondly, pSTS itself might be differently involved in processing fearful faces and bodies. Van de Riet et al (2009) showed that when a face is expressing fear this leads to higher activation levels of pSTS than when a body is expressing fear, but not when faces and bodies convey a neutral expression. Thus, transient inhibition of right pSTS could lead to a differential effect on the perception of fearful faces and bodies, because pSTS contributes differently to the processing network depending on the stimulus category conveying the emotion.

Thirdly, as reviewed by Allison (2000) and Puce and Perrett (2003), the pSTS is not a homogeneous brain area. It cannot be excluded that a region within pSTS was stimulated being more sensitive to processing bodies than faces that could have lead to a differential effect. However, future research is needed to further unravel the specific functions of the sub regions of the pSTS.

Evidently, the present results trigger a lot of questions. Future research using fMRI should shed more light on these findings in order to formulate an accurate explanation on what mechanism can explain this interesting pattern of results.

8. Summary and conclusions

The goal of this thesis was to study the relation between the recognition of bodily expressions and visual awareness, as well as to explore whether multisensory integration depends on visual awareness of the bodily expressions. As part of the same quest, we explored the neurocorrelates of perception of bodily expressions with and without awareness.

Here, I will summarize and discuss briefly the findings. First, I will discuss the behavioral experiments and then the neurological findings.

8.1 The relation between recognizing bodily expressions and visual awareness

Many research reports have concluded that emotional information can be processed without observers being aware of it. The case for emotional perception without awareness has mostly been made with the use of facial expressions in experimental designs. In view of the similarities between facial and bodily expressions for rapid perception and communication of emotional signals, we conjectured that perception of bodily expressions may also not necessarily require visual awareness.

In chapter 2 we investigated the role of visual awareness in the perception of bodily expressions using a masking technique in combination with confidence ratings on a trial by trial basis. Participants had to detect masked fearful, angry and happy bodily expressions in three separate experiments and subsequently the participants had to indicate their confidence. The onset between target and mask (Stimulus Onset Asynchrony, SOA) varied from -50 to +133 ms. Sensitivity measurements (d-prime) as well as the confidence of the participants showed that the bodies could be detected reliably in all SOA conditions and that there was a high degree of similarity between the masking curves of the different emotions as indicated by the similar shape of the masking curves; recognition performance and confidence ratings was lowest when the SOA was 0 milliseconds for all bodily expressions and highest when the SOA latencies were longer.

Following an interesting study of Lau and Passingham (2006), a lack of covariance was observed between the objective and subjective measurements when the participants had to detect fearful bodily expressions while this was not the case when participants had to detect happy or angry bodily expressions. Therefore, we concluded that the detection of fearful bodily expressions is less dependent on visual awareness than happy and angry bodily expressions. It is important to note that the concerns about masking techniques as described by, for example, Milders et al. (2008) and Wiens (2006) were taken into account. However, it is still possible that happy and angry bodily expressions are differently effected by the mask on a visual level than fearful bodily expressions. Future experiments using different sets of stimuli or e.g. dynamic masks would be useful to explore this issue. In addition, testing the SOA conditions in isolation could shed light on the specific masking effects that could explain our findings. Stimulus specific properties in masking studies are known to modulate the sensitivity of the masking effect. See for a thorough review Wiens (2006).

In the three studies described in chapter 2 all parameters were kept constant except for the emotions expressed by the bodies. This made it possible to compare the difference in covariance between the objective detection rates and the subjectively reported awareness and to compare the differences between experiments.

However, the statistical procedure followed in that chapter is debatable. An alternative analysis, not described in chapter 2, was also explored by first objectively finding the SOA conditions just before the detection performance decreases and after it is back to its optimum again. The SOA points before this falling edge of the curve and after the rising edge was found by taking the d’ value of the most negative SOA condition (-50 ms) and then subtract the d’ value of the subsequent SOA condition. This was done for all the SOAs until arriving at the most positive SOA. These values were tested against zero. After a Bonferroni correction this resulted in the isolation of SOAs -33 and +33 ms for fearful body detection, for angry body detection the SOAs -50 and +50 ms, and for happy body detection the SOAs -50 and +50 ms. This finding already points to a difference between detecting the different emotions. It seems that fearful bodies are less well masked than the other emotions. However, when taking into account the reported confidence, data shows again the interesting pattern that participants were less sure when detecting fearful expressions when the SOA was 33 ms than when it was -33 ms while their detection performance was not different. This was not observed when participants had to detect the other emotions. However, we have chosen to report the statistical procedure in chapter 2 to give insight how the measures covary on similar SOAs across emotions including the ones identified by using the described method.

Social interaction depends not only on the body, but also on the face. However, the relative importance of facial and bodily signals is still poorly understood. To better understand the relative contribution of affective signals from the face only or from the rest of the body, we used a binocular rivalry experiment (chapter 3). For this goal we performed two behavioral experiments. In the first experiment we directly contrasted fearful, angry and neutral bodies and faces. We always simultaneously presented bodies in one eye and faces in the other for 60 seconds and asked participants to report what they perceived. In addition, to create a more sensitive design we repeated the first experiment but this time with only three conditions; one baseline condition in which neutral facial and bodily expressions were contrasted and two experimental conditions in which either the face or the body was expressing fear. Taken together, the results of the two experiments showed that there is no clear bias towards either the face or body, when the expression of the body and face are neutral or angry. The perceptual dominance in favor of either the face of the body seemed to be a function of the stimulus class expressing fear.

The results of chapter 2 and chapter 3 converge to the conclusion that we process fearful expressions of the body and the face differently than other emotional or neutral expressions. Öhman (2002, 2005) suggests that fear stimuli automatically activate fear responses and captures the attention as shown in visual search tasks where participants had to detect spiders, snakes or faces among neutral distracters (Öhman, Flykt et al., 2001; Öhman, Lundqvist et al., 2001). However, the special status of fear stimuli is still a matter of debate, specifically in relation to the role of the amygdale (Duncan & Barrett, 2007; Pessoa, 2005; Pessoa & Adolphs, 2010). More research is necessary using other emotions such as sad, surprise and disgust to confirm the conclusions in this thesis. Also varying the stimulus set would be valuable to study the influence of the specific visual stimulus properties or representativeness of the emotional expression on the rivalry pattern or detection performance when the stimuli are masked.

In chapter 4 we presented emotional auditory information at the onset of masked bodily expressions to investigate whether multisensory integration of affective information occurs independently of visual awareness of the bodily expressions. Multisensory integration may occur independently of visual attention as previously shown with compound face-voice stimuli (de Gelder & Vroomen, 2000). We investigated in two experiments whether the perception of whole body expressions and the perception of voices influenced each other, when observers were not aware of the presentation of the bodily expression. In the first experiment participants categorized masked happy and angry bodily expressions while ignoring congruent or incongruent emotional voices. The onset between target and mask varied from -50 to +133 ms. Results show that the objective recognition of the bodily expressions were facilitated regardless the visibility of the bodily expressions while the confidence of the subjects was not facilitated when the SOA latencies were short and thus the visibility was low. I concluded that the congruency between the emotion in the voice and the bodily expressions influenced audiovisual perception independently of the visibility of the stimuli. In the second experiment participants categorized emotional voices combined with masked bodily expressions as fearful or happy. This experiment showed that bodily expressions presented outside visual awareness still influenced prosody perception. These studies show that audiovisual integration between bodily expressions and affective prosody can take place outside and independent of visual awareness.

However, the results leave us wondering why fear responses increased when the voice was slightly more fearful but the unseen bodily expression was neutral. It may be the case that this is caused by the mismatching of the emotional dimensions of the two sensory signals. The ambiguity that is introduced when the voice is fearful but the visual stimulus is neutral could have confused the participants. The unseen neutral bodily expressions did not, in fact, deliver extra information which could help processing the auditory signal. Alternatively, it might be that although the validation results were good, on an unreportable level the neutral bodily expression might be perceived as being fearful. This is a possibility which suggests further research on this intriguing question, like developing a stimulus set which is not only validated explicitly but also with the use of autonomous responses such as pupil dilation or skin conductance.

Choosing fearful expressions in the second experiment over angry or happy expressions was partly pragmatic. Our study in Chapter 2 showed interesting effects with these stimuli. The particular vocal stimuli were chosen because they were already used elsewhere (de Gelder & Vroomen, 2000) and nicely fitted with our fear stimuli. However, it would be interesting to repeat the same experiment with other emotions as well and study the specific effects of different emotional expressions on the audiovisual integration between bodily expressions and affective prosody and its dependency on visual awareness. For example, the second experiment of chapter 4 could be repeated with masked happy and neutral expressions.

The combination of neurophysiologic methods and behavioral experiments could add more strength to the conclusion that actual integration occurred rather than late decision-based effects (de Gelder & Vroomen, 2000), although the uniformity of the facilitating influence of the congruent auditory affective information across all SOAs including the short SOAs might suggests that such a decision bias is absent, but more research is needed to confirm this.

It is important to note that while the same stimulus set is used in the first experiment described in chapter 2, and the second experiment described in chapter 4, the participants were visually not aware of the bodily expressions in the second experiment of chapter 4, but they were aware of the bodily expressions in chapter 2 when using the same SOA parameter. An important difference between the studies is that in the second experiment of chapter 4 the main task was the categorization of the emotional prosody in the spoken sentences, and to withhold that response when the centrally presented cross turned 45 degrees clockwise. In other words, attention was not focused on the bodily expressions while this was the case in the studies described in chapter 2, where the participants were instructed to detect the bodily expressions and to indicate their confidence.

The difference between the accuracy rates when the SOA was 33 ms in the behavioral masking experiment in chapter 2 and chapter 5 has a different cause. In chapter 5 we used a completely different set of stimuli than in chapter 2 and 4. The biggest difference between these sets is the size of the set. While in the experiments in chapter 2 and the first experiment in chapter 4, the trials containing the same stimuli are repeated 56 times, whereas the study described in chapter 5 was only repeated 6 times. This can explain the difference in accuracy rates between chapter 1 and 5 indicating that learning processes might be important in the relation between visual awareness and the detection of (emotional) stimuli.

8.2 The neurocorrelates of processing bodily expressions

The behavioral experiments mentioned in the previous section suggest that fearful bodily expressions can be processed, independent of visual awareness and that the awareness of a bodily expression is a function of the expression of fear. This is in line with previous studies showing that full awareness of the visual stimulus or intact visual cortex is not essential for processing facial and bodily expressions (de Gelder, Vroomen et al., 1999; Tamietto et al., 2009; Tamietto & de Gelder, 2010) and with fMRI studies using binocular rivalry which showed that suppressed images of fearful faces still activated the amygdalae (Pasley et al., 2004; Williams et al., 2004). Whether this is the case in respect to the behavioral experiments reported in chapter 2-4 remains unclear and needs further research using neurological or neurophysiologic measures.

However, in chapter 5-7 studies are described that can shed more light on the cortical and subcortical processing of bodily expressions with and without awareness. Chapter 5 studies the possibility of recognizing bodily expressions using only cortical feed-forward processing. Given the presence of massive feedback loops in brain networks, it is difficult to disentangle the contribution of feedforward and feedback processing on the recognition of visual stimuli, in this case, of emotional body expressions. By means of parametric masking it was possible to control the contribution of feedback activity in human participants. A close comparison was possible between their recognition performance of masked bodily expressions and the performance of a computational neural model which exclusively modeled feed-forward processing and was engineered to fulfill the computational requirements of categorization. Results showed that the longer the SOA (Stimulus Onset Asynchrony) the closer the performance of the human subjects was to the predicted values by the model with an optimum when the SOA was 100 ms. On short SOAs, however, the human performance deteriorated, but the categorization of the emotional expressions was still above baseline. We suggested that although theoretically it is likely that feedback arising from inferotemporal cortex was blocked when the SOA was 100 ms, human participants seem to rely on more local visual feedback processing to equal the model's performance.

That the performance of humans does not match the model's performance when the SOA latency is short, could be hinting at an active subcortico-cortical pathway as the model is not able to predict this performance. As explained in section 1.2, when visual signals are prevented from being processed by the cortical mechanisms, the visual emotional stimulus could still be processed via a colliculo-thalamo-amygdala pathway. However, this study lacks the additional measurement of e.g. subjective awareness to be conclusive on this topic (see e.g. Cheesman & Merikle, 1986).

In chapter 6 we explored how brain areas associated with processing bodily postures contribute to the automatic perception of bodily expressions. While imaging studies indicate that perceiving body emotions relies upon a wide network of subcortico and cortical neural regions, little is known on the causative role of different nodes of this network. We applied event-related rTMS over the body- and action-related extrastriate (EBA), temporal (pSTS) and premotor (vPM) cortices to test their active contribution in perceiving changes between two successive images of either threatening or neutral human body or animal postures. While stimulation of EBA and vPM showed no selective effect on threatening stimuli, with respect to neutral ones, rTMS over EBA and pSTS selectively impaired neutral posture detection and stimulation of pSTS facilitated the detection of changes in threatening human postures with respect to all other experimental conditions. No such effect was found for animal stimuli. Here we demonstrated, for the first time, that pSTS plays an active, crucial role in detecting emotional postural changes early in time after stimulus onset.

An alternative explanation for the present results that cannot be excluded, is that interference with pSTS facilitates activation of the amygdala to threatening postures, thus facilitating the individual's ability to detect threatening postural changes rather than neutral ones. Although it must be noted that it has been shown that monkeys' premotor cortex and extrastriate cortex receives neural projections from the amygdaloid complex as well (Amaral et al., 2003; Amaral & Price, 1984; Avendano et al., 1983).

This function of pSTS in processing expressions was further examined by using TMS in combination of a binocular rivalry design to study the role of pSTS in visual awareness of facial and bodily expressions. Smooth perception of the information conveyed by the body and face is essential for fluent social communication, but the relative importance and the neurocorrelates of facial and bodily expressions are still poorly understood. In the study described in chapter 7 images of bodily and facial expressions are contrasted with houses using a binocular rivalry design. The perceptual sensitivity to either stimulus class was measured after inhibition of neural activity in right posterior superior temporal sulcus (r-pSTS) and vertex with offline 1 Hz rTMS. The results showed a clear pattern in which the transient lesion of r-pSTS facilitates the conscious perception of fearful bodies and suppresses that of fearful faces, while leaving unaffected the perception of their neutral posture. Thus, r-pSTS plays a dissociated role in processing facial and bodily emotional expressions. We propose that the differential adaptive function of perceiving and responding to facial and bodily fearful expressions may be reflected in a dissociation of the link between pSTS and emotion and action related systems. Posterior STS seems to regulate action programming for socially relevant stimuli and possibly plays a major role in behavioral control.

Chapter 6 and 7 suggest that pSTS regulates action programming for socially relevant stimuli and possibly plays a major role in behavioral control. Oram and Perrett (1996) made the interesting observation that STPa, an area in the macaque brain within the rostral superior temporal cortex, deals with the integration of form and motion. This role is also hinted at by the cortico-cortical connections between STS with ventral and dorsal areas (Ungerleider & Haxby, 1994). In addition, studies with macaque monkeys using anterograde tracers, connections are found between amygdala and all levels of visual cortex, as well as to the temporal lobe (Amaral et al., 2003; Amaral & Price, 1984). This is in line with human neuroimaging studies showing activation of pSTS by both facial and bodily expressions (Allison et al., 2000; LaBar et al., 2003; Pichon et al., 2009; van de Riet et al., 2009). The presentation of static bodily expressions in the blind field of a hemianopic patient showed not only activity in the pulvinar but also in cortical regions like MT, premotor cortex, and in STS (de Gelder & Hadjikhani, 2006). However, the contribution of the subcortical pathway and the cortical pathway in the observed results remains unknown. Future research using, for example, our binocular rivalry design, in a neuroimaging study could shed more light on this issue.

It is important to note the difference between the experimental setup of chapter 3 and chapter 7. In chapter 3 bodies and faces are contrasted directly with each other, while in chapter 7 a common neutral stimulus, a house, is used to contrast with the faces and bodies separately. In chapter 3 one of the questions was whether humans have a differential processing sensitivity to both categories and how emotion modulates this sensitivity. There are low-level factors playing a role here, but surprisingly, in the first experiment we did not find any differences in processing sensitivity to either neutral bodily or facial expressions. Something you would expect if you assume that low-level differences between bodies and faces are important factors. However, the pattern changed as a function of the emotion expressed by the body or the face. In chapter 7 the main goal was to test the functionality of pSTS in processing the emotion of the face and the body. To achieve this goal, a common neutral image, a house, was used to keep the visual stimulation the same within the stimulus category, but to be able to manipulate the displayed emotion.

8.3 General conclusions

This thesis revealed that fearful bodily expressions, similar as fearful facial expressions, can be recognized and processed independent of visual awareness, and affective auditory information can influence the recognition of bodily expressions independent of visual awareness. Moreover, unseen bodily expressions can influence the categorization of prosody in the voice. In addition it is shown that the perceptual dominance in favor of either the face or the body is a function of the stimulus class expressing fear.

In addition to the available literature that has already firmly established that emotional bodily expressions clearly and rapidly convey the emotional, intentional and mental state of a person (Meeren et al., 2005; Stekelenburg & de Gelder, 2004), the categorization of bodily expressions seems to rely on feed-forward processing of the stimulus although more local feedback processing in the visual cortex still could be still necessary.

It is shown that posterior STS plays an important role in the processing of emotional bodily expressions. When transiently lesioned, pSTS facilitated the conscious percept of fearful bodies while suppressing that of fearful faces. In addition, when we detect postural changes in our environment, it seems lack of input from pSTS could enhance our sensitivity. Posterior STS seems to regulate action programming for socially relevant stimuli and possibly plays a major role in behavioral control.

Nederlandse samenvatting

Het verwerken van lichaams- en gezichtsexpressies met en zonder visueel bewustzijn

Sociale communicatie is afhankelijk van het verwerken van een verscheidenheid aan informatie. Deze informatie reikt van lichaams- en gezichtsexpressies tot de emotionele kleur van de stem, een ruwe aanraking, of zelfs de geur van iemand.

Er zijn veel onderzoeksartikelen waarin is geconcludeerd dat emotionele informatie verwerkt kan worden zonder dat mensen zich daar visueel bewust van hoeven te zijn. Het ontbreken van visueel bewustzijn wil hier zeggen dat je niet kan rapporteren wat je eigenlijk wel hebt opgepikt via de retina. Deze emotionele informatieverwerking zonder visueel bewustzijn is vooral aangetoond door foto's van gezichtsexpressies te tonen en de zichtbaarheid ervan te manipuleren in talrijke verschillende experimenten. De beschikbare literatuur heeft al aangetoond dat emotionele lichaamsexpressies ook duidelijk en efficiënt de emotionele, intentionele en mentale staat van een persoon kunnen overbrengen (Meeren et al., 2005; Stekelenburg & de Gelder, 2004) en ook dat visueel bewustzijn niet noodzakelijk is voor deze perceptie (de Gelder, Vroomen et al., 1999; Tamietto et al., 2009; Tamietto & de Gelder, 2010).

Tot nu toe is deze verwerking van emotionele lichaamsexpressies zonder visueel bewustzijn alleen nog maar aangetoond bij een speciale patiëntengroep, namelijk hemianope patiënten. Dit zijn patiënten die door een bepaalde medische oorzaak een deel van hun visuele cortex missen waardoor zij bijvoorbeeld in hun rechter visuele veld klinisch blind zijn. Deze patiënten kunnen emoties van gezichts- en lichaamsexpressies boven kans correct classificeren terwijl zij aangeven dat zij het antwoord gokken. In de literatuur zijn verschillende mechanismen voorgesteld die dit fenomeen verklaren. In dit proefschrift ga ik vooral uit van een evolutionair oud subcorticaal mechanisme dat bestaat uit de superior colliculus, de pulvinar en de amygdala die mogelijkerwijs emotionele visuele stimuli verwerkt ook wanneer de toegang tot de visuele cortex wordt geblokkeerd. Dit lijkt namelijk het geval te zijn bij hemianope patiënten. Of dit ook het geval is bij het verklaren van de fenomenen geobserveerd in mijn experimenten moet nog blijken uit toekomstig onderzoek.

In het tweede hoofdstuk van mijn thesis beschrijf ik een studie waarin wij foto's van emotionele lichaamstaal gemaskeerd presenteren en in elke trial vragen om de emotie te detecteren en aan te geven hoe zeker de proefpersonen zijn van hun antwoord. In drie verschillende experimenten liet ik hen bange, boze en blije lichaamsexpressies detecteren. De tijd tussen de presentatie van de foto van de emotionele lichaamsexpressie en het masker (SOA) werd parametrisch gevarieerd tussen de -50 tot +133 ms. Zowel een gevoelige detectie maat (de d-prime) als de zekerheid van de proefpersonen lieten zien dat zij de emotionele lichaamsexpressies betrouwbaar konden detecteren ongeacht de SOA. Een belangrijk verschil tussen de detectie van de emoties was dat wanneer de proefpersonen bange lichaamsexpressies moesten detecteren de objectieve detectiemaat en de zekerheidsindex minder covarieerde dan wanneer zij de andere emoties moesten detecteren.

Deze special relatie tussen bange expressies en bewustzijn was al aangetoond voor gezichten (bijvoorbeeld, Esteves, Dimberg et al., 1994; Esteves & Öhman, 1993; Morris et al., 1996; Öhman, Lundqvist et al., 2001). Om de relatie tussen de niet-bewuste verwerking van gezichts- en lichaamsexpressies beter te begrijpen hebben wij een studie uitgevoerd waarin we twee binocular rivalry experimenten uitvoerden. Binocular rivalry is een methode waarin twee beelden in de twee ogen apart worden aangeboden. Afhankelijk van bepaalde eigenschappen zal er conflict tussen de beide informatiestromen plaatsvinden waardoor je afwisselend de ene of de andere stimulus zal waarnemen. Het gebruik van deze methode staat beschreven in hoofdstuk 3. De methode is geschikt om twee stimulus categorieën met elkaar te contrasteren om de gevoeligheid van ons visueel systeem te testen voor deze stimuli. In het eerste experiment contrasteerden wij bange, boze en neutrale lichamen en gezichten direct met elkaar. We presenteerden altijd een lichaam aan het ene oog en een gezicht aan het andere oog 60 seconden lang en vroegen de proefpersonen om te rapporteren wat ze zagen. In het tweede experiment isoleerden wij de bange lichaamsexpressies om dit beter te kunnen bestuderen. Samengenomen lieten de twee experimenten een duidelijk patroon zien waarin er geen duidelijke voorkeur bestond voor zien van gezichten en lichamen, maar waarin de dominantie van de stimulus werd bepaald door de bange expressie.

In hoofdstuk 4 beschrijf ik twee studies waarin wordt bestudeerd of lichaamsexpressies en emotioneel gesproken zinnen of uitroepingen elkaar beïnvloeden zonder of onafhankelijk van visueel bewustzijn. We onderzochten dit in twee experimenten. In het eerste experiment moesten proefpersonen gemaskeerde blije en boze lichaamexpressies categoriseren terwijl zij emotionele stemmen moesten negeren. De SOA werd weer parametrisch gevarieerd en de proefpersonen moesten zowel de emotie categoriseren als aangeven of zij zeker van hun antwoorden waren. De resultaten lieten zien dat de mate waarin overeenstemming was tussen de emotie van de stem en de lichaamsexpressie invloed had op de herkenning van de lichaamsexpressie, maar dat dit onafhankelijk was van de zichtbaarheid van de lichaamsexpressie. In het tweede experiment moesten de proefpersonen juist de emotie in een stem categoriseren terwijl bange en neutrale lichaamsexpressies gemaskeerd werden aangeboden. Er werd in dit tweede experiment zorgvuldig nagegaan of de proefpersonen de lichaamsexpressies hadden gezien, wat niet de bedoeling was. De resultaten lieten zien dat hoewel de lichaamsexpressies niet gezien waren de perceptie van de emotie in de stem toch werd beïnvloed door deze lichaamsexpressies.

Het brein verwerkt visuele emoties op meerdere manieren. Door de aanwezigheid van vele *feedback* verbindingen in het brein die informatie terugsturen van hersengebieden die hoger in de verwerkingshiërarchie liggen naar verschillende lagere hersengebieden is het moeilijk om de bijdrage van de *feedforward* en *feedback* verbindingen in kaart te brengen. Het laatste mechanisme wordt vooral in verband gebracht met bewuste verwerking van stimuli (Lamme & Roelfsema, 2000). Om meer te weten te komen over welke breinmechanismen wij gebruiken om de gemaskeerde lichaamsexpressies te verwerken werden de resultaten van een experiment waarbij proefpersonen gemaskeerde lichaamsexpressies moesten categoriseren vergeleken met de categorisatie prestaties van lichaamsexpressies door een computationeel neuraal model dat uitsluitend feedforward verbindingen modelleert. Dit experiment is beschreven in hoofdstuk 5. Het bleek dat wanneer de SOA langer was de resultaten beter overeenkwamen met het model. Het optimum lag wanneer deze waarde 100 ms was. Op kortere latenties bleek dat de categorisatie door de proefpersonen verslechterde, maar nog steeds boven kans waren. De data suggereert dat hoewel het theoretisch mogelijk is dat mensen geen feedback nodig hebben vanuit hoger gelegen hersengebieden die een rol spelen bij objectherkenning, nog wel feedback nodig is uit andere gebieden om net zo goed te presteren als het model dat zuiver feedforward verbindingen modelleert.

Studies die gebruik maken van EEG, fMRI en TMS laten zien dat de perceptie van lichaamsexpressies afhankelijk is van een wijd gedistribueerd subcortico-corticaal neuraal netwerk. Om te testen welke hersengebieden noodzakelijk zijn voor het verwerken van lichaamsexpressies gebruikten wij TMS om activiteit te beïnvloeden in lichaams- en actiegerelateerde extrastriate (EBA), temporale (pSTS) en premotorische (vPM) hersengebieden terwijl proefpersonen kleine lichaamsveranderingen in menselijk of dierlijk postuur moesten detecteren die bedreigend of neutraal konden zijn. De resultaten ondersteunen de notie dat pSTS betrokken is bij de detectie van sociaal relevante informatie. Dit zou kunnen wijzen op amygdalo-temporo-corticaal modulerende verbindingen. Deze studie staat beschreven in hoofdstuk 6.

In hoofdstuk 7 beschrijf ik een experiment waarin wij deze bevindingen verder hebben getest door net als in hoofdstuk 3 binocular rivalry toe te passen. Dit keer contrasteerden wij foto's van bange en neutrale lichaams- en gezichtsexpressies met huizen nadat wij doormiddel van offline 1 Hz rTMS de activiteit in de rechter pSTS of de vertex hadden geïnhibeerd. Het resultaat was dat deze virtuele laesie het bewuste percept van bange lichaamsexpressies faciliteerde en dat van bange gezichtsexpressies onderdrukte, terwijl het geen effect had op de neutrale expressies. Oftewel, pSTS maakt onderscheidt of het emotionele signaal overgebracht wordt door het gezicht dan wel het lichaam. Wij stellen voor dat deze differentiële functie voortkomt uit een verschil hoe de pSTS zich in het netwerk gedraagt wanneer verschillende emotie en actiegerelateerde netwerken worden aangesproken. Posterieure STS lijkt belangrijk te zijn in het reguleren van acties voor sociaal relevante stimuli en speelt mogelijk een belangrijke rol in controle over het gedrag.

Dit proefschrift beschrijft dat bange lichaamsexpressies, net zoals bange gezichtsexpressies herkent en verwerkt kunnen worden zonder en onafhankelijk van visueel bewustzijn. Ook kan auditieve emotionele informatie de herkenning van lichaamsexpressies beïnvloeden onafhankelijk van visueel bewustzijn en kunnen op hun beurt niet geziene lichaamsexpressies de herkenning van emoties in de stem beïnvloeden. Daarnaast is er aangetoond dat wanneer lichamen en gezichten tegelijk worden gepresenteerd wij bewust worden van het gezicht of het lichaam dat angst uitdrukt.

Ook is er in dit proefschrift een experiment behandeld waarin de rol werd onderzocht van feedforward en feedback verbindingen in de hersenen in het verwerken van lichaamexpressies wanneer de zichtbaarheid belemmerd wordt. Het blijkt dat mensen toch nog lokale feedback verbindingen nodig hebben om net zo goed te kunnen presteren als een computationeel neuraal model dat alleen feedforward verbindingen modelleert.

Tot slot is er aangetoond doormiddel van TMS studies dat pSTS een belangrijke rol speelt in het verwerken van emotionele lichaamsexpressies. Wanneer activiteit in dit hersengebied werd beïnvloed werden de proefpersonen zich juist bewuster van bange lichaamsexpressies en waren zij beter in het detecteren van kleine veranderingen in het postuur wanneer het een bedreigende houding toonde. Dit terwijl het bewuste percept van bange gezichtexpressies juist werd onderdrukt.

Acknowledgments

Here, I will take a few lines to thank some of the people without whom this doctoral thesis would never have reached the state in which it is worthy to be published, as it is now. Of course, as most scientists will agree with me, earning your PhD is much more than writing this little book, which is merely a selection of some of the work that I have been doing over the past 4 years.

The first year I spent a lot of my time wondering: what is science and what is the best way to practice it? As I matured more in being a scientist just simply by practicing it, especially my supervisor, Beatrice de Gelder, I owe my sincere gratitude. She taught me that you can identify a good experiment by the quality of the questions that it triggers. Simply stated, a good experiment should coin new and interesting questions generating new ways to study phenomena. Especially this attitude opened a lot of doors for me. I traveled to Rome to perform three TMS experiment, I traveled each month up and down to Paris to perform our masking studies with amygdalectomy patients, and I collaborated with excellent scientists from Switzerland and Japan. Bea encourages a dynamic approach of practicing science and taught me to value your colleagues, to interact with them, and to perform experiments in a high pace. Thank you Bea, although sometimes the bar was set very high, you made my experience of practicing science much more intense and dynamic.

I started my PhD course approximately around the same time with Charlotte Sinke and Mariska Kret. Thank you both for the heartwarming and encouraging talks whenever it was needed. I also thank Geert van Boxtel for challenging me to get into the details of the equipment and software I used. I also thank the committee for their valuable comments when reviewing my thesis. In addition, I would like to thank Jeroen Stekelenburg, Hanneke Meeren, Mirjam Keetels, Jan Van den Stock, Marco Tamietto, and Wobbe Zijlstra, and many others for the feedback on experimental or statistical issues.

My father passed away in the second year. There are no words to express my sadness. This affected my scientific progress, as well as my view on life itself. My father always motivated me to continue studying and supported me wherever possible. Without the comfort of my family and friends, I surely would have been squashed by emotions and would have dropped out of the PhD course. I am sure you know who you are when I say: thanks a million!

Special thanks to my paranimfen and very good friends Michiel en Jolle. I hope you did a good job being a paranimf! And last but not least, I thank myself for not losing my faith and working harder than I ever could imagine was possible.

It is impossible for me to name (and remember) everybody who directly or indirectly helped me reaching my goal, but I am sure when I meet you in person, I will tell you directly.

References

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. *Neuron, 21*(2), 373-383.

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in Cognitive Science, 4*(7), 267-278.

Alpers, G. W., & Gerdes, A. B. (2007). Here is looking at you: emotional faces predominate in binocular rivalry. *Emotion, 7*(3), 495-506.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology, 15*(9), 839-843.

Amaral, D. G., Behniea, H., & Kelly, J. L. (2003). Topographic organization of projections from the amygdala to the visual cortex in the macaque monkey. *Neuroscience, 118*(4), 1099-1120.

Amaral, D. G., & Price, J. L. (1984). Amygdalo-cortical projections in the monkey (Macaca fascicularis). *The Journal of Comparative Neurology, 230*(4), 465-496.

Avendano, C., Price, J. L., & Amaral, D. G. (1983). Evidence for an amygdaloid projection to premotor cortex but not to motor cortex in the monkey. *Brain Research, 264*(1), 111-117.

Bannerman, R. L., Milders, M., de Gelder, B., & Sahraie, A. (2008). Influence of emotional facial expressions on binocular rivalry. *Ophthalmic & Physiological Optics, 28*(4), 317-326.

Barrett, L. F., Ochsner, K. N., & Gros, J. J. (2007). On the automaticity of emotion. In J. Bargh (Ed.), *Social psychology and the unconscious: the automaticity of higher mental processes.* New York: Psychology Press.

Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *Journal of Neuroscience, 30*(7), 2414-2417.

Berry, J. W. (1969). Ecology and socialization as factors in figural assimilation and the resolution of binocular rivalry. *International Journal of Psychology, 54*, 331–334.

Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nature Reviews Neuroscience, 3*(1), 13-21.

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

Boroojerdi, B., Bushara, K. O., Corwell, B., Immisch, I., Battaglia, F., Muellbacher, W., et al. (2000). Enhanced excitability of the human visual cortex induced by short-term light deprivation. *Cerebral Cortex, 10*(5), 529-534.

Breitmeyer, B. G., & Ganz, L. (1976). Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychological Review, 83*(1), 1-36.

Briggs, G. G., & Nebes, R. D. (1975). Patterns of hand preference in a student population. *Cortex, 11*(3), 230-238.

Calvo, M. G., & Nummenmaa, L. (2008). Detection of emotional faces: salient physical features guide effective visual search. *Journal of Experimental Psychology: General, 137*(3), 471-494.

Candidi, M., Stienen, B. M. C., Aglioti, S. M., & de Gelder, B. (2011). Event related repetitive transcranial magnetic stimulation of posterior superior temporal sulcus improves the detection of threatening postural changes in human bodies. *Journal of Neuroscience, 31*(48), 17547-17555.

Candidi, M., Urgesi, C., Ionta, S., & Aglioti, S. M. (2008). Virtual lesion of ventral premotor cortex impairs visual perception of biomechanically possible but not impossible actions. *Social Neuroscience, 3*(3-4), 388-400.

Cattaneo, L., Sandrini, M., & Schwarzbach, J. (2010). State-dependent TMS reveals a hierarchical representation of observed acts in the temporal, parietal, and premotor cortices. *Cerebral Cortex, 20*(9), 2252-2258.

Cheesman, J., & Merikle, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics, 36*(4), 387-395.

Cheesman, J., & Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology, 40*(4), 343-367.

Chouchourelou, A., Matsuka, T., Harber, K., & Shiffrar, M. (2006). The visual analysis of emotional actions. *Social Neuroscience, 1*(1), 63-74.

Cooke, D. F., & Graziano, M. S. (2004). Sensorimotor integration in the precentral gyrus: polysensory neurons and defensive movements. *Journal of Neurophysiology, 91*(4), 1648-1660.

Darwin, C. (1872/1965). *The Expression of the Emotions in Man and Animals*. Chicago, USA: University of Chicago.

Day-Brown, J. D., Wei, H., Chomsung, R. D., Petry, H. M., & Bickford, M. E. (2010). Pulvinar projections to the striatum and amygdala in the tree shrew. *Frontiers in Neuroanatomy, 4*, 143.

de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience, 7*(3), 242-249.

de Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 364*(1535), 3475-3484.

de Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences, 7*(10), 460-467.

de Gelder, B., Böcker, K. B., Tuomainen, J., Hensen, M., & Vroomen, J. (1999). The combined perception of emotion from voice and face: early interaction revealed by human electric brain responses. *Neuroscience Letters, 260*(2), 133-136.

de Gelder, B., & Hadjikhani, N. (2006). Non-conscious recognition of emotional body language. *Neuroreport, 17*(6), 583-586.

de Gelder, B., & Partan, S. (2009). The neural basis of perceiving emotional bodily expressions in monkeys. *Neuroreport, 20*(7), 642-646.

de Gelder, B., Pourtois, G., & Weiskrantz, L. (2002). Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. *Proceedings of the National Academy of Sciences of the United States of America, 99*(6), 4121-4126.

de Gelder, B., Snyder, J., Greve, D., Gerard, G., & Hadjikhani, N. (2004). Fear fosters flight: A mechanism for fear contagion when perceiving emotion expressed by a whole body. *Proceedings of the National Academy of Sciences of the United States of America, 101*(47), 16701-16706.

de Gelder, B., van Honk, J., & Tamietto, M. (2011). Emotion in the brain: of low roads, high roads and roads less travelled. *Nature Reviews Neuroscience, 12*(7), 425; author reply 425.

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition & Emotion, 14*(3), 289-311.

de Gelder, B., Vroomen, J., Pourtois, G., & Weiskrantz, L. (1999). Non-conscious recognition of affect in the absence of striate cortex. *Neuroreport, 10*(18), 3759-3763.

Decety, J., & Grezes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Science, 3*(5), 172-178.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research, 91*, 176-180.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science, 11*(1), 86-89.

Dolan, R. J., Morris, J. S., & de Gelder, B. (2001). Crossmodal binding of fear in voice and face. *Proceedings of the National Academy of Sciences of the United States of America, 98*(17), 10006-10010.

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.

Downing, P. E., Peelen, M. V., Wiggett, A. J., & Tew, B. D. (2006). The role of the extrastriate body area in action perception. *Social Neuroscience, 1*(1), 52-62.

Duncan, S., & Barrett, L. F. (2007). The role of the amygdala in visual awareness. *Trends in Cognitive Sciences, 11*(5), 190-192.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences, 4*(9), 345-352.

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

Esteves, F., Dimberg, U., & Öhman, A. (1994). Automatically elicited fear: conditioned skin conductance responses to masked facial expressions. *Cognition & Emotion, 9*, 99-108.

Esteves, F., & Öhman, A. (1993). Masking the face: recognition of emotional facial expressions as a function of the parameters of backward masking. *Scandinavian Journal of Psychology, 34*(1), 1-18.

Esteves, F., Parra, C., Dimberg, U., & Öhman, A. (1994). Nonconscious associative learning: Pavlovian conditioning of skin conductance responses to masked fear-relevant facial stimuli. *Psychophysiology, 31*(4), 375-385.

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience, 13*(2), 171-180.

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. (2007). Masking disrupts reentrant processing in human visual cortex. *Journal of Cognitive Neuroscience, 19*(9), 1488-1497.

Ghazanfar, A. A., & Santos, L. R. (2004). Primate brains in the wild: the sensory bases for social interactions. *Nature Reviews Neuroscience, 5*(8), 603-616.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience, 4*(3), 179-192.

Graziano, M. S., Taylor, C. S., & Moore, T. (2002). Complex movements evoked by microstimulation of precentral cortex. *Neuron, 34*(5), 841-851.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Grèzes, J., Pichon, S., & de Gelder, B. (2007). Perceiving fear in dynamic body expressions. *Neuroimage, 35*(2), 959-967.

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

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

Hadjikhani, N., & de Gelder, B. (2003). Seeing fearful body expressions activates the fusiform cortex and amygdala. *Current Biology, 13*(24), 2201-2205.

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

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. *Journal of Neuroscience, 14*(11 Pt 1), 6336-6353.

Hodsoll, S., Viding, E., & Lavie, N. (2011). Attentional capture by irrelevant emotional distractor faces. *Emotion, 11*(2), 346-353.

Iacoboni, M., Koski, L. M., Brass, M., Bekkering, H., Woods, R. P., Dubeau, M. C., et al. (2001). Reafferent copies of imitated actions in the right superior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 98*(24), 13995-13999.

James, W. (1890). *Principles of Psychology*. New York, NY: Holt.

Jolij, J., & Lamme, V. A. (2005). Repression of unconscious information by conscious processing: evidence from affective blindsight induced by transcranial magnetic stimulation. *Proceedings of the National Academy of Sciences of the United States of America, 102*(30), 10747-10751.

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

Kilner, J. M., Neal, A., Weiskopf, N., Friston, K. J., & Frith, C. D. (2009). Evidence of mirror neurons in human inferior frontal gyrus. *Journal of Neuroscience, 29*(32), 10153-10159.

Koivisto, M., Railo, H., Revonsuo, A., Vanni, S., & Salminen-Vaparanta, N. (2011). Recurrent processing in V1/V2 contributes to categorization of natural scenes. *Journal of Neuroscience, 31*(7), 2488-2492.

Kret, M., & de Gelder, B. (2010). Social context influences recognition of bodily expressions. *Experimental Brain Research, 203*, 169-180.

Kret, M. E., Pichon, S., Grèzes, J., & de Gelder, B. (2011). Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study. *Neuroimage, 54*(2), 1755-1762.

Krettek, J. E., & Price, J. L. (1977). Projections from the amygdaloid complex to the cerebral cortex and thalamus in the rat and cat. *The Journal of Comparative Neurology, 172*(4), 687-722.

Kunst-Wilson, W. R., & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science, 207*(4430), 557-558.

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

Lamme, V. A. (2006). Zap! Magnetic tricks on conscious and unconscious vision. *Trends in Cognitive Sciences, 10*(5), 193-195.

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

Lamme, V. A., Zipser, K., & Spekreijse, H. (2002). Masking interrupts figure-ground signals in V1. *Journal of Cognitive Neuroscience, 14*(7), 1044-1053.

Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences of the United States of America, 103*(49), 18763-18768.

LeDoux, J. E. (1996). *The emotional brain: The mysterious underpinnings of emotional life*. New York, NY, US: Simon and Schuster.

Levelt, W. J. (1965). Binocular Brightness Averaging and Contour Information. *British Journal of Psychology, 56*, 1-13.

Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center-periphery organization of human object areas. *Nature Neuroscience, 4*(5), 533-539.

Long, G. M., & Olszweski, A. D. (1999). To reverse or not to reverse: when is an ambiguous figure not ambiguous? *American Journal of Psychology, 112*(1), 41-71.

Lorteije, J. A., Kenemans, J. L., Jellema, T., van der Lubbe, R. H., de Heer, F., & van Wezel, R. J. (2006). Delayed response to animate implied motion in human motion processing areas. *Journal of Cognitive Neuroscience, 18*(2), 158-168.

Luppino, G., Calzavara, R., Rozzi, S., & Matelli, M. (2001). Projections from the superior temporal sulcus to the agranular frontal cortex in the macaque. *European Journal of Neuroscience, 14*(6), 1035-1040.

Lyon, D. C., Nassi, J. J., & Callaway, E. M. (2010). A disynaptic relay from superior colliculus to dorsal stream visual cortex in macaque monkey. *Neuron, 65*(2), 270-279.

Macmillan, N. A., & Creelman, C. D. (1991). Detection theory: A user's guide. New York: Cambridge University Press.

Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin and Review, 3*, 215-221.

McDonald, A. J. (1991). Organization of amygdaloid projections to the prefrontal cortex and associated striatum in the rat. *Neuroscience, 44*(1), 1-14.

McHugh, J. E., McDonnell, R., O'Sullivan, C., & Newell, F. N. (2011). Perceiving emotion in crowds: the role of dynamic body postures on the perception of emotion in crowded scenes. *Experimental Brain Research, 204*(3), 361-372.

Meeren, H. K., van Heijnsbergen, C. C., & de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences of the United States of America, 102*(45), 16518-16523.

Merabet, L., Thut, G., Murray, B., Andrews, J., Hsiao, S., & Pascual-Leone, A. (2004). Feeling by sight or seeing by touch? *Neuron, 42*(1), 173-179.

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.

Milders, M., Sahraie, A., & Logan, S. (2008). Minimum presentation time for masked facial expression discrimination. *Cognition & Emotion, 22*(1), 63-82.

Moro, V., Urgesi, C., Pernigo, S., Lanteri, P., Pazzaglia, M., & Aglioti, S. M. (2008). The neural basis of body form and body action agnosia. *Neuron, 60*(2), 235-246.

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

Morris, J. S., Öhman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proceedings of the National Academy of Sciences of the United States of America, 96*(4), 1680-1685.

Mottaghy, F. M., Gangitano, M., Sparing, R., Krause, B. J., & Pascual-Leone, A. (2002). Segregation of areas related to visual working memory in the prefrontal cortex revealed by rTMS. *Cerebral Cortex, 12*(4), 369-375.

Nishitani, N., & Hari, R. (2000). Temporal dynamics of cortical representation for action. *Proceedings of the National Academy of Sciences of the United States of America, 97*(2), 913-918.

Nishitani, N., & Hari, R. (2002). Viewing lip forms: cortical dynamics. *Neuron, 36*(6), 1211-1220.

Öhman, A. (2002). Automaticity and the amygdala: Nonconscious responses to emotional faces. *Current Directions in Psychological Science, 11*(2), 62-66.

Öhman, A. (2005). The role of the amygdala in human fear: Automatic detection of threat. *Psychoneuroendocrinology, 30*(10), 953-958.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: detecting the snake in the grass. *Journal of Experimental Psychology: General, 130*(3), 466-478.

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: a threat advantage with schematic stimuli. *Journal of Personality and Social Psychology, 80*(3), 381-396.

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

Panksepp, J. (2004). *Textbook of biological psychiatry*. Hoboken, New Jersey, US: Wiley-Liss.

Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science, 292*(5516), 510-512.

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

Peelen, M. V., Atkinson, A. P., Andersson, F., & Vuilleumier, P. (2007). Emotional modulation of body-selective visual areas. *Social, Cognitive and Affective Neurosciences, 2*, 274-283.

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

Pelphrey, K. A., Viola, R. J., & McCarthy, G. (2004). When strangers pass: processing of mutual and averted social gaze in the superior temporal sulcus. *Psychological Science, 15*(9), 598-603.

Perrett, D. I., Xiao, D., Barraclough, N. E., Keysers, C., & Oram, M. W. (2009). Seeing the future: Natural image sequences produce "anticipatory" neuronal activity and bias perceptual report. *Quarterly Journal of Experimental Psychology (Colchester)*, 1-24.

Pessoa, L. (2005). To what extent are emotional visual stimuli processed without attention and awareness? *Current Opinion in Neurobiology 15*(2), 188-196.

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance. *Nature Reviews Neuroscience, 11*(11), 773-783.

Pessoa, L., Japee, S., & Ungerleider, L. G. (2005). Visual awareness and the detection of fearful faces. *Emotion, 5*(2), 243-247.

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

Pichon, S., de Gelder, B., & Grezes, J. (2008). Emotional modulation of visual and motor areas by dynamic body expressions of anger. *Social Neuroscience, 3*(3-4), 199-212.

Pichon, S., de Gelder, B., & Grezes, J. (2009). Two different faces of threat. Comparing the neural systems for recognizing fear and anger in dynamic body expressions. *Neuroimage, 47*(4), 1873-1883.

Pichon, S., de Gelder, B., & Grezes, J. (in press). Threat prompts defensive brain responses independently of attentional control. *Cerebral Cortex*.

Pinto, J., & Shiffrar, M. (2009). The visual perception of human and animal motion in point-light displays. *Social Neuroscience, 4*(4), 332-346.

Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007). TMS evidence for the involvement of the right occipital face area in early face processing. *Current Biology, 17*(18), 1568-1573.

Pourtois, G., Sander, D., Andres, M., Grandjean, D., Reveret, L., Olivier, E., et al. (2004). Dissociable roles of the human somatosensory and superior temporal cortices for processing social face signals. *European Journal of Neuroscience, 20*(12), 3507-3515.

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *Journal of Neuroscience, 16*(16), 5205-5215.

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

Puce, A., & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philosophical Transactions of the Royal Society B: Biological Sciences, 358*(1431), 435-445.

Rennie, C. J., Wright, J. J., & Robinson, P. A. (2000). Mechanisms of cortical electrical activity and emergence of gamma rhythm. *Journal of Theoretical Biology, 205*(1), 17-35.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*(11), 1019-1025.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169-192.

Robichaud, L., & Stelmach, L. B. (2003). Inducing blindsight in normal observers. *Psychonomic Bulletin Review, 10*(1), 206-209.

Rotshtein, P., Malach, R., Hadar, U., Graif, M., & Hendler, T. (2001). Feeling or features: different sensitivity to emotion in high-order visual cortex and amygdala. *Neuron, 32*(4), 747-757.

Sah, P., Faber, E. S., Lopez De Armentia, M., & Power, J. (2003). The amygdaloid complex: anatomy and physiology. *Physiological Reviews, 83*(3), 803-834.

Saygin, A. P. (2007). Superior temporal and premotor brain areas necessary for biological motion perception. *Brain, 130*(Pt 9), 2452-2461.

Schiller, P. H., & Malpeli, J. G. (1977). Properties and tectal projections of monkey retinal ganglion cells. *Journal of Neurophysiology, 40*(2), 428-445.

Schindler, K., Van Gool, L., & de Gelder, B. (2008). Recognizing emotions expressed by body pose: a biologically inspired neural model. *Neural Networks, 21*(9), 1238-1246.

Schurger, A. (2009). A very inexpensive MRI-compatible method for dichoptic visual stimulation. *Journal of Neuroscience Methods, 177*(1), 199-202.

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

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America, 104*(15), 6424-6429.

Shewmon, D. A., Holmes, G. L., & Byrne, P. A. (1999). Consciousness in congenitally decorticate children: developmental vegetative state as self-fulfilling prophecy. *Developmental Medicine & Child Neurology, 41*(6), 364-374.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*(1), 34-50.

Spiridon, M., Fischl, B., & Kanwisher, N. (2006). Location and spatial profile of category-specific regions in human extrastriate cortex. *Human Brain Mapping, 27*(1), 77-89.

Stekelenburg, J. J., & de Gelder, B. (2004). The neural correlates of perceiving human bodies: an ERP study on the body-inversion effect. *Neuroreport, 15*(5), 777-780.

Stienen, B. M. C., & de Gelder, B. (2011). Fear detection and visual awareness in perceiving bodily expressions *Emotion, 11*(5), 1182-1189.

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.

Suzuki, W. A., & Amaral, D. G. (1994). Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *Journal of Comparative Neurology, 350*(4), 497-533.

Talairach, J., & Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain*. New York: Thieme Medical Publishers.

Tamietto, M., Castelli, L., Vighetti, S., Perozzo, P., Geminiani, G., Weiskrantz, L., et al. (2009). Unseen facial and bodily expressions trigger fast emotional reactions. *Proceedings of the National Academy of Sciences of the United States of America, 106*(42), 17661-17666.

Tamietto, M., & de Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. *Nature Reviews Neuroscience, 11*(10), 697-709.

Tamietto, M., Geminiani, G., Genero, R., & de Gelder, B. (2007). Seeing fearful body language overcomes attentional deficits in patients with neglect. *Journal of Cognitive Neuroscience, 19*(3), 445-454.

Tanaka, A., Koizumi, A., Imai, H., Hiramatsu, S., Hiramoto, E., & de Gelder, B. (2010). I feel your voice: cultural differences in the multisensory perception of emotion. *Psychological Science, 21*(9), 1259-1262.

Taylor, J. C., Roberts, M. V., Downing, P. E., & Thierry, G. (2010). Functional characterisation of the extrastriate body area based on the N1 ERP component. *Brain and Cognition, 73*(3), 153-159.

Taylor, J. C., Wiggett, A. J., & Downing, P. E. (2007). Functional MRI analysis of body and body part representations in the extrastriate and fusiform body areas. *Journal of Neurophysiology, 98*(3), 1626-1633.

Tong, F., Meng, M., & Blake, R. (2006). Neural bases of binocular rivalry. *Trends in Cognitive Sciences, 10*(11), 502-511.

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

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. (2003). Faces and objects in macaque cerebral cortex. *Nature Neuroscience, 6*(9), 989-995.

Tsao, D. Y., Moeller, S., & Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences of the United States of America, 105*(49), 19514-19519.

Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Current Opinion in Neurobiology, 4*(2), 157-165.

Urgesi, C., Berlucchi, G., & Aglioti, S. M. (2004). Magnetic stimulation of extrastriate body area impairs visual processing of nonfacial body parts. *Current Biology, 14*(23), 2130-2134.

Urgesi, C., Calvo-Merino, B., Haggard, P., & Aglioti, S. M. (2007). Transcranial magnetic stimulation reveals two cortical pathways for visual body processing. *Journal of Neuroscience, 27*(30), 8023-8030.

Urgesi, C., Candidi, M., Ionta, S., & Aglioti, S. M. (2007). Representation of body identity and body actions in extrastriate body area and ventral premotor cortex. *Nature Neuroscience, 10*(1), 30-31.

van de Riet, W. A., Grezes, J., & de Gelder, B. (2009). Specific and common brain regions involved in the perception of faces and bodies and the representation of their emotional expressions. *Social Neuroscience, 4*(2), 101-120.

Van den Stock, J., de Jong, J. J., Hodiamont, P. P. G., & de Gelder, B. (in press). Perceiving emotions from bodily expressions and multisensory integration of emotion cues in schizophrenia. *Social Neuroscience*.

Van den Stock, J., Grèzes, J., & de Gelder, B. (2008). Human and animal sounds influence recognition of body language. *Brain Research, 1242*, 185-190.

Van den Stock, J., Peretz, I., Grèzes, J., & de Gelder, B. (2009). Instrumental music influences recognition of emotional body language. *Brain Topography, 21*(3-4), 216-220.

Van den Stock, J., Righart, R., & de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion, 7*(3), 487-494.

Van den Stock, J., Tamietto, M., Sorger, B., Pichon, S., & de Gelder, B. (in press). Cortico-subcortical visual, somatosensory and motor activations for perceiving dynamic whole-body emotional expressions with and without V1. *Proceedings of the National Academy of Sciences of the United States of America*.

Van den Stock, J., van de Riet, W. A., Righart, R., & de Gelder, B. (2008). Neural correlates of perceiving emotional faces and bodies in developmental prosopagnosia: an event-related fMRI-study. *PLoS ONE, 3*(9), e3195.

Vroomen, J., Driver, J., & de Gelder, B. (2001). Is cross-modal integration of emotional expressions independent of attentional resources? *Cognitive, Affective, & Behavioral Neuroscience, 1*(4), 382-387.

Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nature Neuroscience, 7*(11), 1271-1278.

Wassermann, E. M. (1998). Risk and safety of repetitive transcranial magnetic stimulation: report and suggested guidelines from the International Workshop on the Safety of Repetitive Transcranial Magnetic Stimulation, June 5-7, 1996. *Electroencephalography and Clinical Neurophysiology, 108*(1), 1-16.

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. *Journal of Neuroscience, 18*(1), 411-418.

Wiens, S. (2006). Current concerns in visual masking. *Emotion, 6*(4), 675-680.

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. *Journal of Neuroscience, 24*(12), 2898-2904.

Yoon, K. L., Hong, S. W., Joormann, J., & Kang, P. (2009). Perception of facial expressions of emotion during binocular rivalry. *Emotion, 9*(2), 172-182.

Yu, K., & Blake, R. (1992). Do recognizable figures enjoy an advantage in binocular rivalry? *Journal of Experimental Psychology: Human Perception and Performance, 18*(4), 1158-1173.

Zhang, P., Jamison, K., Engel, S., He, B., & He, S. (2011). Binocular rivalry requires visual attention. *Neuron, 71*(2), 362-369.

List of papers on which the thesis is based, their status and other papers

Chapter 2

Stienen, B. M. C., & de Gelder, B. (2011). Fear detection and visual awareness in perceiving bodily expressions. *Emotion, 11*(5), 1182-1189

Chapter 3

Stienen, B. M. C., & de Gelder, B. (2011). Fear modulates visual awareness similarly for facial and bodily expressions. *Frontiers in Human Neuroscience, 5:132,* doi:10.3389/fnhum.2011.00132.

Chapter 4

Stienen, B. M. C., Tanaka, A, & de Gelder, B. (2011). Emotional voice and emotional body postures influence each other independently of visual awareness. *PLoS ONE, 6*(10), e25517.

Chapter 5

Stienen, B. M. C., Konrad, S., & de Gelder, B. (in press). A computational feed-forward model predicts categorization of masked emotional body language for longer, but not for shorter latencies. *Neural Computation.*

Chapter 6

Candidi[2], M., Stienen[2], B. M. C., Aglioti, S. M., & de Gelder, B. (2011). Event related repetitive transcranial magnetic stimulation of posterior superior temporal sulcus improves the detection of threatening postural changes in human bodies. Journal of Neuroscience, 31(48), 17547-17555.

Chapter 7

Stienen, B.M.C., Candidi, M., de Gelder, B., Alioti, S.M.A. (in preparation). Dissociating conscious perception of fearful faces and bodies by transient inhibition of right pSTS.

Other papers:

de Gelder, B., Tamietto, M., van Boxtel, G., Goebel, R., Sahraie, A., Van den Stock, J., Stienen, B. M. C., Weiskrantz, L., & Pegna, A. (2008). Intact navigation skills after bilateral loss of striate cortex. *Current Biology*, 18(24), R1128-R1129

Stienen, B. M. C., & de Gelder, B. (2010). Relative affective blindsight for fearful bodily expressions. *Proceedings of Measuring Behavior 2010*.

de Gelder, B., Stienen, B. M. C., & Van den Stock, J. (in press). Integrating face and voice. In P. Belin (Ed.), *Person Perception*. New York: Springer.

---

[2] Equal contribution