From Representation to Recognition: MEG Studies of Face Perception

by

Jia Liu

M.S., Cognitive Psychology,
Beijing University, China, 1997
B.A., Psychology
Beijing University, China, 1995

Submitted to the Department of Brain and Cognitive Sciences in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Cognitive Neuroscience

at the

Massachusetts Institute of Technology

February 2003

Signature of Author: ____
Department of Brain and Cognitive Sciences
September 23, 2002

Certified by: ____
Nancy G. Kanwisher
Professor of Cognitive Neuroscience
Thesis Supervisor

Accepted by: ____
Earl K. Miller
Professor of Neuroscience
Chairman, Department Graduate Committee

From Representation to Recognition: MEG Studies of Face Perception

by

Jia Liu

Submitted to the Department of Brain and Cognitive Sciences
on September 23, 2002, in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy in Cognitive Neuroscience

Abstract

Face recognition is one of the most important problems our visual system must solve. Here I used magnetoencephalography (MEG) in an effort to characterize the sequence of cognitive and neural processes underlying this remarkable ability.

This work is designed to answer several questions. First, how long does it take for the human visual system to recognize a stimulus as a face? Second, what are the stages of processing in face perception? Finally, what is the nature of representations extracted at each of these stages?

MEG provides an ideal tool for addressing these questions, as its high temporal resolution enables us to separately measure perceptual operations that may occur only a few tens of milliseconds apart from each other. Yet, unlike single-unit recording, it can be used in normal human subjects.

Three new findings about human face recognition will be reported in this thesis. First, a face stimulus begins to be categorized as a face within 100 ms after stimulus onset in humans, substantially faster than previously thought. Second, face recognition occurs in two distinct stages: an initial stage at which the stimulus is categorized as a face, and a stage that occurs 70 ms later at which the individual identity of the face is extracted. Finally, the representations extracted at these two stages differ not only in specificity, but also in the aspects of a face represented at each stage.

Thesis Supervisor: **Nancy G. Kanwisher**

Title: **Professor of Cognitive Neuroscience**

Acknowledgements

Thanks to Nancy Kanwisher for being the best advisor I could imagine. I have benefited greatly not only from her keen sense of experimental design and her broad knowledge of this field but also from her extremely detailed and invaluable comments on all of my posters, talks, and papers. Under her guidance and encouragement my PhD training at MIT is a joy.

Thanks to Pawan Sinha, Ken Nakayama and Jim DiCarlo not only for serving on my thesis committee but also for inspiring and sharpening my thoughts in many ways.

Thanks to many BCS people, including past and present members of Kanwisher Lab, for stimulating collaborations, valuable discussion, and everything else. Special thanks are due to Alison Harris for assisting me running experiments and collecting data; without her many findings described in this thesis would have been nothing more than hypothesis. To Bart Anderson for the class that introduced me to the wonderful world of psychophysics and computational vision. To Steven Pinker and Molly Potter for showing me the art of teaching when I served as a teaching assistant in their courses. To Denise Heintze for straightening out my administrative tangles. I count myself lucky to have had a chance to study in this wonderful department.

Finally, thanks to my parents and my wife, for everything. To them this thesis is dedicated.

Contents

1 Introduction 5

1.1 The Speed of face recognition in the human visual system 5
1.2 Stages of processing in face perception: Categorization versus Identification 8
1.3 Face-specific processing 9
1.4 Specific Aims 16

2 The Selectivity of the occipitotemporal M170 for faces 24

2.1 Introduction 25
2.2 Results: Selectivity, Generality and Face Inversion Effect 27
2.3 Discussion 29
2.4 Conclusion 31
2.5 Methods 32
2.6 Acknowledgement 34

3 Stages of processing in face perception 40

3.1 Introduction 41
3.2 Results 43
  3.2.1 A Face-Selective Response at a Latency of 100 ms 43
  3.2.2 Decoupling categorization and identification 45
3.3 Discussion 49
3.4 Methods 53
3.5 Acknowledgement 58

4 Parts before Configurations in Face Processing 67

4.1 Introduction 68

4.2 Results 71

4.2.1 Face parts versus face configurations 71

4.2.2 Face parts versus object parts 74

4.2.3 Face Inversion 75

4.3 Discussion 76

4.4 Methods 82

4.4 Acknowledgement 84

5 General Conclusions and Future Work 93

5.1 Early stages of face recognition rely on feedforward network 93

5.2 Two stages of face processing: Categorization begins prior to identification 94

5.3 Multiple levels of face representations: parts before configurations 95

5.4 Remaining Questions 97

5.4.1 The relation between neural substrates underlying the M100 and the M170 98

5.4.2 Face parts versus face configuration 98

5.4.3 Faces versus objects 99

5.5 Bruce & Young model revised 100

Chapter 1

Introduction

1.1 The Speed of face recognition in the human visual system

The human visual system can recognize faces with astonishing speed and accuracy. But exactly how fast is human face recognition? The answer to this question may provide empirical constraints on theories of face recognition (Thorpe & Imbert, 1989).

Several event-related potential (ERP) and magnetoencephalography (MEG) studies suggest that human face and object recognition begins within 150-170 ms after stimulus onset. For example, using a go/no-go categorization task, Thorpe et al. (1996) instructed subjects to release a button if a natural image that was presented for 20 ms contained an animal ("go"), or to inhibit such a motor response if it did not ("no-go"). ERP analysis revealed a more negative response over frontal sensors specific to no-go trials that developed roughly 150 ms after stimulus onset. Since this difference in neural activity was correlated with the decision, it implies that the recognition of objects in the visual scene has begun within 150 ms after stimulus onset. This latency range is in agreement with the earliest significant differences in the ERP response found between faces and nonfaces (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Jeffreys, 1996), vehicles and nonvehicles (VanRullen & Thorpe, 2001), and familiar and novel objects (Kanwisher & DiGirolamo, 1995).

Several recent ERP studies in humans have reported even earlier differences in ERP responses to different object categories ranging from 30 – 130 ms. For example, by recording event-related potentials elicited by repeated versus non-repeated faces, Seeck et al. (Seeck et al., 1997) found that the potentials elicited by the first presentation of novel faces (non-repeated) began to differ from those to the second presentation (repeated faces) as early as 50 ms after stimulus onset. The authors concluded that this effect implies that face recognition may begin within 50 ms after stimulus onset (see also Braeutigam, Bailey, & Swithenby, 2001). However, this difference may instead reflect an effect of repetition of the precise physical properties of the photographs, rather than face recognition *per se*. Evidence favoring this argument comes from a study on repetition effects using two-tone "Mooney" faces (George, Jemel, Fiori, & Renault, 1997). Because inversion of Mooney faces made face recognition impossible, different percepts were created by presenting images upright and inverted whereas many of the physical properties of the images were held constant. If the repetition effect found by Seeck et al. (1996) were specific to face recognition, this early difference should diminish when inverted Mooney faces are repeated. This prediction is violated in George et al.'s experiment, which showed that ERPs to the first presentation of images differed from those to the second one as early as 50 ms after stimulus onset, regardless whether upright or inverted Mooney faces were repeated (see also Debruille, Guillem, & Renault, 1998).

Another vein of research on early category-specific ERP responses (<150 ms) has focused on the comparison of evoked potentials to different object categories. For example, differential neural activity has been reported between written words and faces (Schendan, Ganis, & Kutas, 1998), and female versus male faces and hands (Mouchetant-Rostaing, Giard, Delpuech, Echallier, & Pernier, 2000; Mouchetant-Rostaing, Giard, Bentin, Aguera, & Pernier, 2000) at a latency of 90-125 and 50-65 ms, respectively. However, findings that neural activity is different in response to different categories of stimuli are insufficient to demonstrate that these neural makers reflect categorization of the stimuli. Instead, it could simply reflect low-level differences present in stimuli (VanRullen et al., 2001).

In sum, prior work suggests that human face recognition may start earlier than previously thought. However, one critical issue has not been resolved. In these studies category selectivity was not unconfounded from differential responses to low-level visual features. Dissociating visual inputs from visual percepts of objects is critical for identifying category-specific responses, because objects from different categories may share similar visual properties (e.g. an apple vs. a tennis ball) yet objects from the same category may look quite different (e.g. an apple vs. a banana). Therefore, only when neural activity is correlated with subjects' percept (i.e. decision) rather than with the visual input, can it indicate that the recognition of objects has begun.

1.2 Stages of processing in face perception: Categorization versus Identification

Objects can be recognized at several levels of abstraction, including classifying an object into its basic-level category (categorization), and extracting the critical information necessary to identify and recognize individuals (identification). Rosch et al. (1976) found that subjects preferred to name objects in terms of their basic classes (e.g. "dog") compared to their subordinate (e.g. "golden retriever") or superordinate class (e.g. "animal"), and verified object names more quickly at the basic than subordinate level. In addition, when subjects were asked to list as many attributes as they could that applied to certain objects, the greatest increase in the number of characteristic features that described objects was found at the basic level compared to at other levels. These findings suggest that the basic level category may be the "entry point" in object recognition, where the perceptual stimulus first makes contact with its stored presentation (Jolicoeur, Gluck, & Kosslyn, 1984).

Evidence that a similar sequence may occur in face perception comes from single-unit recording in macaques showing that the initial transient response of face-selective neurons in inferotemporal cortex reflects a rough categorization of face versus nonface, whereas subsequent firing of the same neural population represents finer information such as the facial expression or identity of the person (Sugase, Yamane, Ueno, & Kawano, 1999). However, it has been argued that visual expertise in discriminating exemplars of a specific visual category may shift the point of initial contact with memory representations from the basic level to the subordinate level. For example, airplane mechanic experts were able to identify a stimulus as a "Boeing 707" as quickly as they categorized it as a "plane" (Rosch et al., 1976). Given that we are all experts at face recognition, this hypothesis predicts that we should be able to identify an individual face as fast as we can tell that it is a face at all. Indeed, a recent behavioral study showed that face identification occurred more frequently than and as quickly as face categorization (Tanaka, 2001). However, other studies showed the opposite results that categorizing a stimulus as a face takes substantially less time than identifying the specific face (Grill-Spector & Kanwisher, 2001).

Because behavioral RT measurements reflect the sum of all stages of processing, they do not enable us to determine the absolute amount of time needed for each process. On the other hand, neural measurements can answer these questions by providing a direct on-line measure of processing.

1.3 Face-specific processing

Appreciation of the specialness of the face for human and primates can be traced back at least to Darwin's *The Expression of Emotion in Man and Animals* (1872), but intensive research on face recognition in normal individuals started only a few decades ago. In a seminal study, Yin (1969) found that inversion impaired subjects' ability to recognize faces more than it impaired recognition of other mono-oriented stimuli such as houses (reviewed by Valentine, 1988a). Subsequent developments demonstrated the inversion effect in a variety of tasks. In a famous-nonfamous discrimination task, face inversion increased reaction time and decreased identification accuracy (Valentine, 1988b). In a visual search task, the ability to search for a specific inverted face or reject an inverted distracter face suffered as compared to upright faces (Tong & Nakayama, 1999).

In addition to the effects of inversion, face recognition also appears to be more "holistic" than object recognition. For example, accuracy at remembering individual face parts, such as the nose, was higher when the entire face was presented than when the parts were presented in isolation, but the same holistic advantage was absent for parts of houses (Tanaka & Farah, 1993) or inverted faces (Tanaka & Sengco, 1997). In addition, when the top and bottom halves of different familiar faces were aligned with each other, the reaction time to name the person depicted in the top half was slower, compared to when the two halves were not aligned (Young, Hellawell, & Hay, 1987; Carey & Diamond, 1994). Apparently, the two halves of the composites fused into a new (unfamiliar) face and the identification of the top half was influenced by the bottom half. The fused face, however, was easily disentangled when the composites were inverted (Young et al., 1987; Carey et al., 1994). The fact that both kinds of evidence for holistic processing of faces are found only for upright, not inverted, faces in these and other experiments demonstrates an empirical link between inversion sensitivity and reliance on holistic processing.

The apparently unique characteristics of face processing observed in behavioral studies in humans suggest the existence of specialized neural mechanisms involved in face processing. The first evidence supporting this idea comes from single-unit recording at area TE and the adjacent area STP in monkeys, where a small set of neurons with responses selective for faces was found (Gross, Rocha-Miranda, & Bender, 1972). Follow-up studies found face neurons in STS and IT (Baylis, Rolls, & Leonard, 1987), amygdala (Sanghera, Rolls, & Roper-Hall, 1979; Leonard, Rolls, Wilson, & Baylis, 1985; Rolls, 1992), ventral striatum (Rolls, Tovee, Purcell, Stewart, & Azzopardi, 1994; Williams, Rolls, Leonard, & Stern, 1993), and the inferior convexity of the prefrontal cortex (Wilson & Goldman-Rakic, 1994). Although face neurons vary in the nature of their preference for faces, many give almost no response to other stimuli, including nonface objects, textures, and colors (Bruce, Desimone, & Gross, 1981). Furthermore, for some of these neurons the selectivity of the response for faces generalizes across a variety of stimulus transformations such as changes in size, position and luminance (Perrett, Hietanen, Oram, & Benson, 1992; Hietanen, Perrett, Oram, Benson, & Dittrich, 1992).

Evidence from humans also implicates distinct neural mechanisms for face recognition. In prosopagnosia, patients lose the ability to recognize familiar faces but their ability to categorize a face stimulus as a face is largely intact (reviewed by De Renzi, Perani, Carlesimo, Silveri, & Fazio, 1994; Young, 1992). Although prosopagnosia is often accompanied by varying levels of object agnosia (e.g. Damasio, Damasio, & Van Hoesen, 1982; Gauthier, Behrmann, & Tarr, 1999), a few studies report the absence of impairments in object recognition in their prosopagnosic patients (e.g. Bentin, Deouell, & Soroker, 1999; Farah, Levinson, & Klein, 1995; Henke, Schweinberger, Grigo, Klos, & Sommer, 1998; McNeil & Warrington, 1993; Wada & Yamamoto, 2001). Conversely, Moscovitch et al. (1997) reported a patient who suffered from severe object agnosia yet had normal face recognition performance (see also Farah, McMullen, & Meyer, 1991; Feinberg, Schindler, Ochoa, Kwan, & Farah, 1994). The double dissociation between face and object recognition implies that the face recognition system is dissociable and anatomically distinct from the object recognition system.

Further support for a distinct neural substrate for face processing comes from neuroimaging studies of normal humans. Several early Positron Emission Tomography (PET) studies showed that regions in the occipital and occipitotemporal cortex became active during face recognition tasks (Haxby et al., 1994; Sergent, Ohta, & MacDonald, 1992), and this activation was later localized in the fusiform gyrus by functional Magnetic Resonance Imaging (fMRI) (Puce, Allison, Gore, & McCarthy, 1995; Puce, Allison, Asgari, Gore, & McCarthy, 1996; Clark et al., 1996). More recent fMRI studies have revealed several regions that show selectivity for faces, i.e., stronger responses to faces than control objects: a region in the fusiform gyrus (the Fusiform Face Area, FFA), the middle occipital lobe (the Occipital Face Area, OFA), and the superior temporal sulcus (STS) (Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997; Hoffman & Haxby, 2000; Gauthier, Skudlarski, Gore, & Anderson, 2000). Follow-up fMRI studies tested a wide variety of object categories under different cognitive tasks and characterized the selectivity of these neural substrates, especially the FFA. For example, the FFA generalizes across a variety of face stimuli (e.g., front-view photographs of faces, line drawings of faces, cat faces, cartoon faces, and upside-down faces) but responds less strongly to nonface control stimuli (e.g. animals, flowers, cars houses, and hands) (Tong, Nakayama, Moscovitch, Weinrib, & Kanwisher, 2000; reviewed by Kanwisher, 2000).

FMRI findings indicate neuroanatomical specificity for face processing. However, these studies provide no information about the temporal structure of face perception. On the other hand, event-related electromagnetic recordings can address these questions by continuously measuring neural activity with high temporal resolution. Scalp ERP recordings from normal subjects have revealed a posterior-lateral negative response component peaking at a latency of 170 ms (N170; Bentin et al., 1996; George, Evans, Fiori, Davidoff, & Renault, 1996) and a positive response component at the scalp vertex peaking at 190 ms (Vertex Positive Potential or VPP; Jeffreys, 1996), whose amplitude is considerably larger in response to human faces than in response to a variety of nonface objects, such as cars, scrambled faces, items of furniture, or human hands. This finding indicates that by about 170 ms after stimulus onset face processing begins to differ neurally from object processing.

In accordance with scalp ERP recordings, direct measurements from the cortical surface using subdural electrodes in epilepsy patients reveal that discrete regions in the human ventral pathway produce a selective negative response to faces at a latency of 200 ms (N200), but not to cars, butterflies, scrambled faces, and letterstrings (Allison, McCarthy, Nobre, Puce, & Belger, 1994; Allison et al., 1994). This response presumably corresponds approximately to the N170 component in scalp ERP studies; the later latencies in the intracranial case may arise from medication and/or histories of epilepsy typical of that subject population. A more recent study has shown that the face-specific N200 sites are clustered in the ventral occipitotemporal cortex, primarily in the lateral fusiform gyrus (Puce, Allison, & McCarthy, 1999; McCarthy, Puce, Belger, & Allison, 1999; Allison, Puce, Spencer, & McCarthy, 1999), consistent with the locus of the face-sepcific neural mechanisms suggested by fMRI studies (Kanwisher et al., 1997; McCarthy et al., 1997; Halgren et al., 1999). Furthermore, electrical stimulation at the N200 sites (lateral fusiform gyrus) impairs face recognition to briefly presented face images (Puce et al., 1999), indicating the necessity of these regions for face recognition.

These studies indicate the existence of face-specific processing mechanisms that are anatomically and functionally distinct from those involved in processing nonface objects. However, it has been argued that these mechanisms may not be selective for faces *per se*, but rather for any stimulus domain in which observers have acquired expertise through extensive training on within-category discrimination (Tarr & Gauthier, 2000). Previous studies on expertise found a disproportional impairment in recognizing over-learned and highly similar nonface objects when they were inverted (e.g. different breeds of dog; Diamond & Carey, 1986; Tanaka & Gauthier, 1997). More recent studies have used artificial creature-like "Greeble" stimuli which shared similar parts in a common spatial configuration to test the "expert-subordinate-level recognition" hypothesis, and some effects thought to reflect face-exclusive mechanisms were found with Greebles (Gauthier, Anderson, Tarr, Skudlarski, & Gore, 1997; Gauthier & Tarr, 1997; Gauthier & Tarr, 2002). In addition, the activity of the right FFA was enhanced for cars in car experts and birds in bird experts (Gauthier et al., 2000). On the other hand, visual expertise alone cannot explain the unique characteristics of face processing. For example, McKone et al. (2001) showed that the physical continuum of face stimuli created by morphing one individual to another was perceived as two categories (categorical perception) and accuracy was higher at discriminating a pair of morphed face stimuli which crossed the category boundary than a physically equidistant pair which fell within the same perceptual category (see also Beale & Keil, 1995). However, even after around 30,000 exposures to the same pair of faces for 2.5 years of training, the increase in discriminability at the categorical boundary found for upright faces was still absent for inverted ones, indicating face recognition is not necessarily affected by extensive training in adults (McKone et al., 2001). Indeed, in patients whose bilateral cataracts were surgically corrected at 2~6 months of age, discrimination between faces that differed in the relative positions of facial features was impaired even after an average of 14 years of experience with faces after surgery (Le Grand, Mondloch, Maurer, & Brent, 2001).

1.4 Specific Aims

I. One of the basic aims of this work is to determine the exact speed of face recognition in the human visual system. Several previous studies suggest that face recognition may occur within 100 ms after stimulus onset, but they are open to explanations in terms of nonspecific repetition effects or differences in the low-level features present in the stimuli. Here we will unconfound stimulus category from its low-level visual properties to explore when the first truly face-selective response occurs. The answer to this question may constrain models of face recognitions, for example, by strengthening the evidence that face recognition is a largely feedforward process.

II. Evidence from behavioral and neurophysiological studies suggests that basic-level object categorization occurs before subordinate-level identification. Will these findings hold also for face recognition, given that we are all experts in face recognition and given that the primary goal of face perception is to identify individuals?

III. Are categorization and identification simply different points on a continuous spectrum of discrimination, as the selectivity of the underlying neural substrates continuously sharpens over time, permitting coarse discriminations early, and more fine-grained discriminations only later? Alternatively, are categorization and identification two distinct processes relying on qualitatively different representations?

Reference List

Allison, T., Ginter, H., McCarthy, G., Nobre, A. C., Puce, A., Luby, M. et al. (1994). Face recognition in human extrastriate cortex. *J.Neurophysiol., 71,* 821-825.

Allison, T., McCarthy, G., Nobre, A., Puce, A., & Belger, A. (1994). Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cereb.Cortex, 4,* 544-554.

Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb.Cortex, 9,* 415-430.

Baylis, G. C., Rolls, E. T., & Leonard, C. M. (1987). Functional subdivisions of the temporal lobe neocortex. *J.Neurosci., 7,* 330-342.

Beale, J. M. & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition, 57,* 217-239.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perceptions in humans. *Journal of Cognitive Neuroscience, 8,* 551-565.

Bentin, S., Deouell, L. Y., & Soroker, N. (1999). Selective visual streaming in face recognition: evidence from developmental prosopagnosia. *Neuroreport, 10,* 823-827.

Braeutigam, S., Bailey, A. J., & Swithenby, S. J. (2001). Task-dependent early latency (30-60 ms) visual processing of human faces and other objects. *Neuroreport, 12,* 1531-1536.

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J.Neurophysiol., 46,* 369-384.

Carey, S. & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition, 1,* 253-274.

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

Damasio, A. R., Damasio, H., & Van Hoesen, G. W. (1982). Prosopagnosia: anatomic basis and behavioral mechanisms. *Neurology, 32,* 331-341.

De Renzi, E., Perani, D., Carlesimo, G. A., Silveri, M. C., & Fazio, F. (1994). Prosopagnosia can be associated with damage confined to the right hemisphere--an MRI and PET study and a review of the literature. *Neuropsychologia, 32,* 893-902.

Debruille, J. B., Guillem, F., & Renault, B. (1998). ERPs and chronometry of face recognition: following-up Seeck et al. and George et al. *Neuroreport, 9,* 3349-3353.

Diamond, R. & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *J.Exp.Psychol.Gen., 115,* 107-117.

Farah, M. J., Levinson, K. L., & Klein, K. L. (1995). Face perception and within-category discrimination in prosopagnosia. *Neuropsychologia, 33,* 661-674.

Farah, M. J., McMullen, P. A., & Meyer, M. M. (1991). Can recognition of living things be selectively impaired? *Neuropsychologia, 29,* 185-193.

Feinberg, T. E., Schindler, R. J., Ochoa, E., Kwan, P. C., & Farah, M. J. (1994). Associative visual agnosia and alexia without prosopagnosia. *Cortex, 30,* 395-411.

Gauthier, I., Anderson, A. W., Tarr, M. J., Skudlarski, P., & Gore, J. C. (1997). Levels of categorization in visual recognition studied using functional magnetic resonance imaging. *Curr.Biol., 7,* 645-651.

Gauthier, I., Behrmann, M., & Tarr, M. J. (1999). Can face recognition really be dissociated from object recognition? *J.Cogn Neurosci., 11,* 349-370.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nat.Neurosci., 3,* 191-197.

Gauthier, I. & Tarr, M. J. (1997). Becoming a "Greeble" expert: exploring mechanisms for face recognition. *Vision Res., 37,* 1673-1682.

Gauthier, I. & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: bridging brain activity and behavior. *J.Exp.Psychol.Hum.Percept.Perform., 28,* 431-446.

George, N., Evans, J., Fiori, N., Davidoff, J., & Renault, B. (1996). Brain events related to normal and moderately scrambled faces. *Brain Res.Cogn Brain Res., 4,* 65-76.

George, N., Jemel, B., Fiori, N., & Renault, B. (1997). Face and shape repetition effects in humans: a spatio-temporal ERP study. *Neuroreport, 8,* 1417-1423.

Grill-Spector, K. & Kanwisher, N. (2001). Common cortical mechanisms for different components of visual object recognition: A combined behavioral and fMRI Study [abstract]. *Journal of Vision, 1,* 474a.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *J.Neurophysiol., 35,* 96-111.

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

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

Henke, K., Schweinberger, S. R., Grigo, A., Klos, T., & Sommer, W. (1998). Specificity of face recognition: recognition of exemplars of non-face objects in prosopagnosia. *Cortex, 34,* 289-296.

Hietanen, J. K., Perrett, D. I., Oram, M. W., Benson, P. J., & Dittrich, W. H. (1992). The effects of lighting conditions on responses of cells selective for face views in the macaque temporal cortex. *Exp.Brain Res., 89,* 157-171.

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

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

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: making the connection. *Cognit.Psychol., 16,* 243-275.

Kanwisher, N. (2000). Domain specificity in face perception. *Nat.Neurosci., 3,* 759-763.

Kanwisher, N. & DiGirolamo, G. (1995). Accessing stored representations begins within 155 ms of object recognition. Paper presented at the meeting of the psychonomics Society. 1995. Los Angeles.

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

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Neuroperception. Early visual experience and face processing. *Nature, 410,* 890.

Leonard, C. M., Rolls, E. T., Wilson, F. A., & Baylis, G. C. (1985). Neurons in the amygdala of the monkey with responses selective for faces. *Behav.Brain Res., 15,* 159-176.

McCarthy, G., Puce, A., Belger, A., & Allison, T. (1999). Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex. *Cereb.Cortex, 9,* 431-444.

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-Specific Processing in the Human Fusiform Gyrus. *Journal of Cognitive Neuroscience, 9,* 604-609.

McKone, E., Martini, P., & Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *J.Exp.Psychol.Hum.Percept.Perform., 27,* 573-599.

McNeil, J. E. & Warrington, E. K. (1993). Prosopagnosia: a face-specific disorder. *Q.J.Exp.Psychol.A, 46,* 1-10.

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience, 9,* 555-604.

Mouchetant-Rostaing, Y., Giard, M. H., Bentin, S., Aguera, P. E., & Pernier, J. (2000). Neurophysiological correlates of face gender processing in humans. *Eur.J.Neurosci., 12,* 303-310.

Mouchetant-Rostaing, Y., Giard, M. H., Delpuech, C., Echallier, J. F., & Pernier, J. (2000). Early signs of visual categorization for biological and non-biological stimuli in humans. *Neuroreport, 11,* 2521-2525.

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philos.Trans.R.Soc.Lond B Biol.Sci., 335,* 23-30.

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. *J.Neurosci., 16,* 5205-5215.

Puce, A., Allison, T., Gore, J. C., & McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex studied by functional MRI. *J.Neurophysiol., 74,* 1192-1199.

Puce, A., Allison, T., & McCarthy, G. (1999). Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials. *Cereb.Cortex, 9,* 445-458.

Rolls, E. T. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philos.Trans.R.Soc.Lond B Biol.Sci., 335,* 11-20.

Rolls, E. T., Tovee, M. J., Purcell, D. G., Stewart, A. L., & Azzopardi, P. (1994). The responses of neurons in the temporal cortex of primates, and face identification and detection. *Exp.Brain Res., 101,* 473-484.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382-349.

Sanghera, M. K., Rolls, E. T., & Roper-Hall, A. (1979). Visual responses of neurons in the dorsolateral amygdala of the alert monkey. *Exp.Neurol., 63,* 610-626.

Schendan, H. E., Ganis, G., & Kutas, M. (1998). Neurophysiological evidence for visual perceptual categorization of words and faces within 150 ms. *Psychophysiology, 35,* 240-251.

Seeck, M., Michel, C. M., Mainwaring, N., Cosgrove, R., Blume, H., Ives, J. et al. (1997). Evidence for rapid face recognition from human scalp and intracranial electrodes. *Neuroreport, 8,* 2749-2754.

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

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

Tanaka, J. W. (2001). The entry point of face recognition: evidence for face expertise. *J.Exp.Psychol.Gen., 130,* 534-543.

Tanaka, J. W. & Farah, M. J. (1993). Parts and wholes in face recognition. *Q.J.Exp.Psychol.A, 46,* 225-245.

Tanaka, J. W. & Gauthier, I. (1997). Expertise in Object and Face Recognition. *The Psychology of Learning and Motivation, 36,* 83-125.

Tanaka, J. W. & Sengco, J. A. (1997). Features and their configuration in face recognition. *Mem.Cognit., 25,* 583-592.

Tarr, M. J. & Gauthier, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nat.Neurosci., 3,* 764-769.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381,* 520-522.

Thorpe, S. & Imbert, M. (1989). Biological Constraints on Connectionist Modeling. In R.Pfeifer, Z. Schreter, F. Fogelman-Soulie, & L. Steels (Eds.), *Connectionism in Perspective* (pp. 63-92). Amsterdam: Elsevier.

Tong, F. & Nakayama, K. (1999). Robust representations for faces: evidence from visual search. *J.Exp.Psychol.Hum.Percept.Perform., 25,* 1016-1035.

Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., & Kanwisher, N. (2000). Response Properties of the Human Fusiform Face Area. *Cognitive Neuropsychology, 17,* 257-279.

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

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

VanRullen, R. & Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *J.Cogn Neurosci., 13,* 454-461.

Wada, Y. & Yamamoto, T. (2001). Selective impairment of facial recognition due to a haematoma restricted to the right fusiform and lateral occipital region. *J.Neurol.Neurosurg.Psychiatry, 71,* 254-257.

Williams, G. V., Rolls, E. T., Leonard, C. M., & Stern, C. (1993). Neuronal responses in the ventral striatum of the behaving macaque. *Behav.Brain Res., 55,* 243-252.

Wilson, F. A. & Goldman-Rakic, P. S. (1994). Viewing preferences of rhesus monkeys related to memory for complex pictures, colours and faces. *Behav.Brain Res., 60,* 79-89.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81,* 141-145.

Young, A. W. (1992). Face recognition impairments. *Philos.Trans.R.Soc.Lond B Biol.Sci., 335,* 47-53.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16,* 747-759.

Chapter 2

The Selectivity of the Occipitotemporal M170 for Faces

Abstract

Evidence from fMRI, ERPs, and intracranial recordings suggests the existence of face-specific mechanisms in the primate occipitotemporal cortex. The present study used a 64-channel MEG system to monitor neural activity while normal subjects viewed a sequence of grayscale photographs of a variety of unfamiliar faces and non-face stimuli. In fourteen out of fifteen subjects, face stimuli evoked a larger response than non-face stimuli at a latency of 160 msec after stimulus onset at bilateral occipitotemporal sensors. Inverted face stimuli elicited responses that were no different in amplitude but 13 msec later in latency than upright faces. The profile of this M170 response across stimulus conditions is largely consistent with prior results using scalp and subdural ERPs.

2.1 Introduction

Extensive evidence from a wide variety of techniques suggests the existence of face-specific mechanisms in primate occipitotemporal cortex. The aim of this study was to provide a detailed characterization of the neural response to face stimuli using MEG.

Behavioral evidence from normal and brain-damaged subjects has suggested a functional dissociation between face and nonface processing. The strongest evidence comes from the double dissociation between face and object recognition, with prosopagnostic patients impaired at face but not object recognition (De Renzi, 1986), and other patients showing the opposite pattern of deficit (Moscovitch, Winocur, & Behrmann, 1997).

Many techniques have been used to explore face-processing mechanisms. Functional brain imaging studies have localized a focal region in the fusiform gyrus called the fusiform face area (FFA) that responds in a highly selective fashion to faces, compared to a wide variety of other stimulus types (Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997; Sergent, Ohta, & MacDonald, 1992; Haxby et al., 1999). However, fMRI provides little information about the temporal characteristics of face processing. Electrical recordings from the scalp surface have revealed a posterior-lateral negative peak at a latency of 170 msec elicited by human faces but not by animal faces, cars, scrambled faces, items of furniture, or human hands (Jeffreys, 1996; George, Evans, Fiori, Davidoff, & Renault, 1996; Bentin, Allison, Puce, Perez, & McCarthy, 1996). However, the poor spatial resolution of ERPs prevents precise localization of the neural source(s) of the N170. In contrast, intracranial recording has provided impressive evidence for selective neural responses to faces, with both high temporal and spatial resolution. Specifically, multiple distinct regions in the temporal lobes and hippocampus of epilepsy patients have been found to produce an N200 response to faces but not to cars, butterflies and scrambled faces, or letterstrings (Allison et al., 1994; Allison, Puce, Spencer, & McCarthy, 1999; Fried, MacDonald, & Wilson, 1997; Seeck et al., 1997). Nevertheless, these recordings are possible only from severely epileptic patients, where the degree of cortical re-organization caused by the seizures is not known.

In contrast to the techniques described, MEG provides excellent temporal resolution, good spatial resolution, and can be used safely in neurologically normal subjects. Several recent studies have found a strong magnetic response (M170) to face stimuli in comparison to nonface stimuli over occipitotemporal brain regions (Lu et al., 1991; Sams, Hietanen, Hari, Ilmoniemi, & Lounasmaa, 1997; Linkenkaer-Hansen et al., 1998; Swithenby et al., 1998; Streit et al., 1999; Watanabe, Kakigi, Koyama, & Kirino, 1999). These studies suggest that the M170 is quite selective for faces; however, only a few stimulus conditions were compared in each study. To provide a stronger test of face selectivity, the present study tested the amplitude and latency of the M170 to 13 different stimulus types. In Experiment 1, we compared the magnetic response elicited by faces and a variety of non-face images, in order to test whether the M170 is in fact specific to face processing as opposed to a more general process such as subordinate-level categorization, or processing of anything animate or human. In Experiment 2, we tested the generality of the M170 response across faces that varied in format, surface detail, and viewpoint. In Experiment 3, we tested whether the M170 is sensitive to stimulus inversion.

Three critical design features were used in the present study. First, we ran all subjects on a "localizer" experiment with face, object and hand stimuli in order to identify candidate face-selective sensors for each subject on a data set independent from the data collected in the main experiments. Second, in the three main experiments subjects performed a one-back task (pressing a button whenever two identical images were repeated consecutively), which obligated them to attend to all stimuli regardless of inherent interest. Finally, all stimulus classes in each experiment were interleaved in a random order to eliminate any effects of stimulus predictability.

2.2 Results

The most face-selective sensor (i.e. the one showing the greatest increase in response to faces compared to hands and objects) was identified independently for each subject and hemisphere from an inspection of the data from the localizer experiment (see Figure1, top, for an example). The independent definition of our sensor of interest (SOI) allowed us to objectively characterize the response properties of the M170 in the three following experiments which were run on the same subjects in the same session. Figure1 (top) shows the MEG response in each channel to faces and objects in a typical subject, with the face-selective SOI in each hemisphere indicated. Figure 1 (below) shows the magnetic responses in the SOIs from the left and right hemisphere for this subject. Only one subject's data did not show any clear face-selective SOI; this subject was excluded from further analyses. In all other subjects, a clear face-selective SOI was found in the ventral occipitotemporal region of each hemisphere.

The response to each stimulus type for each sensor of interest was averaged across the subjects in each experiment and is shown in Figure 2. For each subject individually, the amplitude and latency of the M170 was determined for each stimulus in each hemisphere. These values were then analyzed in six different ANOVAs (three experiments and two dependent measures, amplitude and latency), with hemisphere and stimulus condition as factors in each. All six ANOVAs found main effects of stimulus condition (all *p*s<0.02), but the main effects of hemisphere did not reach significance (all *p*s >0.05). Because there was no hint of an interaction of condition by hemisphere in any of six ANOVAs (all Fs<1), in subsequent analyses the data from the left and right hemisphere were averaged within each subject. The averages across subjects of each individual subject's M170 amplitude and latency for each condition are shown in Figure 3.

In Experiment 1, the amplitude of the M170 was significantly larger for faces than for animals (t[13]=7.69, *p*<0.0001), human hands (t[13]=5.72, *p*<0.0001), houses (t[13]=5.18, *p*<0.0001) and common objects (t[13]=8.34, *p*<0.0001). In addition, the M170 latency was significant later (by 9 msec on average) to animals than to human faces (t[13]=3.78, *p*<0.001).

For Experiment 2, all face stimuli produced a significantly larger response than the response to objects (all *p*s<0.005), except for cartoon faces where this difference did not reach significance (t[6]=1.84, *p*>0.05). The amplitude of the M170 elicited by front-view human faces was significantly larger than that for profile faces (t[6]=2.6, *p*<0.05) and cartoon faces (t[6]=4.93, *p*<0.001), but not significantly different from cat faces or line-drawing faces (both *p*s > 0.2). In addition, the M170 latency was significant later to cat face than to human faces (t[6]=4.48, *p*<0.001); however, the latencies for line-drawing and profile faces did not differ from that for human front-view faces (all *p*s>0.1).

In the third experiment, the M170 latency was significant later (13 msec on average) to inverted faces than to upright ones (t[13]=8.99, P<0.0001), but no significant difference was revealed in amplitude (t[13]=0.47, P>0.1). In addition, two-tone Mooney faces failed to elicit as large an M170 as human faces did (t[13]=6.07, P<0.0001).

2.3 Discussion

The main results of this study can be summarized as follows. A clear and bilateral M170 response to faces was found at occipitotemporal sites in 14 out of 15 subjects tested. Neither animal stimuli nor human hands elicited an M170 as large as that elicited by faces, showing that the M170 is selective for faces, not for human or animal forms. Further, because the M170 response was low to houses and hands yet the task required subjects to discriminate between exemplars of these categories, our data suggest that the M170 does not simply reflect subordinate-level categorization for any stimulus class. Experiment 2 found that the M170 was not significantly lower in amplitude for cat faces and outline faces than for grayscale human front-view faces, demonstrating that the M170 generalizes across face stimuli with very different low-level features. On the other hand, the longer latency of the M170 elicited by cat faces and the lower amplitude of the

M170 elicited by profile-view human faces suggest that any deviation from the configuration of human front-view faces reduces the efficiency of the processing underlying this response. In Experiment 3, the M170 to inverted faces was as large as that to upright faces, but it was delayed 13 msec in latency.

Our results are generally consistent with prior studies of the M170 (see also Halgren et al. (2000)) except that where we find bilateral face-specific M170s, other studies have found the M170 to be larger (Watanabe et al., 1999) or exclusively located (Swithenby et al., 1998) in the right hemisphere. The most direct and extensive investigation of face-specific neural responses have been carried out using subdural electrode recordings from the surface of the human ventral pathway (Allison et al., 1999; McCarthy, Puce, Belger, & Allison, 1999; Puce, Allison, & McCarthy, 1999). These studies have included most of the stimulus conditions tested in the present study. The response profile we observed for the M170 and the bilaterality of the M170 response are both consistent with the results from direct electrical recordings reported by Allison and colleagues. Our results are also consistent with prior studies using scalp ERPs, except that animal faces did not produce an N170 (Bentin et al., 1996) but they did produce an M170 in the present study.

Although the response properties of the M170 are similar to those of the FFA observed with fMRI in most respects, there are several apparent differences. First, M170 responses were bilateral and if any thing larger over the left hemisphere, whereas the FFA is typically larger in the right hemisphere. Second, our experiment showed that the M170 elicited by cartoon and Mooney faces was more like that elicited by common objects than by human faces. These results differ from the pattern of results found for the

FFA using fMRI (Tong, Nakayama, Moscovitch, Weinrib, & Kanwisher, 2000). These differences suggest that the M170 may reflect processing that occurs not only in the FFA but also in other neural sites.

Does the M170 reflect face detection or face recognition, or both? Behavioral studies have shown that surface information is critical in face recognition (Davies, Ellis, & Shepherd, 1978). However in our MEG study, linedrawing faces elicited as large a magnetic response as did grayscale faces. Furthermore, inversion of faces did not reduce the amplitude of M170, although face recognition performance is greatly reduced by inversion (Farah, 1995). These considerations suggest that the M170 may be engaged in detecting the presence of faces, instead of extracting the critical stimulus information necessary for face recognition.

2.4 Conclusions

Our results strongly suggest that the M170 response is tuned to the broad stimulus category of faces. The face-selective M170 is similar in many respects to the N170 and N200 observed with scalp and subdural ERPs (Bentin et al., 1996; Allison et al., 1999; McCarthy et al., 1999; Puce et al., 1999). Our study lays the groundwork for future MEG investigations into the number and locus of neural sources that generate the M170. Further, the evidence provided here for the selectivity of the M170 enables us to use the M170 as a marker of face processing in future work.

2.5 Methods

Seventeen healthy normal adults, aged 19-40, volunteered or participated for payment in all four experiments in a single testing session. All were right-handed and reported normal or corrected-to-normal vision. The data from two subjects were omitted from further analysis because they fell asleep during the experiment.

Subjects lay on the scanner bed in a dimly lit, sound attenuated, and magnetically shielded room, with a response button under their right hands. A mirror was placed 120 cm in front of the subject's eyes and the screen center was positioned on the subject's horizontal line of sight. The stimuli consisted of gray-scale photographs (256 levels) of a variety of unfamiliar faces and non-face stimulus categories. Each image subtended 5.7 × 5.7 degrees of visual angle and was presented at the center of gaze for a duration of 200 msec by a projector. The onset-to-onset interstimulus interval between stimuli was randomized from 600 to 1000 msec and stimuli were presented in a pseudorandom order. During the experiment, a small fixation cross was continuously present at the screen center.

The experiment consisted of 8 experimental blocks, divided into 4 experiments (the localizer experiment, plus Experiments 1-3). Each subject was first run on the localizer experiment which involved passively viewing 200 trials each of faces, objects and hands (intermixed). In this experiment, subjects were simply instructed to attentively view the sequence of images. In the following three experiments, subjects performed a one-back task in which they were asked to press a button whenever two consecutive images were identical. In each of the three main experiments, subjects performed 110 trials for each of 5 or 6 different stimulus categories. Only 7 subjects participated in Experiment 2. On average, 10% of trials were repetition targets; these were excluded from the analysis.

The magnetic brain activity was digitized continuously (1000 Hz sampling rate with 1Hz high pass and 200Hz low-pass cutoff, and 60 Hz notch) from a 64-channel whole head system with SQUID based first-order gradiometer sensors (KIT MEG SYSTEM). Five hundred millisecond epochs (100 msec pre-stimulus baseline and 400 msec post-stimulus) were acquired for each stimulus. All 200 trials (localizer experiment) or 100 trials (the following three experiments) of each type were averaged together, separately for each sensor, stimulus category, and subject.

2.6 Acknowledgements

This work was supported by grants to N.K. from Human Frontiers, NIMH (56037), and the Charles E. Reed Faculty Initiatives Fund. We thank Alison Harris for preparing the line drawing faces, John Kanwisher for assistance with the optics, and Anders Dale for discussions of the research.

Reference List

Allison, T., Ginter, H., McCarthy, G., Nobre, A. C., Puce, A., Luby, M. et al. (1994). Face recognition in human extrastriate cortex. *J.Neurophysiol., 71,* 821-825.

Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb.Cortex, 9,* 415-430.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perceptions in humans. *Journal of Cognitive Neuroscience, 8,* 551-565.

Davies, G. M., Ellis, H. D., & Shepherd, J. W. (1978). Face recognition accuracy as a function of mode of presentation. *J.Appl.Psychol., 63,* 180-187.

De Renzi, E. (1986). Current issues in prosopagnosia. In H.D.Ellis, M. A. Jeeves, F. Newcombe, & A. Young (Eds.), *Aspects of face processing* (pp. 153-252). Dordrecht: Martinus Nijhoff.

Farah, M. J. (1995). Dissociable system for visual recognition: A cognitive neuropsychology approach. In S.M.Kosslyn & D. N. Osherson (Eds.), *Visual Cognition* (pp. 101-119). Cambridge: MIT Press.

Fried, I., MacDonald, K. A., & Wilson, C. L. (1997). Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron, 18,* 753-765.

George, N., Evans, J., Fiori, N., Davidoff, J., & Renault, B. (1996). Brain events related to normal and moderately scrambled faces. *Brain Res.Cogn Brain Res., 4,* 65-76.

Halgren, E., Raij, T., Marinkovic, K., Jousmaki, V., & Hari, R. (2000). Cognitive response profile of the human fusiform face area as determined by MEG. *Cereb.Cortex, 10,* 69-81.

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

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

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

Linkenkaer-Hansen, K., Palva, J. M., Sams, M., Hietanen, J. K., Aronen, H. J., & Ilmoniemi, R. J. (1998). Face-selective processing in human extrastriate cortex around 120 ms after stimulus onset revealed by magneto- and electroencephalography. *Neurosci.Lett., 253,* 147-150.

Lu, S. T., Hamalainen, M. S., Hari, R., Ilmoniemi, R. J., Lounasmaa, O. V., Sams, M. et al. (1991). Seeing faces activates three separate areas outside the occipital visual cortex in man. *Neuroscience, 43,* 287-290.

McCarthy, G., Puce, A., Belger, A., & Allison, T. (1999). Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex. *Cereb.Cortex, 9,* 431-444.

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-Specific Processing in the Human Fusiform Gyrus. *Journal of Cognitive Neuroscience, 9,* 604-609.

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience, 9,* 555-604.

Puce, A., Allison, T., & McCarthy, G. (1999). Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials. *Cereb.Cortex, 9,* 445-458.

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

Seeck, M., Michel, C. M., Mainwaring, N., Cosgrove, R., Blume, H., Ives, J. et al. (1997). Evidence for rapid face recognition from human scalp and intracranial electrodes. *Neuroreport, 8,* 2749-2754.

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

Streit, M., Ioannides, A. A., Liu, L., Wolwer, W., Dammers, J., Gross, J. et al. (1999). Neurophysiological correlates of the recognition of facial expressions of emotion as revealed by magnetoencephalography. *Brain Res.Cogn Brain Res., 7,* 481-491.

Swithenby, S. J., Bailey, A. J., Brautigam, S., Josephs, O. E., Jousmaki, V., & Tesche, C. D. (1998). Neural processing of human faces: a magnetoencephalographic study. *Exp.Brain Res., 118,* 501-510.

Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., & Kanwisher, N. (2000). Response Properties of the Human Fusiform Face Area. *Cognitive Neuropsychology, 17,* 257-279.

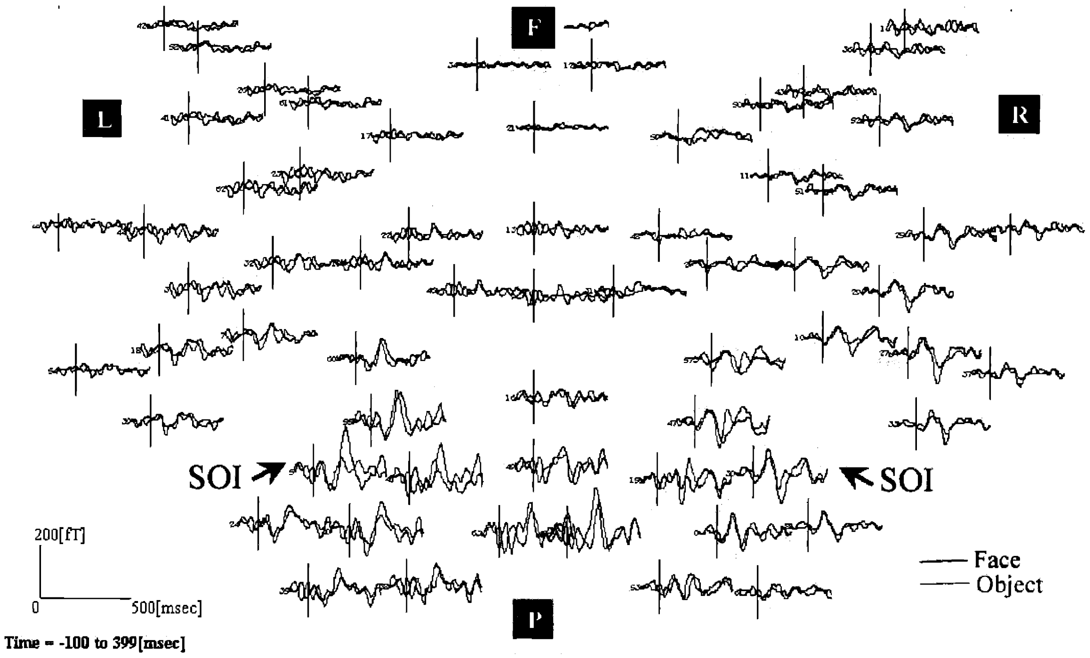
Watanabe, S., Kakigi, R., Koyama, S., & Kirino, E. (1999). Human face perception traced by magneto- and electro-encephalography. *Brain Res.Cogn Brain Res., 8,* 125-142.

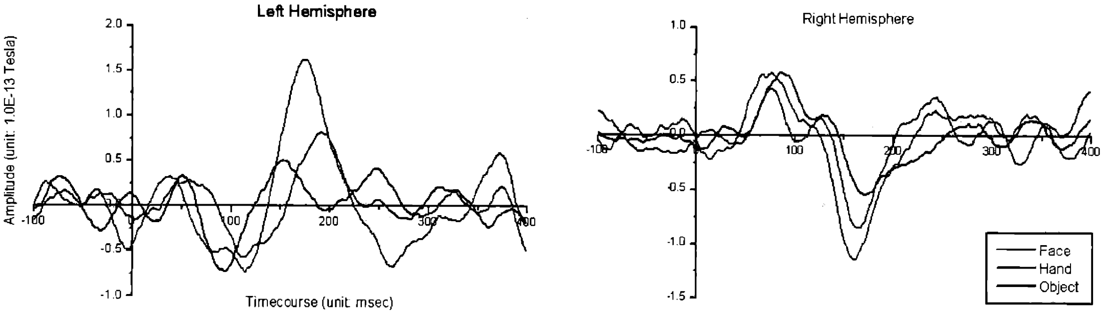
Figure Legends

Figure 1: (Top) The average response of each of 64 channels elicited by faces (black waveform) and objects (red waveform) from a typical subject in the localizer experiment. As can be seen, at least one sensor in each hemisphere shows a much stronger response to faces than objects; these were selected as the SOIs for analyses of subsequent experiments in the same subject. (L: Left; R: Right; F: Frontal; P: Posterior). (Bottom) The response to faces (Red), hands (green), and objects (blue) at these two SOIs in the localizer experiment are shown below.

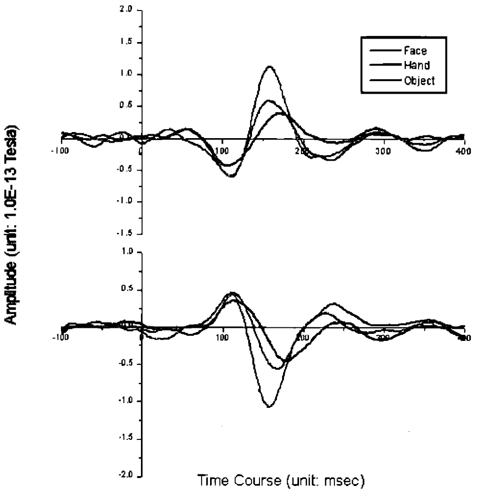
Figure 2: The M170 response from SOIs in the left (top) and right (bottom) hemispheres averaged across subjects for each experiment.

Figure 3: The average amplitudes and latencies for each condition from three main experiments.

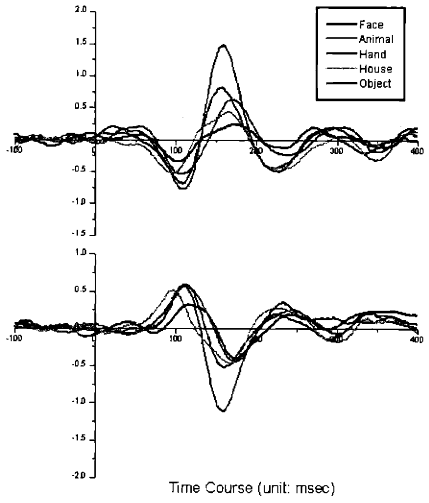

F
L
R
SOI
SOI
200[fT]
0
500[msec]
Face
Object
P
Time = -100 to 399[msec]


Left Hemisphere
Amplitude (unit: 1.0E-13 Tesla)
2.0
1.5
1.0
0.5
0
-0.5
-1.0
-100
100
200
300
400
Timecourse (unit: msec)
Right Hemisphere
1.0
0.5
0.0
-0.5
-1.0
-1.5
-100
100
200
300
400
Face
Hand
Object

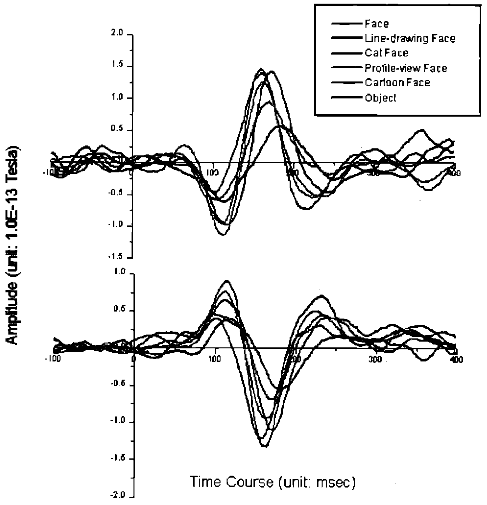
Localizer Experiment



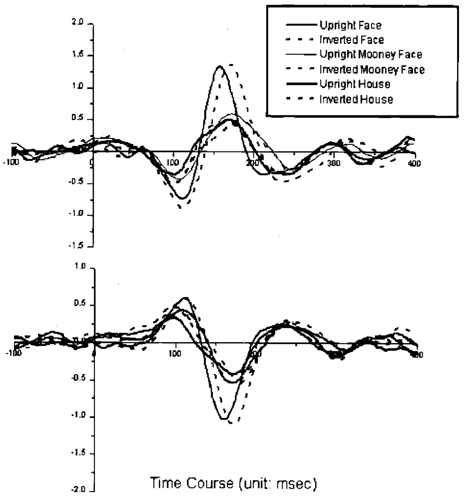
Experiment 1



Experiment 2

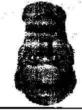


Experiment 3



		Face	Animal	Hand	House	Object
						
Exp 1	Amplitude	1.44	0.72	0.96	0.75	0.63
	Latency	161	170	167	158	167

		Face	Line drawing	Cat face	Profile	Cartoon face	Object
							
Exp 2	Amplitude	1.50	1.42	1.36	1.21	0.91	0.72
	Latency	165	163	178	166	173	185

		Face	Inverted face	Mooney	Inverted Mooney	House	Inverted house
							
Exp 3	Amplitude	1.36	1.41	0.84	0.78	0.81	0.77
	Latency	159	172	171	173	159	161

(Amplitude unit: 1.0E-13 Tesla; Latency unit: msec)

Chapter 3

Stages of processing in face perception: An MEG study

Abstract

Magnetoencephalography (MEG) was used to investigate stages of processing in face perception. We found a face-selective MEG response occurring only 100 ms after stimulus onset (the "M100"), 70 ms earlier than previously reported. Further, the M100 response amplitude was correlated with successful categorization of stimuli as faces, but not with successful recognition of individual faces, whereas the previously-described face-selective "M170" response was correlated with both processes. These data suggest that face processing proceeds through two stages: an initial stage of face categorization, and a later stage at which the identity of individual faces is extracted.

3.1 Introduction

Face recognition is one of the most important problems our visual system must solve. Here we used magnetoencephalography (MEG) in an effort to characterize the sequence of cognitive and neural processes underlying this remarkable ability.

Two candidate stages of face processing are the categorization of a stimulus as a face, and the identification of a specific individual. Several studies of object recognition suggest that objects are first categorized at a "basic level" (e.g., dog, bird), before a finer "subordinate level" identification is achieved (e.g., poodle, sparrow)(Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Jolicoeur, Gluck, & Kosslyn, 1984). Evidence that a similar sequence may occur in face perception comes from single-unit recordings in macaques showing that the initial transient response of face-selective neurons in inferotemporal cortex reflects a rough categorization of face versus nonface, whereas subsequent firing of the same neural population represents finer information such as the facial expression or identity of the person (Sugase, Yamane, Ueno, & Kawano, 1999). However, it has been argued that visual expertise in discriminating exemplars of a specific visual category may shift the point of initial contact with memory representations from the basic level to the subordinate level (Rosch et al., 1976; Tanaka & Taylor, 1991). Given that we are all experts at face recognition, this hypothesis predicts that subjects should be able to identify an individual face as fast as they can tell that it is a face at all. Although some behavioral evidence is consistent with this hypothesis(Tanaka, 2001), other evidence is not (Grill-Spector & Kanwisher, 2001).

MEG is an ideal technique for addressing these questions, as its high temporal resolution enables us to tease apart processing stages that may occur within tens of milliseconds of each other. Prior work with MEG and event-related potentials (ERPs) has characterized a response component called the N170 (or "M170" in MEG) occurring around 170 ms after stimulus onset that is about twice as large for face stimuli as for a variety of control nonface stimuli such as hands, houses, or animals (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Jeffreys, 1996; Sams, Hietanen, Hari, Ilmoniemi, & Lounasmaa, 1997; Liu, Higuchi, Marantz, & Kanwisher, 2000). This response is thought to reflect the "structural encoding" of a face (Bruce & Young, 1986; Bentin et al., 1996; Eimer, 2000), that is, the extraction of a perceptual representation of the face.

Although several reports of even earlier category-selective responses have been published (Seeck et al., 1997; Schendan, Ganis, & Kutas, 1998; Mouchetant-Rostaing, Giard, Bentin, Aguera, & Pernier, 2000; Kawasaki et al., 2001; Braeutigam, Bailey, & Swithenby, 2001; Streit, Wolwer, Brinkmeyer, Ihl, & Gaebel, 2000), they are open to explanations in terms of nonspecific repetition effects (George, Jemel, Fiori, & Renault, 1997) or differences in the low-level features present in the stimuli (VanRullen & Thorpe, 2001). Here we asked when the first truly face-selective responses occur. To do this, we recorded MEG responses while subjects passively viewed a sequence of photographs of faces and a variety of control stimuli (Experiments 1 and 2). These experiments found a new face-selective response occurring only 100 ms after stimulus onset (the "M100"), generated from extrastriate cortex (Experiment 3). In Experiment 4, we tested whether the M100 and M170 amplitudes are correlated with success in face categorization and/or face identification. Finally, we tested for further qualitative differences in the processes underlying the M100 and the M170 by measuring the responses of each component to face configurations and face parts.

3.2 Results

3.2.1 A Face-Selective Response at a Latency of 100 ms (the "M100")

An interpolated map of the t-values comparing the MEG response to faces versus houses for a typical subject in Experiment 1 is shown in **Fig. 1a**. As expected, this map shows the previously-described face-selective response (the "M170") occurring at a latency of about 160 ms post-stimulus onset (Liu et al., 2000); this response presumably corresponds approximately to the “N170” component in ERP studies (Bentin et al., 1996) and the “N200” in intracranial ERP studies (Allison, Puce, Spencer, & McCarthy, 1999).

In addition to the M170, we found a smaller response peaking at a mean latency of 105 ms (range: from 84.5 to 130.5 ms; SD = 16.1) that was significantly higher for faces than houses (the “M100”). This result was seen with the same polarity in 13 out of 15 subjects. The scalp distribution of the face-selective M100 response was slightly posterior to that of the M170, but the sensors showing the strongest face-selectivity for the M100 largely overlapped with those showing the strongest face-selectivity for the M170. The MEG response of a representative subject at a typical overlapping face-selective sensor in the right hemisphere is shown in **Fig. 1b**.

To provide a stronger test of face selectivity, the magnitude of the M100 response to a variety of control stimuli was measured in Experiment 2. Here, subjects were asked to press a button whenever two consecutive images were identical, obligating them to attend to all stimuli regardless of inherent interest. Target trials containing such repeated stimuli were excluded from the analysis. Accuracy on the 1-back matching task was high for all categories (>90% correct) except for hands (76%), which are visually very similar to each other. The MEG data from Experiment 1 were first used to define "sensors of interest" (SOIs) in each subject that produced significantly face-selective responses for both the M100 and the M170 (see Methods). Both the amplitudes and latencies of peak responses to the new stimulus conditions in Experiment 2 in these same SOIs were then quantified in the same subject in the same session. The M100 response to photographs of faces was greater (see **Fig. 2**) than to photographs of animals ($F_{1,12} = 10.2$; $P<0.01$), human hands ($F_{1,12} = 9.0$, $P<0.02$), houses ($F_{1,12} = 8.1$, $P< 0.02$), and nonface objects ($F_{1,12} = 10.3$, $P<0.01$). Therefore, the M100 is not selective more generally for anything animate, or for any human body parts; instead, it appears to be selective for faces. However, both the magnitude and selectivity of the M100 response were weaker than for the M170 response.

The earlier latency and somewhat more posterior scalp distribution of the M100 compared to the M170 suggest that the two components may not originate from the same anatomical source. In order to test whether the M100 might originate in retinotopically-organized visual cortex, we separately measured the amplitude of the M100 in each hemisphere to faces presented in the contralateral versus ipsilateral visual field (2.8° off fixation) in Experiment 3. This manipulation is known to affect responses in visual areas V1, V2, V3, VP, V3A and V4v in humans (Tootell, Mendola, Hadjikhani, Liu, & Dale, 1998). No difference was observed in the amplitude of the M100 in each hemisphere for contralaterally versus ipsilaterally presented faces ($F_{1,10}<1$), indicating that the source of this component must be beyond retinotopic cortex.

3.2.2 Decoupling categorization and identification

The results described so far implicate both the M100 and M170 in face processing, but do not indicate what aspect of face processing each component reflects. In Experiment 4, we asked for each component whether it was involved in the categorization of a stimulus as a face, or the extraction of the individual identity of a face, or both. Subjects were instructed to make two judgments about each stimulus, determining both its category (face or house) and its individual identity.

In this experiment, 10 subjects matched front-view test images of faces and houses to profile views (faces) or three-quarter views (houses) of the same stimulus set presented earlier in the same trial (**Fig. 3a**). There were three possible responses: Different Category (i.e., the sample was a face and the test was a house or vice versa); Different Individual but same category; and Same Individual. Correct categorization required discrimination between Different Category trials and either Different Individual or Same Individual trials; correct identification required distinguishing between Different Individual and Same Individual trials.

A set of face and house stimuli (5 exemplars each) were constructed, each of which had identical spatial frequency, luminance, and contrast spectra. Subjects were first trained to match each face and house with its profile or three-quarters view, respectively, at 100% accuracy. Using a technique similar to the recently proposed RISE technique (see Methods) (Sadr & Sinha, 2001; Sadr & Sinha, in press), each subject was then run on a psychophysical staircase procedure in which the percentage of phase coherence of each test stimulus was gradually reduced until the subject reached threshold performance on the matching task (75% correct, 20 steps by QUEST staircase (Watson & Pelli, 1983)). In this way, five threshold face stimuli and five threshold house stimuli were constructed for each subject for each of the two tasks (categorization and identification). Across all stimuli and subjects, the resulting threshold face and house stimuli had a mean percent phase coherence of 14% (face) and 18% (house) for the categorization task and 38% (face) and 51% (house) for the identification task, indicating that more stimulus information was necessary for the identification task than the categorization task, as expected.

Next, each subject performed the same matching task (Different Category, Different Individual, or Same Individual) in the MEG scanner, using face and house stimuli with phase coherence varied across four levels: 0%, 90%, and the two previously-derived sets of thresholds for that subject, one for the categorization task, and the other for the identification task (**Fig. 3b**). In addition, the original version of each image with unmodified spatial frequencies was included to localize face-selective SOIs. By measuring both categorization and identification performance on each trial, the task allowed us to decouple the MEG correlates of successful categorization from those of successful identification. To obtain the MEG correlates of successful categorization, we compared the average MEG response to the same test image when the subject correctly categorized but failed to identify it, versus when they categorized it incorrectly. For identification, we compared the response to the same test image when the subject correctly identified it versus when they incorrectly identified it but categorized it successfully.

MEG waveforms averaged across each subject's face-selective SOIs from the face categorization and identification tasks are shown in **Fig. 4a**. The magnitudes of both the M100 ($F_{1,9} = 9.5$, $P<0.02$) and the M170 ($F_{1,9} = 5.8$, $P<0.05$) were larger for successful than unsuccessful categorization of faces (**Fig. 4b**, top left). However, only the M170 ($F_{1,9} = 43.3$, $P<0.001$), but not the M100 ($F_{1,9} <1$), was higher for correct than incorrect identification of faces (interaction: $F_{1,9} = 8.7$, $P<0.02$) (**Fig. 4b**, bottom left). For house stimuli, neither the M100 nor the M170 differed for correct versus incorrect trials in either task (all $F_{1,9} < 1$) (**Fig. 4b**, top right and bottom right). The finding that the M170 is specific for face identification (not house identification) is further supported by a significant 3-way interaction ($F_{1,9} = 6.73$, $P<0.03$) of face vs. house identification, success vs. failure, and M100 vs. M170 (**Fig. 4b**, bottom left and right). In addition, neither the main effect of hemisphere nor the interaction of task by hemisphere was significant (both Fs<1).

Accuracy on categorization and identification tasks at the two levels of phase coherence (derived from the previous psychophysical measurements) is shown in Table 1. Pairwise t-tests revealed no significant difference in accuracy as a function of stimulus category (faces versus houses) for either task (categorization versus identifications) (all *P*s > 0.2). Therefore, any difference between faces and houses observed in MEG responses cannot be explained in term of differences in behavioral performance. Note that even when the stimuli were degraded to "categorization threshold level", the subjects' performance in the identification task was above chance (both Ps < 0.01).

In sum, both the M100 and M170 are correlated with successful face categorization, but only the later M170 component is correlated with successful face identification, indicating a dissociation between the two processes. One possible account of this dissociation is that the selectivity of the underlying neural population may continuously sharpen over time, permitting crude discriminations (e.g. between a face and a house) earlier, and more fine-grained discriminations (e.g., between two different faces) only later. Indeed, the ratio of the response to faces versus houses is lower for the M100 (1.6) than for the M170 (1.8, interaction $P<0.03$), showing that selectivity does indeed sharpen over time. However, this fact alone does not indicate whether the selectivity changes only in degree, or whether qualitatively different representations underlie the M100 and the M170. This question was addressed in Experiment 5, in which we measured the M100 and M170 responses to information about face configurations and face parts.

In this experiment, two face-like stimulus categories were constructed from veridical human faces by orthogonally eliminating or disrupting either face configuration or face parts (eyes, nose and mouth; **Fig. 5a**). Specifically, to remove the contribution of face parts, face parts in each stimulus were replaced by solid black ovals in their corresponding locations, preserving the face configuration ("Configuration" stimuli, see **Fig. 5a**). Conversely, for "Parts" stimuli, the face parts were kept intact but were rearranged into a novel nonface configuration (see **Fig. 5a**). We measured MEG responses to fourteen subjects who passively viewed these two sets of face-like stimuli (50 exemplars each) presented in a random order while fixating.

The responses to Configuration and Part stimuli recorded at independently defined face-selective sensors, averaged across subjects, are shown in **Fig. 5b**. Critically, we found a significant two-way interaction of M100 versus M170 by Configuration versus Parts stimuli, ($F_{1,13} = 13.4$, $P < 0.005$). This interaction reflects the fact that the amplitude of the M100 was significantly larger for Parts stimuli than for Configuration stimuli ($F_{1,13} = 11.5$, $P<0.005$), whereas a trend in the opposite direction was found for the M170, ($F_{1,13} = 3.35$, $P = 0.09$). This finding shows that it is not merely the degree of selectivity, but the qualitative nature of the selectivity, that differs between the M100 and the M170. Again, neither the main effect of hemisphere nor the interaction of stimulus type by hemisphere was significant (both Fs<1).

3.3 Discussion

In Experiments 1–3, we report an MEG response component occurring over occipitotemporal cortex and peaking at around 100 ms that is significantly larger for faces than for a variety of nonface objects. This result indicates that the categorization of a stimulus as a face begins within 100 ms after stimulus onset, substantially earlier than previously thought (Bentin et al., 1996; Thorpe, Fize, & Marlot, 1996; VanRullen et al., 2001).

Unlike prior reports of very early category-selective ERP responses (Seeck et al., 1997; Schendan et al., 1998; Mouchetant-Rostaing et al., 2000; Kawasaki et al., 2001; Braeutigam et al., 2001; Streit et al., 2000), the M100 reported here cannot be explained in terms of differences in the low-level features present in the stimuli. The M100 response was stronger when the *same face stimulus* was correctly perceived as a face, than when it was wrongly categorized as a nonface. This result shows that the M100 reflects the subject's percept rather than simply the low-level properties of the stimulus.

It is possible that a correlate of the face-selective M100 could be obtained with ERPs. However, because MEG is sensitive to a subset of the neural activity that can be detected with scalp ERPs (Hamalainen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993), there is no simple correspondence between MEG responses and scalp ERP responses, and selectivities that are clear in the MEG response may be diluted with ERPs. Similarly, the M170 response measured with MEG probably corresponds to only one of the two sources hypothesized to underlie the N170 response (Bentin et al., 1996; Sagiv & Bentin, 2001). On the other hand, direct correspondences may exist between the M100 and M170 and the more focal intracranial P150 and N200 ERPs (Allison et al., 1999), respectively, assuming that the later latencies in the intracranial case arise from medication and/or histories of epilepsy typical of that subject population. Unfortunately the limitations in current source localization techniques leave these correspondences only tentative at present.

The latency of the M100 response is not directly comparable to the category-selective response that occurs at a latency of 100 ms in IT neurons in macaques (Oram & Perrett, 1992; Rolls, 1984), because all cortical response latencies are shorter in macaques than humans. For example, V1 responses occur 40–60 ms after stimulus presentation in macaques (Thorpe & Fabre-Thorpe, 2001), about 20 msec earlier than that in humans (Thorpe et al., 1996; Gomez Gonzalez, Clark, Fan, Luck, & Hillyard, 1994).

Given that at least 60–80 ms are thought to be necessary for visual information to reach primary visual cortex in humans (Gomez Gonzalez et al., 1994), this leaves only an additional 20–40 ms for the first face-selective responses to be generated in cortex. Such latencies are hard to reconcile with models of visual categorization that rely heavily on iterative feedback loops and/or recurrent processing, and strengthen the claim that initial stimulus categorization is accomplished by largely feedforward mechanisms (Thorpe & Imbert, 1989).

The second major finding of this study is that both the M100 and the M170 are correlated with successful face categorization, but only the later M170 component is correlated with successful face identification. This finding indicates that processes critical for the identification of a face begin at a substantially later latency than processes critical for the categorization of a stimulus as a face. Evidently, our expertise with faces has not led us to be able to identify individual faces as fast as we can tell they are faces at all (as argued by Tanaka (2001)).

The dissociation we report here between the processes underlying face categorization and these underlying face identification do not simply reflect the greater difficulty of identification compared to categorization, because our results were obtained even when the stimuli were adjusted to produce identical performance in the categorization and identification tasks (Experiment 4). Further, the difference in the response for successful versus unsuccessful trials on face stimuli cannot be explained by general processes such as attention or association, because neither the M100 nor the M170 amplitude differed for correct versus incorrect trials on house stimuli. Thus, our data argue strongly that the categorization of a stimulus as a face begins substantially earlier than the identification of the particular face.

Are these two stages, categorization and identification, simply different points on a continuous spectrum of discrimination, with cruder discriminations occurring at earlier latencies and finer discriminations occurring later, perhaps as initially coarse neural population codes get sharpened over time (Kovacs, Vogels, & Orban, 1995; Keysers, Xiao, Foldiak, & Perrett, 2001)? Consistent with this hypothesis, the M170 shows stronger face selectivity than the M100. However, this hypothesis predicts that the rank ordering of preferred stimuli must be the same for the M100 and the M170. Contrary to this prediction, Experiment 5 found that the M100 shows a stronger response to stimuli depicting face parts than face configurations, whereas the M170 shows the opposite response profile. If neural populations simply sharpened the selectivity of their response over time, this preference reversal would not be found. Instead, our data suggest that qualitatively different information is extracted from faces at 100 ms versus 170 ms after stimulus onset. Finally, the observed change in response profile cannot be easily explained in terms of a progression from coarse or global information to fine or local information or a progression from less to more clear face features. Instead, information about relatively local face parts is more important in determining the M100 response, whereas information about relatively global face configurations is more important in the later M170 response. Thus the most natural account of our data is that the M100 and the

M170 reflect qualitatively distinct stages of face perception: an earlier stage that is critical for categorizing a stimulus as a face, and that relies more on information about face parts, and a later stage that is critical for identifying individual faces, and that relies more on information about face configurations.

Will the findings reported here hold for the recognition of stimulus classes other than faces? Given the numerous sources of evidence that faces are "special"(Farah, 1995), we cannot simply assume that they will. Unfortunately we cannot run experiments comparable to those reported here on other stimulus categories, because we have not found MEG markers selective for other categories. However, recent behavioral experiments suggest that the stages of processing reported here for face recognition will generalize to the visual recognition of nonface objects as well (Grill-Spector et al., 2001).

3.5 Methods

MEG Recordings for Experiments 1–3 were made using a 64-channel whole head system with SQUID based first-order gradiometer sensors (Kanazawa Institute of Technology MEG system at the KIT/MIT MEG Joint Research Lab on the MIT campus); Experiments 4 and 5 were run after the system was upgraded to 96 channels. Magnetic brain activity was digitized continuously at a sampling rate of 1000 Hz (500 Hz for Experiment 4) and was filtered with 1Hz high pass and 200Hz low-pass cutoff and a 60 Hz notch. Informed consent was obtained from all subjects, and the study was approved by MIT Committee on the Use of Humans as Experimental Subjects (COUHES).

Experiments 1–3: The face-selective M100 response

15 subjects (age 18–40) passively viewed 100 intermixed trials of faces and houses (50 exemplars each) in Experiment 1; two additional subjects' data were discarded due to self-reported sleepiness. The thirteen of the fifteen subjects who showed the early face preference over occipitotemporal cortex also performed a one-back task on faces and a variety of nonface objects (50 exemplars each) in Experiment 2. Each image subtended 5.7 × 5.7 degrees of visual angle and was presented at the center of gaze for 200 ms, followed by an 800 ms ISI. The design for Experiment 3 is described in the text.

For each subject in Experiment 1, t-tests were conducted between the MEG responses to faces and houses at each time point (from –100 to 400 msec; 500 time points) and each sensor (64 channels) separately. Sensors where the magnetic fields evoked by faces were significantly larger than those by houses ($Ps<0.05$) for at least 5 consecutive time points both within the time window centered at the latency of the M100 and within that of the M170 were defined as Sensors of Interest (SOIs). P levels for these SOI-defining statistics were not corrected for multiple sensors or multiple time point comparisons. All critical claims in this paper are based on analyses of the average responses over these sensors in independent data sets, and thus require no correction for multiple spatial hypotheses.

The peak amplitude of the M100 (maximum deflection) was determined for each stimulus type in each hemisphere within a specified time window (width > 40 msec) for each subject individually. Because there was no main effect of hemisphere ($P>0.05$) and no interaction of condition by hemisphere (F<1), in subsequent analyses the data from the left and right hemisphere were averaged within each subject (after flipping the sign of the data from the right hemisphere to match the polarities).

Experiment 4: Categorization versus Identification

10 subjects (age 22–32) participated in Experiment 4. The MEG recordings were preceded by a training session (<10 min) and then a psychophysical staircase adjustment session conducted in the MEG room. MEG data from two additional subjects were discarded, one due to performance that was more than two standard deviations below the mean, the other due to polarities of both M100 and M170 that were reversed compared to all other subjects (though inclusion the data from this subject did not change the pattern or significance of the results).

Noise images were generated by inverse Fourier transformation of the mean amplitude spectra with randomized phase spectra (Sadr et al., 2001; Rainer, Augath, Trinath, & Logothetis, 2001; Sadr et al., in press). Intermediate images containing x% phase spectra of original images and (100–x)% random phase spectra were generated using linear interpolation (phase spectra levels of 0%, and 90% along with categorization and identification thresholds). This procedure ensured that all images were equated for spatial frequency, luminance and contrast.

Analyses were performed on only the subset of data for which subjects responded both correctly and incorrectly to an identical stimulus. That is, for each stimulus, equal numbers of successful and unsuccessful trials were chosen, and the extra trials were omitted from the analysis from whichever condition had more trials. In particular, because the number of correct trials was larger than incorrect trials, each incorrect trial was paired with the temporally closest correct trial. This analysis was conducted for each stimulus, guaranteeing that average MEG responses on successful and unsuccessful trials were derived from identical stimuli. Finally, the MEG recordings were averaged across stimuli separately for successful and unsuccessful trials. Note that the trials used to generate the waveform for face categorization were only selected from the MEG responses to those stimuli degraded to "categorization thresholds", and the trials used to generate the waveform for face identification were only selected from the MEG responses to those stimuli degraded to "identification thresholds". The same held for house categorization and identification. The exact number of success and failure trials for each task varied across subjects, but ranged from 20 to 30 trials each for successful and unsuccessful categorization and from 15 to 20 trials for successful and unsuccessful identification.

Experiment 5: Face Configuration versus Face Parts

Two stimulus categories were constructed from veridical faces. In Configuration stimuli, face parts in each veridical face were replaced by solid black ovals in their corresponding locations, whereas for Parts stimuli the face parts were kept intact but were rearranged into a novel nonface configuration. The size of the black ovals in the Configuration stimuli was matched to the actual size of corresponding face parts, and the arrangements of nonface configurations varied across all exemplars of Parts stimuli. Photographs of Configuration and Parts stimuli were presented at the center of gaze for 200ms with an 800 ms ISI. Fourteen subjects (age 18–41) passively viewed 100 intermixed trials of each stimulus category (50 exemplars each); three additional subjects' data were discarded due to lack of a face-selective MEG response in the independent localizer scan.

3.5 Acknowledgements

We thank J. Sadr and P. Sinha for helpful discussions on their RISE technique, M. Eimer, S. Hillyard, A. Marantz, M. Valdes-Sosa, P. Downing, W. Freiwald, K. Grill-Spector, Y. Jiang and the rest of Kanwisher Lab for comments on the manuscript. Supported by the Reed fund and NEI grant (EY13455) to N.K.

Reference List

Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb.Cortex, 9,* 415-430.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perceptions in humans. *Journal of Cognitive Neuroscience, 8,* 551-565.

Braeutigam, S., Bailey, A. J., & Swithenby, S. J. (2001). Task-dependent early latency (30-60 ms) visual processing of human faces and other objects. *Neuroreport, 12,* 1531-1536.

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

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

Farah, M. J. (1995). Dissociable system for visual recognition: A cognitive neuropsychology approach. In S.M.Kosslyn & D. N. Osherson (Eds.), *Visual Cognition* (pp. 101-119). Cambridge: MIT Press.

George, N., Jemel, B., Fiori, N., & Renault, B. (1997). Face and shape repetition effects in humans: a spatio-temporal ERP study. *Neuroreport, 8,* 1417-1423.

Gomez Gonzalez, C. M., Clark, V. P., Fan, S., Luck, S. J., & Hillyard, S. A. (1994). Sources of attention-sensitive visual event-related potentials. *Brain Topogr., 7,* 41-51.

Grill-Spector, K. & Kanwisher, N. (2001). Common cortical mechanisms for different components of visual object recognition: A combined behavioral and fMRI Study [abstract]. *Journal of Vision, 1,* 474a.

Hamalainen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography -- theory, instrumentation, and applications to noninvasive studies of the working human brain. *Review of Modern Physics, 65,* 413-497.

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

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: making the connection. *Cognit.Psychol., 16,* 243-275.

Kawasaki, H., Kaufman, O., Damasio, H., Damasio, A. R., Granner, M., Bakken, H. et al. (2001). Single-neuron responses to emotional visual stimuli recorded in human ventral prefrontal cortex. *Nat.Neurosci., 4,* 15-16.

Keysers, C., Xiao, D. K., Foldiak, P., & Perrett, D. I. (2001). The speed of sight. *J.Cogn Neurosci., 13,* 90-101.

Kovacs, G., Vogels, R., & Orban, G. A. (1995). Cortical correlate of pattern backward masking. *Proc.Natl.Acad.Sci.U.S.A, 92,* 5587-5591.

Liu, J., Higuchi, M., Marantz, A., & Kanwisher, N. (2000). The selectivity of the occipitotemporal M170 for faces. *Neuroreport, 11,* 337-341.

Mouchetant-Rostaing, Y., Giard, M. H., Bentin, S., Aguera, P. E., & Pernier, J. (2000). Neurophysiological correlates of face gender processing in humans. *Eur.J.Neurosci., 12,* 303-310.

Oram, M. W. & Perrett, D. I. (1992). Time course of neural responses discriminating different views of the face and head. *J.Neurophysiol., 68,* 70-84.

Rainer, G., Augath, M., Trinath, T., & Logothetis, N. K. (2001). Nonmonotonic noise tuning of BOLD fMRI signal to natural images in the visual cortex of the anesthetized monkey. *Curr.Biol., 11,* 846-854.

Rolls, E. T. (1984). Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. *Hum.Neurobiol., 3,* 209-222.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382-349.

Sadr, J. & Sinha, P. (2001). Exploring object perception with Random Image Structure Evolution. *MIT Artificial Intelligence Laboratory Memo, 2001-6, (2001)., 2001.*

Sadr, J. & Sinha, P. Object recognition and random image structure evolution. *Cognit.Sci.,* (in press).

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

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

Schendan, H. E., Ganis, G., & Kutas, M. (1998). Neurophysiological evidence for visual perceptual categorization of words and faces within 150 ms. *Psychophysiology, 35,* 240-251.

Seeck, M., Michel, C. M., Mainwaring, N., Cosgrove, R., Blume, H., Ives, J. et al. (1997). Evidence for rapid face recognition from human scalp and intracranial electrodes. *Neuroreport, 8,* 2749-2754.

Streit, M., Wolwer, W., Brinkmeyer, J., Ihl, R., & Gaebel, W. (2000). Electrophysiological correlates of emotional and structural face processing in humans. *Neurosci.Lett., 278,* 13-16.

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

Tanaka, J. W. (2001). The entry point of face recognition: evidence for face expertise. *J.Exp.Psychol.Gen., 130,* 534-543.

Tanaka, J. W. & Taylor, M. (1991). Object categories and expertise: is the basic level in the eye of the beholder? *Cognitive Psychology, 23,* 457-482.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381,* 520-522.

Thorpe, S. & Imbert, M. (1989). Biological Constraints on Connectionist Modeling. In R.Pfeifer, Z. Schreter, F. Fogelman-Soulie, & L. Steels (Eds.), *Connectionism in Perspective* (pp. 63-92). Amsterdam: Elsevier.

Thorpe, S. J. & Fabre-Thorpe, M. (2001). Neuroscience. Seeking categories in the brain. *Science, 291,* 260-263.

Tootell, R. B., Mendola, J. D., Hadjikhani, N. K., Liu, A. K., & Dale, A. M. (1998). The representation of the ipsilateral visual field in human cerebral cortex. *Proc.Natl.Acad.Sci.U.S.A, 95,* 818-824.

VanRullen, R. & Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *J.Cogn Neurosci., 13,* 454-461.

Watson, A. B. & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33,* 113-120.

Figure Legends

Figure 1: MEG data from a typical subject. (**a**) Pair-wise t-tests between the responses at each sensor reveal an early (M100) and late (M170) significant difference in the MEG response to faces versus houses over occipitotemporal cortex. (**b**) The MEG waveforms are averaged across all face and house trials at a typical sensor of interest in the right hemisphere. Red: faces. Blue: houses. Black: t value. The left vertical scale indicates the amplitude of the MEG response (10^{-13} Tesla) whereas the right one shows the t-value. A value $t = 1.99$ (green) corresponds to $p<0.05$ (uncorrected for comparisons at multiple time points).

Figure 2: Amplitudes of the peak M100 response, averaged across subjects, to faces and a variety of nonface objects at predefined sensors of interest. The error bars show the standard deviation across subjects of the difference of the M100 amplitudes between faces and each category of nonface object.

Figure 3: Stimulus and Task of Exp 4. (**a**) In each trial, a sample stimulus was followed after a delay by a test stimulus. The sample images (5 exemplars for each category) were profile-view faces or 3/4-view houses. (**b**) Test stimuli were frontal views of the sample stimuli. The phase coherence of the test stimuli was varied from 0% (visual noise) to 90% in four levels; original images with 100%

coherence were also included. Here we report the data for the stimuli presented at categorization and identification thresholds only.

Figure 4: Categorization versus Identification. (**a**) MEG waveforms from the face categorization (left) and identification (right) tasks. Blue: success; Red: failure. The waveforms were generated by averaging the selected raw data (see Methods) from independently defined SOIs in 10 subjects. (**b**) The amplitudes of the M100 and M170 at SOIs averaged across subjects. Successful categorization of faces elicits higher amplitudes at both M100 and M170 (top left), but no significant difference was found between successfully and unsuccessfully categorized houses at the predefined SOIs (top right). Correctly identified (compared to incorrectly identified) faces produce a significantly larger amplitude only of the M170, but not of the M100 (bottom left). The amplitude elicited by houses is not affected by success or failure in the identification task (bottom right).

Figure 5: Face configuration versus face parts. (**a**) Example: “Configuration” and “Parts” stimuli. (**b**) Amplitudes of the M100 and the M170 response, averaged across subjects, to Configuration and Parts stimuli at predefined sensors of interest.

Table 1: Accuracy as a function of task and stimulus category (Exp 4, guessing corrected)

	Categorization Task		Identification Task	
	Face	**House**	**Face**	**House**
Categ. Thresh. Level	**74%**	**72%**	**26%**	**19%**
Ident. Thresh. Level	**95%**	**95%**	**73%**	**65%**

Figure 1

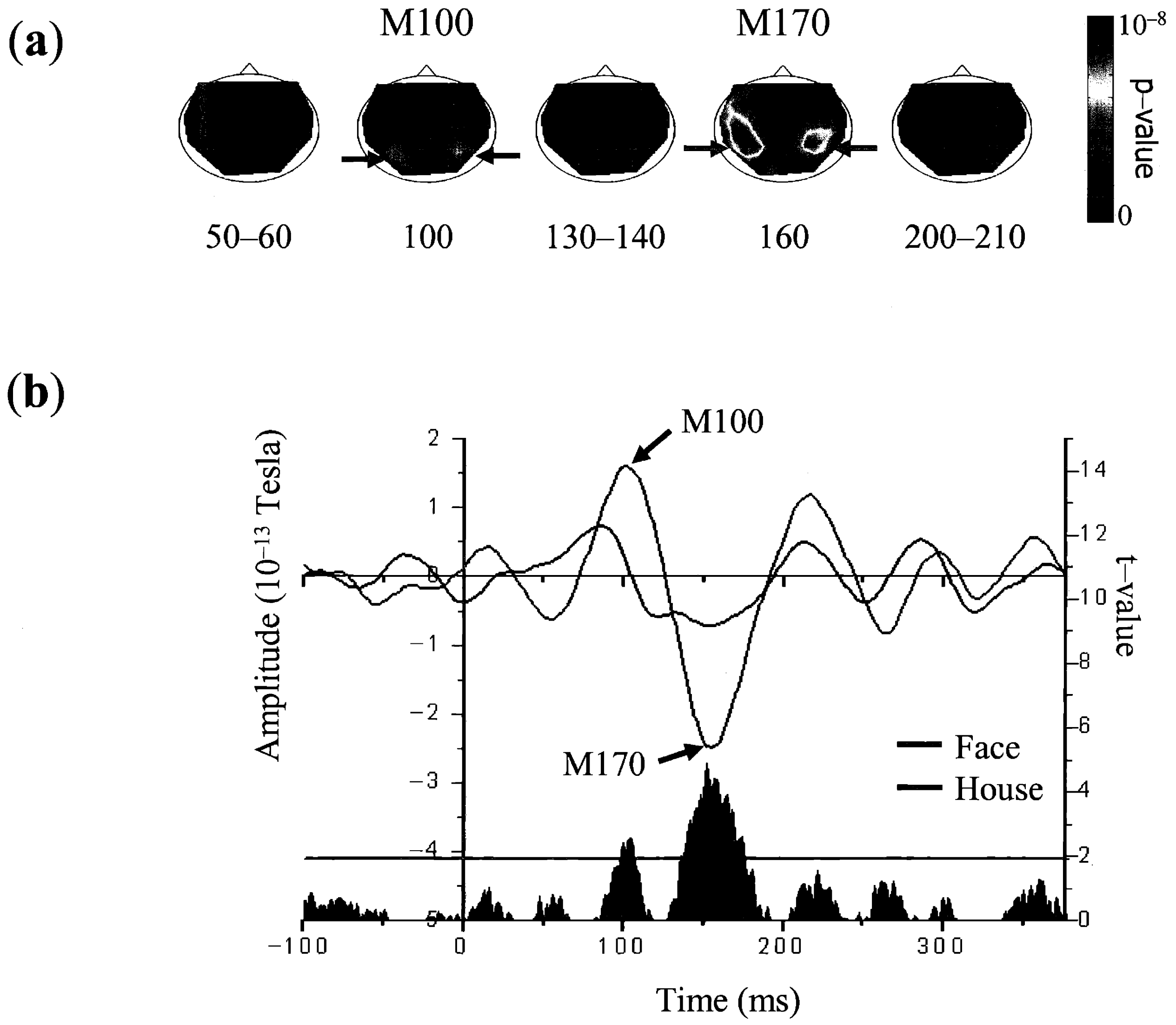

(a)
M100
M170
50–60
100
130–140
160
200–210
10^{-8}
p–value
0
(b)
M100
M170
Amplitude (10^{-13} Tesla)
2
1
0
−1
−2
−3
−4
−5
t–value
14
12
10
8
6
4
2
0
−100
0
100
200
300
Time (ms)
Face
House

Figure 2

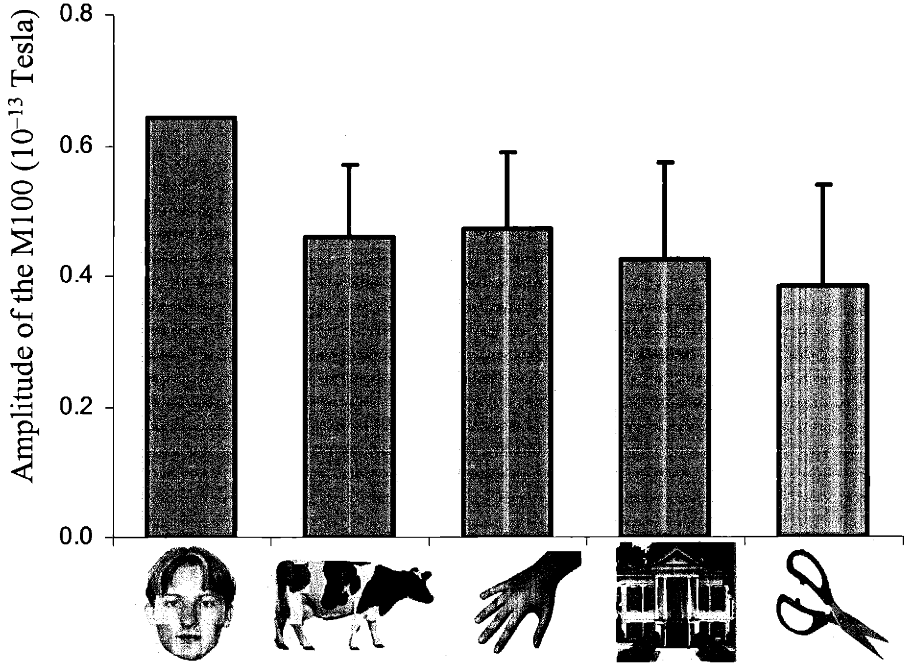


Figure 3

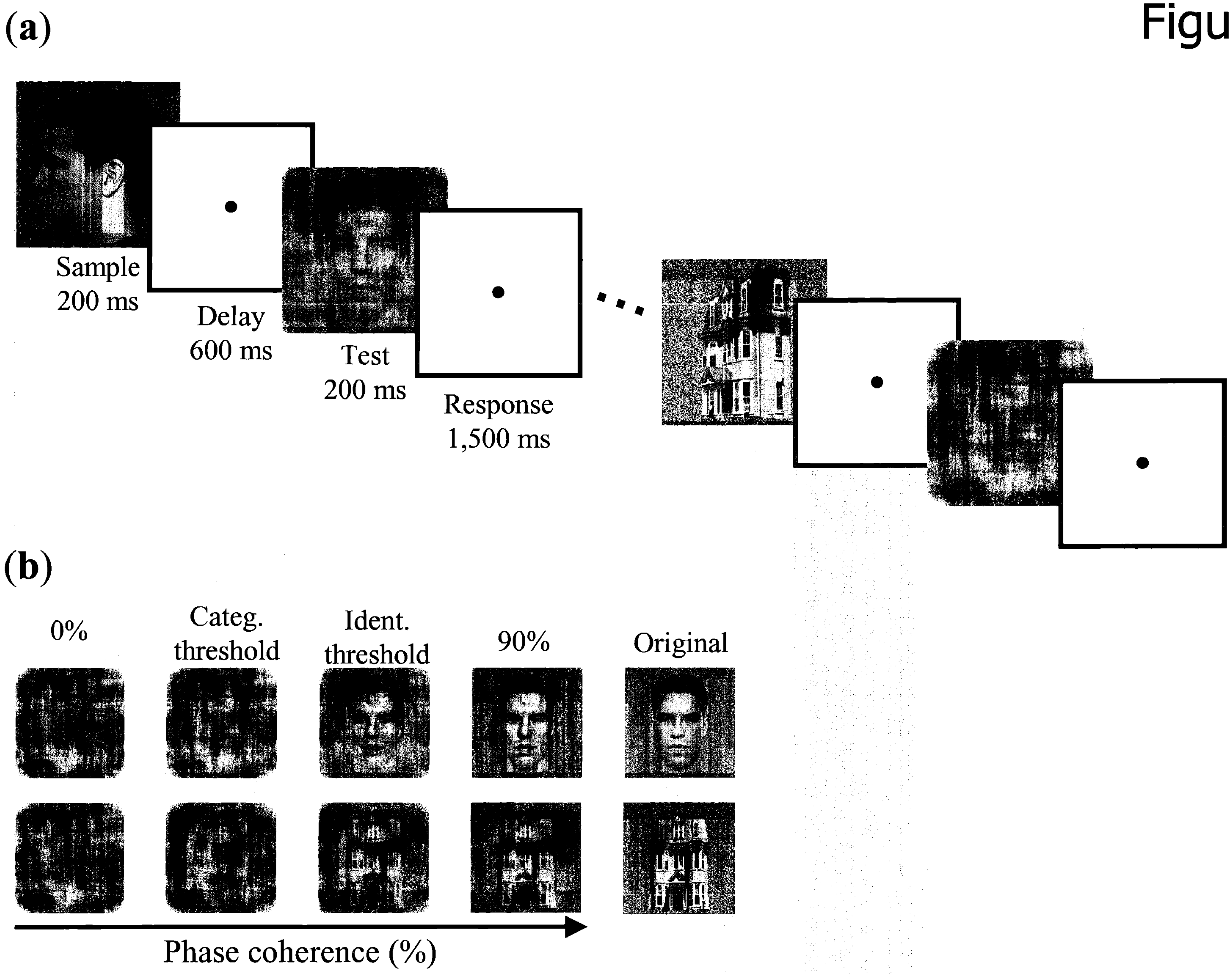

(a)
Sample
200 ms
Delay
600 ms
Test
200 ms
Response
1,500 ms
(b)
0%
Categ.
threshold
Ident.
threshold
90%
Original
Phase coherence (%)

Figure 4

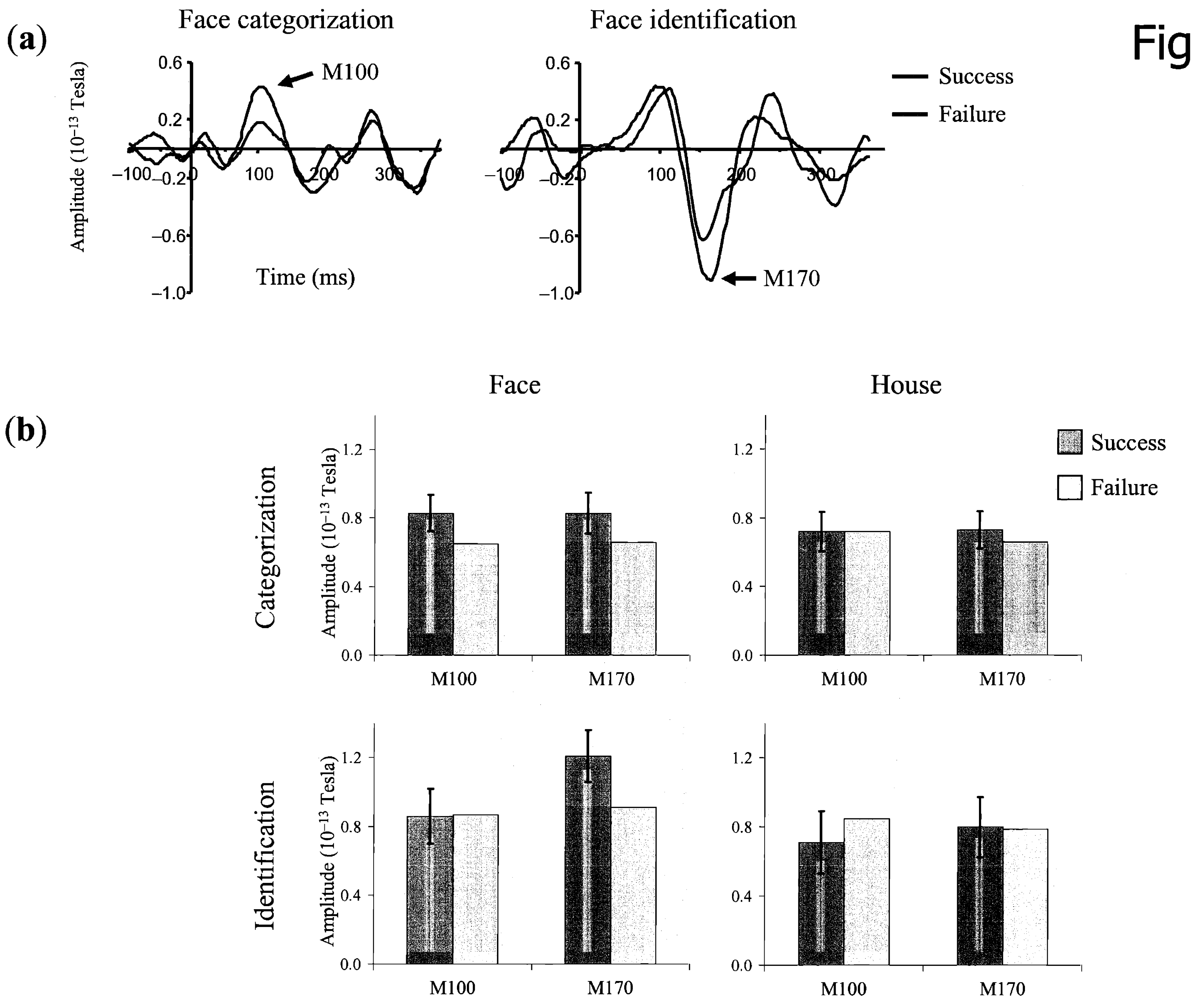
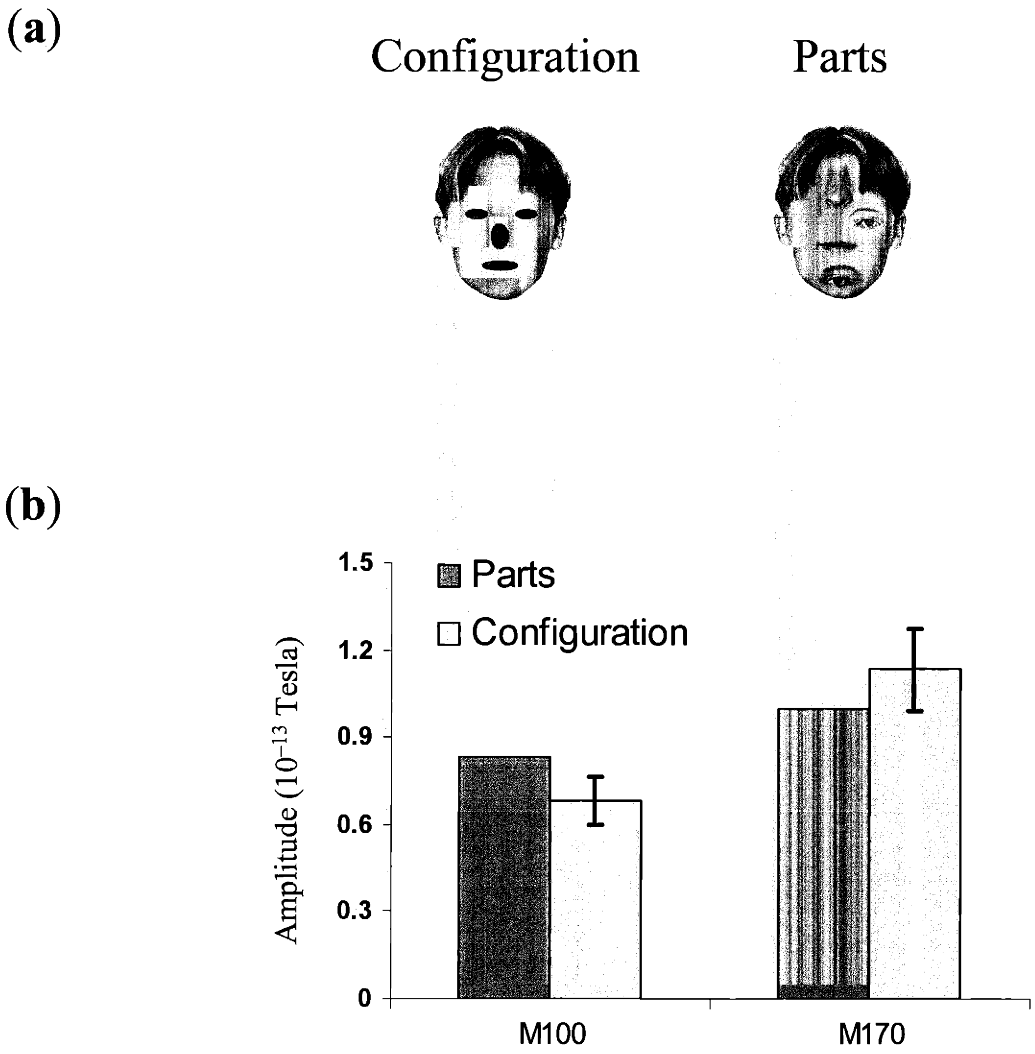

(a)
Face categorization
Face identification
Amplitude (10^{-13} Tesla)
0.6
0.2
−0.2
−0.6
−1.0
−100
0
100
200
300
M100
M170
Time (ms)
Success
Failure
(b)
Face
House
Categorization
Identification
Amplitude (10^{-13} Tesla)
1.2
0.8
0.4
0.0
M100
M170
Success
Failure

Figure 5


(a)
Configuration
Parts
(b)
Amplitude (10^{-13} Tesla)
1.5
1.2
0.9
0.6
0.3
0
▨ Parts
☐ Configuration
M100
M170

Chapter 4

Parts before Configurations in Face Processing: An MEG Study

Abstract

Prior studies using MEG have identified two face-selective responses, the M100, which is correlated with successful categorization of a stimulus as a face, and the M170, which is correlated with successful identification of individual faces (Liu, Harris, & Kanwisher, 2002). Here we measured the sensitivity of each of these response components to stimuli containing face parts (eyes, noses, and mouths) and face configurations. In Experiment 1, both the M100 and the M170 showed a larger response to stimuli with face parts than stimuli with black ovals in the corresponding locations, whereas only the M170 responded more strongly to stimuli with veridical face configurations than stimuli without. In Experiment 2, pairs of eyes alone elicited larger M100 and M170 responses than pairs of car wheels, indicating that these processing

stages do not encode just any object parts. Experiment 3 found that the latency of the M170, but not the M100, was delayed when faces were inverted. These findings provide evidence that the processing of face parts begins 70 ms before the processing of face configurations.

4.1 Introduction

Within a fraction of a second of encountering a familiar person, we already know who it is. How is this accomplished? What is the sequence of cognitive and neural processes underlying our ability to recognize faces?

Prior studies using magnetoencephalography (MEG) in humans have identified two neural responses over occipitotemporal cortex that are larger for face stimuli than for a variety of nonface stimuli. The 'M100', occurring only 100 ms after stimulus onset, is correlated with successful categorization of a stimulus as a face, but not with successful identification of individual faces (Liu et al., 2002). In contrast, the 'M170' (Sams, Hietanen, Hari, Ilmoniemi, & Lounasmaa, 1997; Liu, Higuchi, Marantz, & Kanwisher, 2000; Halgren, Raij, Marinkovic, Jousmaki, & Hari, 2000; or the 'N170' in ERP, Bentin, Allison, Puce, Perez, & McCarthy, 1996; Jeffreys, 1996; or the 'N200' in intracranial ERP, Allison, Puce, Spencer, & McCarthy, 1999; McCarthy, Puce, Belger, & Allison, 1999; Puce, Allison, & McCarthy, 1999), occurring 70 ms later, is correlated both with successful face categorization and with the identification of individual faces (Liu et al., 2002). These data suggest that face perception undergoes distinct stages of processing (see aslo Sugase, Yamane, Ueno, & Kawano, 1999; Itier & Taylor, 2002). In an effort to better characterize the sequence of processing steps involved in face perception, the present study asked what kind of information is extracted from face stimuli at each of these processing stages.

A substantial behavioral literature indicates that faces are processed in a more "holistic" or "configural" fashion than other visual stimuli. For example, accuracy of discriminating individual face parts, such as the nose, is higher when the entire face is presented than when the parts are presented in isolation; however, the same holistic advantage is absent for parts of houses (Tanaka & Farah, 1993). Another line of evidence that faces are processed holistically comes from studies of "composite" face stimuli in which the top half of one face is placed above and contiguous with the bottom half of a different face. Subjects are slower and less accurate in identifying either half when the two halves are aligned, compared to when they are misaligned, presumably because processing of the "whole" face in the aligned case is automatic, increasing interference from the other half of the face (Young, Hellawell, & Hay, 1987; see aslo Carey & Diamond, 1994; Hole, 1994). Neither of these effects is found when the same tasks are carried out on inverted faces (Tanaka & Sengco, 1997; Young et al., 1987), indicating that configural/holistic face processing occurs only for upright faces.

While these and other studies suggest that configural/holistic processing plays a special role in face recognition, the shapes of individual face parts (i.e. eyes, noses, and mouths) are also important in face perception. First, we can distinguish between faces that differ only in the shapes of their component parts (Macho & Leder, 1998; Cabeza & Kato, 2000; Le Grand, Mondloch, Maurer, & Brent, 2001; Leder & Bruce, 2000). Second, a recent computational study (Ullman, Vidal-Naquet, & Sali, 2002) shows that face parts ("features of intermediate complexity") are more useful than whole face configurations in distinguishing between faces and nonfaces, suggesting that face parts may play a particularly important role in basic-level categorization of a stimulus as a face versus a nonface.

For the present study we predicted that the M100 would be more sensitive to face parts, and the M170 would be more sensitive to face configurations, based on prior evidence that: 1) the M100 is more involved in categorizing a stimulus as a face whereas the M170 is more involved in identifying individual faces (Liu et al., 2002), and 2) local face parts may be more important for categorizing a stimulus as a face (Ullman et al., 2002), whereas configural information is more informative for identifying individual faces (McKone, Martini, & Nakayama, 2001). To test this hypothesis we measured the extent to which face parts and face configurations drive the M100 and M170.

Specifically, in Experiment 1 we measured the amplitudes and latencies of the M100 and the M170 responses to stimuli in which we orthogonally varied whether the images contained: a) real face parts versus solid black ovals in the corresponding locations; and b) veridical face configurations versus rearranged nonface configurations (Figure 1). In Experiment 2, we further tested the selectivity of both the M100 and the M170 to face parts, by comparing the response to pairs of eyes alone with the response to stimuli that were visually similar to eyes, and were also object parts, but had nothing to do with faces: partial side views of cars showing pairs of wheels. Finally, in Experiment 3, we measured the amplitude and latency of the M100 and the M170 to upright and inverted faces. Given that inversion is known to impair configural processing (Tanaka et al., 1997; Young et al., 1987) while sparing encoding of local features (Freire, Lee, &

Symons, 2000; Leder et al., 2000; Le Grand et al., 2001), we predicted that the M170 would be more sensitive than the M100 to face inversion.

4.2 Results

4.2.1 Experiment 1: Face parts versus Face configuration

In Experiment 1, we examined the extent to which face parts and face configurations drive the M100 and M170 by presenting four categories of stimuli in which face parts and configuration were orthogonally manipulated in a 2x2 design (Figure 1). In addition, full face and house stimuli were presented as controls.

Face-selective sensors of interest ("SOIs") over occipitotemporal cortex were identified separately for each subject and hemisphere from an independent localizer experiment; these SOIs showed significantly larger responses for both the M100 and the M170 to faces compared to houses (see Methods). Figure 2 shows the MEG response of a typical face-selective sensor in the right hemisphere from one subject in this localizer experiment.

Amplitude of M100 and M170 Responses

The amplitude data for each condition in Experiment 1 (Figure 3) were analyzed in a 4-way ANOVA, where the factors were MEG component (M100 versus M170), hemisphere (left versus right), face parts (real versus black ovals), and face configurations (veridical versus scrambled). This ANOVA found significant main effects of MEG component ($F(1,13) = 6.0$, $p < 0.03$), face parts ($F(1,13) = 32.3$, $p < 0.0001$), and face configuration ($F(1,13) = 7.3$, $p < 0.02$), and significant interactions of the M100 and the M170 by face configuration ($F(1,13) = 34.8$, $p< 0.0001$). The interaction of the M100 versus M170 by face parts did not reach significance ($F(1, 13) = 2.3$, $p > 0.1$), and neither did the interaction of hemisphere by face parts ($F(1,13) = 1.7$, $p > 0.2$). No other main effects or interactions reached significance (all Fs < 1).

The significant main effect of face parts indicates that both the M100 and M170 are sensitive to face parts. In fact, despite the trend of an interaction between MEG component and face parts, both M100 ($F(1,13) = 21.7$, $p < 0.001$) and M170 ($F(1,13) = 14.5$, $p < 0.002$) responses were significant larger when real face parts rather than black ovals were present (Figure 4a). The effect of face parts was observed independent of whether the face configuration was present ($F(1,13) = 32.1$, $p < 0.001$) or absent ($F(1,13) = 18.9$, $p < 0.001$).

However, only the M170, not the M100, was sensitive to face configurations. This was revealed by the previous–mentioned significant interaction of MEG component by face configuration, along with a significant effect of face configurations for the M170 ($F(1,13) = 29.7$, $p < 0.0001$), but not for the M100 ($F(1,13) < 1$) (Figure 4b). This interaction of M100 versus M170 by face configuration was observed independent of whether the face parts were present ($F(1,13) = 17.3$, $p < 0.001$) or absent ($F(1,13) = 12.0$, $p < 0.005$).

A separate test on the amplitude of the M170 alone found no interaction of face parts by face configuration ($F(1,13) < 1$).

In addition, the main effect of MEG components indicates that the M170 amplitude was larger than the M100 amplitude in general. In addition, the ratio of the response to full faces versus houses was lower for the M100 (1.6) than for the M170 (2.1, interaction $p<0.001$). This finding indicates that face selectivity sharpens over time.

Latency

The 4-way ANOVA on latency found a significant main effect of a later response to stimuli containing face parts than stimuli without ($F(1,13) = 12.0$, $p < 0.004$); the interaction of the M100 versus M170 by face configuration did not reach significance ($F(1, 13) = 2.3$, $p = 0.15$). A main effect of hemisphere ($F(1,13) = 4.7$, $p < 0.05$) showed that latencies were faster in the right than left hemisphere. No other main effects or interactions reached significance.

The significant main effect of face parts in latency indicates that both the M100 and M170 are sensitive to face parts. In fact, the presence of real face parts significantly delayed the M100 and the M170 (5 ms on average in both cases), compared to black ovals in the corresponding locations (both $Ps < 0.05$); the faster responses to the black ovals might be due to their simpler pattern. The absence of an interaction of MEG components by face parts ($F(1, 13) < 1$) suggests that the slower response for face parts seen in the M170 response may be inherited from the M100 response. In addition, a separate test failed to find any significant difference in latency between face parts and black ovals in the MEG response components occurring before the M100 (i.e. < 100ms) either at the SOIs or at other non-face selective sensors (all Fs<1), suggesting the delay in latency was induced at the M100 stage.

Following upon the trend of an interaction between MEG component and face configuration (not significant), separate tests on the latency of each these two MEG responses found that the latency of the M100 was not sensitive to face configuration information ($F(1, 13) < 1$) whereas the M170 latency was significantly delayed (7 ms on average) when local parts were rearranged in nonface configurations ($F(1, 13) = 7.9$, $p < 0.02$). This result accords with the amplitude data in showing that only the M170 is sensitive to face configuration.

The main effect of hemisphere on latency reflects the fact that both the M100 and the M170 in the right hemisphere reached their peak responses about 6 ms earlier on average than those in the left hemisphere. No significant interaction of MEG component by hemisphere was found ($F(1,13) = 2.6$, $p = 0.13$). Given that the same pattern was not found for houses ($F(1,13) < 1$), this result suggests that the hemispheric asymmetry in latency is specific for face processing.

4.2.2 Experiment 2: face parts versus object parts

The results from Experiment 1 demonstrate that both the M100 and M170 are sensitive to face parts. However, it is possible that the neural mechanisms underlying these responses are actually involved in more general encoding of object parts, rather than representing face parts specifically. To test this hypothesis, we compared the response to eyes alone with the response to object parts that were visually similar to pairs of eyes but had nothing to do with faces: partial side views of cars showing pairs of wheels (see Figure 5). Full faces and whole cars were also included as control conditions.

The M100 and M170 responses to these stimulus conditions recorded at pre-defined face-selective sensors, average across subjects, are shown in Figure 5. The M100 response to pairs of eyes was significantly larger (Figure 5A) than that to pairs of wheels ($F(1,12) = 9.0$, $p < 0.02$) and whole cars ($F(1,12) = 5.0$, $p < 0.05$), and not significantly smaller than that to whole faces ($F(1,12) = 1.1$, $p > 0.1$). Similar results were found for the M170 response (Figure 5B). In addition, neither interaction of stimulus condition by hemisphere nor by MEG response component was found (all $Fs < 1$).

4.2.3 Experiment 3: face inversion

In Experiment 1, we disrupted configural processing by rearranging face parts into a novel nonface configuration. Although this manipulation controlled the amount of pre-experimental exposure to face parts (e.g. upright noses have been seen as often as upright full faces), it changed other aspects of the stimulus images, such as their spatial frequency spectra. In Experiment 3, we used the well-known phenomenon of face inversion (Yin, 1969; Valentine, 1988), which is known to impair configural processing of faces while preserving most low-level properties of the stimulus, to further test whether only the M170 is sensitive to configural information.

Figure 6 shows the average across subjects of the response recorded from face-selective SOIs to upright and inverted faces. The M170 latency was significantly delayed (11 ms on average) for inverted compared to upright faces ($F(1, 12) = 36.50$, $p < 0.0001$), but no significant difference in M100 latency was found ($F(1, 12) = 1.25, p > 0.2$). The finding that inversion delays only the M170 is further supported by a significant interaction of M100 versus M170 by upright and inverted faces ($F(1, 12) = 16.56, p < 0.002$) . No significant difference was observed in amplitude for upright versus inverted faces (all $Fs < 1$).

4.3 Discussion

In the present study, we examined the sensitivity of two consecutive face-selective MEG responses, the M100 and the M170, to two kinds of face information: face parts (eyes, nose, and mouth) and face configurations (the standard T–shaped arrangement of these parts in faces). In Experiment 1, amplitudes of both the M100 and M170 responses were larger to stimuli that contained actual face parts than to stimuli containing black ovals in the corresponding locations. However, only the M170 was sensitive to face configurations (regardless of whether that configuration is comprised of real face parts or black ovals). A similar pattern was also observed in latency: the replacement of face parts with black ovals affected the latencies of both the M100 and the M170 responses, whereas the rearrangement of face configurations delayed the latency of the M170, but not the M100. The results from Experiment 2 rule out the possibility that the M100 and the M170 are sensitive to any object parts, given that the amplitudes of both M100 and M170 responses to pairs of eyes were significantly larger than to pairs of car wheels. Finally, in Experiment 3 inversion of faces, which impairs holistic processing but spares encoding of face parts, only delayed the latency of the M170, not the M100. In sum, the results from these three experiments are consistent with our prediction that the M100 is more sensitive to face parts and the M170 is more sensitive to face configurations. Furthermore, these findings suggest that face parts are encoded as early as 100 ms after stimulus onset, whereas face configurations are extracted 70ms later.

Second, we also found a right hemisphere advantage in face recognition by showing that neural activity elicited by faces, but not houses, reached its peak response about 6 ms earlier in the right than the left hemisphere (see also Allison et al., 1999; Yovel, Levy, Grabowecky, & Paller, 2002). This hemispheric asymmetry was seen in both M100 and M170 responses, implying that processing of faces first happens in the right hemisphere and then possibly spreads to the left hemisphere. However, no significant differences between hemispheres were found in processing face parts versus face configurations, suggesting there is no hemispheric bias in processing these two kinds of face information at the early stages of face processing. Behavioral (Leehey, Carey, Diamond, & Cahn, 1978; Hillger & Koenig, 1991; Rhodes, Brake, & Atkinson, 1993) and PET (Rossion et al., 2000a) findings that the right hemisphere is superior at processing face configurations whereas the left hemisphere is more engaged in part-based processing (see also Navon, 1977; Marsolek, Kosslyn, & Squire, 1992; Martinez et al., 1997) may reflect the properties of face processing at the later stages (Fink et al., 1996; Yovel et al., 2002).

Relationship to Previous ERP studies

The relationship between the face-selective M170 and face-selective responses reported with ERP methods is unclear. Because MEG can detect only a subset of the neural activity that can be detected with scalp ERPs (Hamalainen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993), the two measures do not necessarily reflect identical underlying neural sources. Nonetheless, because each measure reveals a face-selective response at the same latency after stimulus presentation in similar scalp locations, it is likely there is some relationship between them. Our M170 data is consistent with prior N170 data from scalp ERPs in showing a delayed latency of response to 1) inverted, compared to upright, faces (Bentin et al., 1996; Linkenkaer-Hansen et al., 1998; Rossion et al., 1999; Rossion et al., 2000b; Eimer, 2000a; Sagiv & Bentin, 2001; Taylor, Edmonds, McCarthy, & Allison, 2001; de Haan, Pascalis, & Johnson, 2002; Itier et al., 2002), 2) faces with distorted configurations in which face features remain intact (George, Evans, Fiori, Davidoff, & Renault, 1996; Eimer & McCarthy, 1999), and 3) contrast–reversed faces (Itier et al., 2002). Our M170 data presented here is also consistent with the finding that the N170 is sensitive to face parts, as its amplitude is attenuated and the latency is delayed when face parts are removed (Eimer, 2000b; see also Eimer, 1998; Jemel, George, Chaby, Fiori, & Renault, 1999). Thus, both the N170 and the M170 are sensitive to both face parts and face configurations.

However, the response profiles of the N170 and M170 differ in some respects. Most notably, the amplitude of the M170 is larger in response to upright full faces than to face parts, whereas the N170 shows an opposite pattern (Bentin et al., 1996; Rossion et al., 2000b; Sagiv et al., 2001). It has been argued that the N170 reflects two underlying sources, one more sensitive to whole faces and located in the middle fusiform gyrus, and the other more sensitive to face parts, especially eyes, and located on posterior middle temporal gyrus (Bentin et al., 1996; Allison et al., 1999). Halgren et al. (2000) provide evidence that the equivalent current dipole of the M170 is localized in the fusiform gyrus, which seems to correspond to the posterior fusiform (whole-face-sensitive) source of the N170. Consistent with this hypothesis, direct subdural electric recordings from this region show a larger response (the N200) for full faces than for face parts (McCarthy et al., 1999), suggesting a close correspondence between the M170 and the intracranial N200, but not the scalp N170. This hypothesis is further supported by evidence that the N170 is larger or is even exclusively located in the right hemisphere (Bentin et al., 1996; George et al., 1996), whereas no significant differences in amplitude have been found between hemispheres for the M170 and intracranial N200 (Liu et al., 2000; Allison et al., 1999).

The relationship between the M100 and the P1 that precedes the N170 is also unclear. Many previous ERP studies on face inversion only report its effect on the N170, not the P1, presumably because no effect on the P1 was observed. Among those who do report these results, the findings are contradictory: some report that P1 amplitude and latency were not affected by inversion (Rossion et al., 1999), some report that only P1 latency was delayed (Itier et al., 2002), and others demonstrate both that P1 latency was delayed and that P1 amplitude was enhanced (Linkenkaer-Hansen et al., 1998). Another ERP study on the P1 reported that alterations of face configurations by moving the eyes and the mouth away from the nose affected the amplitude of the P1 (Halit, de Haan, & Johnson, 2000), suggesting that this early response may be sensitive to face configurations. However, because the configurational alterations used by Halit et al

(2000) make the original faces less attractive, the modulation seen in the P1 may actually reflect rapid emotional face processing (Pizzagalli, Regard, & Lehmann, 1999; Eimer & Holmes, 2002); the P1 modulation in that study may also simply reflect low-level differences between the altered and the original faces, as the authors note. In the present study we controlled low-level visual properties (e.g., spatial frequencies) and amount of pre-experimental exposure to all aspects of face stimuli except face configurations by either inverting these faces or rearranging the upright face parts in nonface configurations. The results from these two manipulations of face configuration converge to demonstrate that the M100 is not sensitive to face configuration.

In sum, although many questions remain about the relationship between the M100 and M170 and neural responses measured at similar latencies with scalp and subdural ERPs, the present data are not inconsistent with prior results from ERPs.

4.4 Conclusions

In the present study, we have shown that different kinds of information about face stimuli are extracted at different stages of face processing, with the extraction of information about face parts beginning at 100 ms after stimulus onset, and the extraction of information about face configurations beginning about 70 ms later. The encoding sequence of face parts and then face configurations cannot be explained in terms of a progression from coarse or global information to fine or local information (Navon, 1977; Marr, 1982; Sergent, 1986), given that the encoding of detailed edges and precise shapes of face parts requires if anything higher spatial frequencies than their arrangement (i.e.

face configurations). Instead, combined with our prior finding that the M100 is correlated with successful categorization of a stimulus as a face whereas the M170 is correlated with successful face identification, these results suggest that face parts play a special role in the categorization of a stimulus as a face (consistent with a recent computational result (Ullman et al., 2002)), whereas face configurations play a special role in face identification (consistent with a large behavioral literature (e.g. Young et al., 1987; Tanaka & Farah, 1993; McKone et al., 2001)).

Many questions remain. Do the M100 and the M170 arise from the same neural source, or are the two response components generated by nearby but distinct neural populations? If the two are distinct, does the neural population that generates the M170 get its inputs directly from the population that generates the M100, or do the M100 and M170 reflect parallel processing paths? The serial model would provide a natural explanation of the fact that the M170 seems to "inherit" the response properties of the M100 (i.e., its sensitivity to face parts, and its correlation with successful face categorization), while adding new properties (i.e., a sensitivity to face configurations and face inversion, and a correlation with successful face identification). But even this simple serial model leaves other important questions unanswered. For example, a full account of face perception will ultimately have to determine whether and how information about face parts and face configurations ever becomes integrated into a unified face percept, or whether the two kinds of information remain forever distinct and separate.

4.5 Methods

MEG Recordings were made using a 64–channel (Experiment 1 and 3) or 96–channel (Experiment 2) whole head system with SQUID–based first–order gradiometer sensors (Kanazawa Institute of Technology system) at the KIT/MIT MEG Joint Research Lab at MIT). Magnetic brain activity was digitized continuously at a sampling rate of 1000 Hz (500 Hz for Experiment 2) and was filtered with 1Hz high pass and 200Hz low-pass cutoff and a 60 Hz notch. Informed consent was obtained from all subjects, and the study was approved by the MIT Committee on the use of humans as experimental subjects (COUHES).

Fifteen subjects (age range 18–41) participated in Experiment 1, thirteen (18–30) in Experiment 2, and fifteen (18–40) in Experiment 3. In each experiment, photographs of human faces and of houses were included in a localizer experiment to define face-selective sensors over occipitotemporal cortex bilaterally. Specifically, t-tests were conducted between the MEG responses to faces and to houses at each time point (from –100 to 400 msec) and each sensor separately. Sensors where the magnetic fields evoked by faces were significantly larger than those by houses ($Ps < 0.05$) for at least 5 consecutive time points (both within a 40-ms time window centered at the latency of the M100 and at that of the M170) were defined as Sensors of Interest (SOIs). All data reported here from SOIs are independent from the data used to define SOIs. Thus, the need to perform statistical corrections for comparisons across multiple sensors is eliminated. Second, by defining ROIs separately for each individual, the problems of anatomical variance across individuals can be largely circumvented. Subjects whose data failed to show clear face-selective SOIs (one in Experiment 1 and two in Experiment 3) were excluded from further analyses.

In Experiment 1, four stimulus categories were constructed from veridical faces by orthogonally eliminating either face parts (i.e. eyes, nose and mouth) or face configurations (see Figure 1). To remove the contribution of face parts, face parts were replaced by black ovals in their corresponding locations, whereas to disrupt face configuration, face parts (or black ovals) were rearranged from face configuration into a novel nonface configuration. The size of the ovals was matched to the actual size of corresponding face parts in each face stimulus, and the arrangements of nonface configurations varied across all exemplars. Because external face features, such as hairline, chin and ears, may interact with internal face features (both face configuration and face parts) (Sinha & Poggio, 1996; Liu, Harris, & Kanwisher, 2001), external face features were removed by a square cutout showing the central face region only. In addition, photographs of full faces (i.e. faces with face parts, veridical configuration and external features) and houses were included for comparison. In Experiment 2, photographs of pairs of eyes and pairs of wheels were constructed from front views of faces and side views of cars respectively. The size of wheel images was matched to that of eye images. In Experiment 3, photographs of inverted faces were created by rotating the image plane of upright faces 180 degrees.

In Experiment 1 and 3, fifty exemplars from each stimulus category were tested, and each exemplar was repeated once (i.e. 100 trials per category); In Experiment 2, ten exemplars from each category were tested and each was repeated 5 times (i.e. 50 trials per category). In each experiment, each image subtended 5.7 by 5.7 degree of visual angle and was presented at the center of gaze for 200 msec followed by a fixation-only interval of 800 msec on average (ranging from 600 to 1000 msec). Subjects were instructed to lie on the scanner bed and passively view a sequence of randomly interleaved images of each stimulus category.

4.6 Acknowledgement

We thank K. Nakayma, P. Sinha, J. DiCarlo, W. Freiwald, G. Yovel and the rest of Kanwisher Lab for comments on the manuscript. Supported by the Reed fund and NEI grant (EY13455) to N.K.

Reference List

Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb.Cortex, 9,* 415-430.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perceptions in humans. *Journal of Cognitive Neuroscience, 8,* 551-565.

Cabeza, R. & Kato, T. (2000). Features are also important: contributions of featural and configural processing to face recognition. *Psychol.Sci., 11,* 429-433.

Carey, S. & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition, 1,* 253-274.

de Haan, M., Pascalis, O., & Johnson, M. H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. *J.Cogn Neurosci., 14,* 199-209.

Eimer, M. (1998). Does the face-specific N170 component reflect the activity of a specialized eye processor? *Neuroreport, 9,* 2945-2948.

Eimer, M. (2000a). Effects of face inversion on the structural encoding and recognition of faces. Evidence from event-related brain potentials. *Brain Res.Cogn Brain Res., 10,* 145-158.

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

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

Eimer, M. & McCarthy, R. A. (1999). Prosopagnosia and structural encoding of faces: evidence from event-related potentials. *Neuroreport, 10,* 255-259.

Fink, G. R., Halligan, P. W., Marshall, J. C., Frith, C. D., Frackowiak, R. S., & Dolan, R. J. (1996). Where in the brain does visual attention select the forest and the trees? *Nature, 382,* 626-628.

Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: direct evidence. *Perception, 29,* 159-170.

George, N., Evans, J., Fiori, N., Davidoff, J., & Renault, B. (1996). Brain events related to normal and moderately scrambled faces. *Brain Res.Cogn Brain Res., 4,* 65-76.

Halgren, E., Raij, T., Marinkovic, K., Jousmaki, V., & Hari, R. (2000). Cognitive response profile of the human fusiform face area as determined by MEG. *Cereb.Cortex, 10,* 69-81.

Halit, H., de Haan, M., & Johnson, M. H. (2000). Modulation of event-related potentials by prototypical and atypical faces. *Neuroreport, 11,* 1871-1875.

Hamalainen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography -- theory, instrumentation, and applications to noninvasive studies of the working human brain. *Review of Modern Physics, 65,* 413-497.

Hillger, L. A. & Koenig, O. (1991). Separable mechanisms in face processing: Evidence from hemispheric specialization. *J.Cogn Neurosci., 3,* 42-58.

Hole, G. (1994). Configurational factors in the perception of unfamiliar faces. *Perception, 23,* 65-74.

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

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

Jemel, B., George, N., Chaby, L., Fiori, N., & Renault, B. (1999). Differential processing of part-to-whole and part-to-part face priming: an ERP study. *Neuroreport, 10,* 1069-1075.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Neuroperception. Early visual experience and face processing. *Nature, 410,* 890.

Leder, H. & Bruce, V. (2000). When inverted faces are recognized: the role of configural information in face recognition. *Q.J.Exp.Psychol.A, 53,* 513-536.

Leehey, S., Carey, S., Diamond, R., & Cahn, A. (1978). Upright and inverted faces: the right hemisphere knows the difference. *Cortex, 14,* 411-419.

Linkenkaer-Hansen, K., Palva, J. M., Sams, M., Hietanen, J. K., Aronen, H. J., & Ilmoniemi, R. J. (1998). Face-selective processing in human extrastriate cortex around 120 ms after stimulus onset revealed by magneto- and electroencephalography. *Neurosci.Lett., 253,* 147-150.

Liu, J., Harris, A., & Kanwisher, N. (2001). What makes a face a face: an MEG and fMRI study [Abstract]. *J.Vision, 1,* 340a.

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

Liu, J., Higuchi, M., Marantz, A., & Kanwisher, N. (2000). The selectivity of the occipitotemporal M170 for faces. *Neuroreport, 11,* 337-341.

Macho, S. & Leder, H. (1998). Your eyes only? A test of interactive influence in the processing of facial features. *J.Exp.Psychol.Hum.Percept.Perform., 24,* 1486-1500.

Marr, D. (1982). *Vision*. New York: W.H. Freeman and Company.

Marsolek, C. J., Kosslyn, S. M., & Squire, L. R. (1992). Form-specific visual priming in the right cerebral hemisphere. *J.Exp.Psychol.Learn.Mem.Cogn, 18,* 492-508.

Martinez, A., Moses, P., Frank, L., Buxton, R., Wong, E., & Stiles, J. (1997). Hemispheric asymmetries in global and local processing: evidence from fMRI. *Neuroreport, 8,* 1685-1689.

McCarthy, G., Puce, A., Belger, A., & Allison, T. (1999). Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex. *Cereb.Cortex, 9,* 431-444.

McKone, E., Martini, P., & Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *J.Exp.Psychol.Hum.Percept.Perform., 27,* 573-599.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognit.Psychol., 9,* 353-383.

Pizzagalli, D., Regard, M., & Lehmann, D. (1999). Rapid emotional face processing in the human right and left brain hemispheres: an ERP study. *Neuroreport, 10,* 2691-2698.

Puce, A., Allison, T., & McCarthy, G. (1999). Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials. *Cereb.Cortex, 9,* 445-458.

Rhodes, G., Brake, S., & Atkinson, A. P. (1993). What's lost in inverted faces? *Cognition, 47,* 25-57.

Rossion, B., Delvenne, J. F., Debatisse, D., Goffaux, V., Bruyer, R., Crommelinck, M. et al. (1999). Spatio-temporal localization of the face inversion effect: an event-related potentials study. *Biol.Psychol., 50,* 173-189.

Rossion, B., Dricot, L., Devolder, A., Bodart, J. M., Crommelinck, M., De Gelder, B. et al. (2000a). Hemispheric asymmetries for whole-based and part-based face processing in the human fusiform gyrus. *J.Cogn Neurosci., 12,* 793-802.

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S. et al. (2000b). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport, 11,* 69-74.

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

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

Sergent, J. (1986). Microgenesis of face perception. In H.D.Ellis, M. A. Jeeves, F. Newcombe, & A. Young (Eds.), *Aspects of face processing* (Dordrecht: Martinus Nijhoff.

Sinha, P. & Poggio, T. (1996). I think I know that face.. *Nature, 384,* 404.

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

Tanaka, J. W. & Farah, M. J. (1993). Parts and wholes in face recognition. *Q.J.Exp.Psychol.A, 46,* 225-245.

Tanaka, J. W. & Sengco, J. A. (1997). Features and their configuration in face recognition. *Mem.Cognit., 25,* 583-592.

Taylor, M. J., Edmonds, G. E., McCarthy, G., & Allison, T. (2001). Eyes first! Eye processing develops before face processing in children. *Neuroreport, 12,* 1671-1676.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nat.Neurosci.*.

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

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81,* 141-145.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16,* 747-759.

Yovel, G., Levy, J., Grabowecky, M., & Paller, K. A. Neural correlates of the left-visual-field superiority in face perception appear at multiple stages of face perception. *J.Cogn Neurosci.,* (in press).

Figure Legends

Figure 1: Stimulus Manipulations in Experiment 1. The 2x2 design involved orthogonal manipulation of the presence versus absence of face parts (eyes, nose, and mouth; horizontal axis) and face configurations (the placement of these parts in a face arrangement versus scrambled arrangement).

Figure 2: The MEG waveforms from the localizer experiment in a typical subject, averaged across all face and house trials at a typical face-selective sensor in the right hemisphere. Red: faces. Blue: houses. Black region: t value. The left vertical scale indicates the amplitude of the MEG response (10^{-13} Tesla); the right scale shows the t-value. The green horizontal line corresponds to $p = 0.05$ (uncorrected for comparisons at multiple time points).

Figure 3: Amplitudes of the peak M100 (A) and M170 (B) responses, averaged across subjects, to each stimulus category at predefined sensors of interest in Experiment 1. The error bars show the standard deviation across subjects of the difference of the MEG amplitudes between full faces and each category tested.

Figure 4: The averaged amplitudes of the M100 and the M170 response to face parts (A) and face configurations (B) in Experiment 1.

Figure 5: The averaged amplitudes of the M100 (A) and M170 (B) responses to each stimulus category in Experiment 2. The error bars show the standard deviation across subjects of the difference of the MEG amplitudes between pairs of eyes and each category tested.

Figure 6: The averaged MEG waveforms elicited by the presentation of upright (Red) and inverted (Blue) faces.

Figure 1

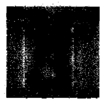
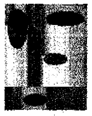
		Face Parts	
		Present	Absent
Face Configuration	Present		
	Absent		

Figure 2

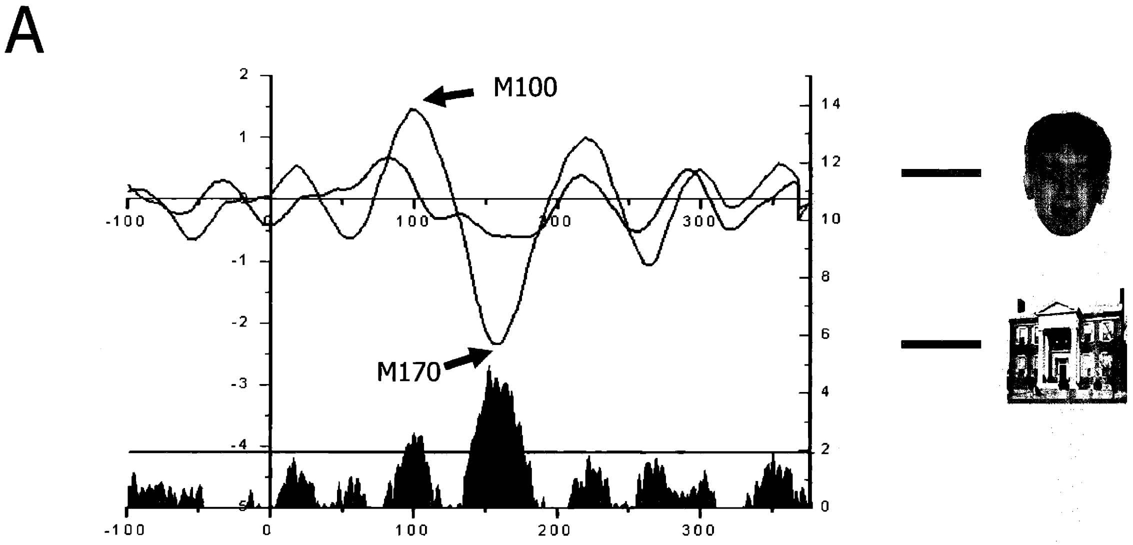
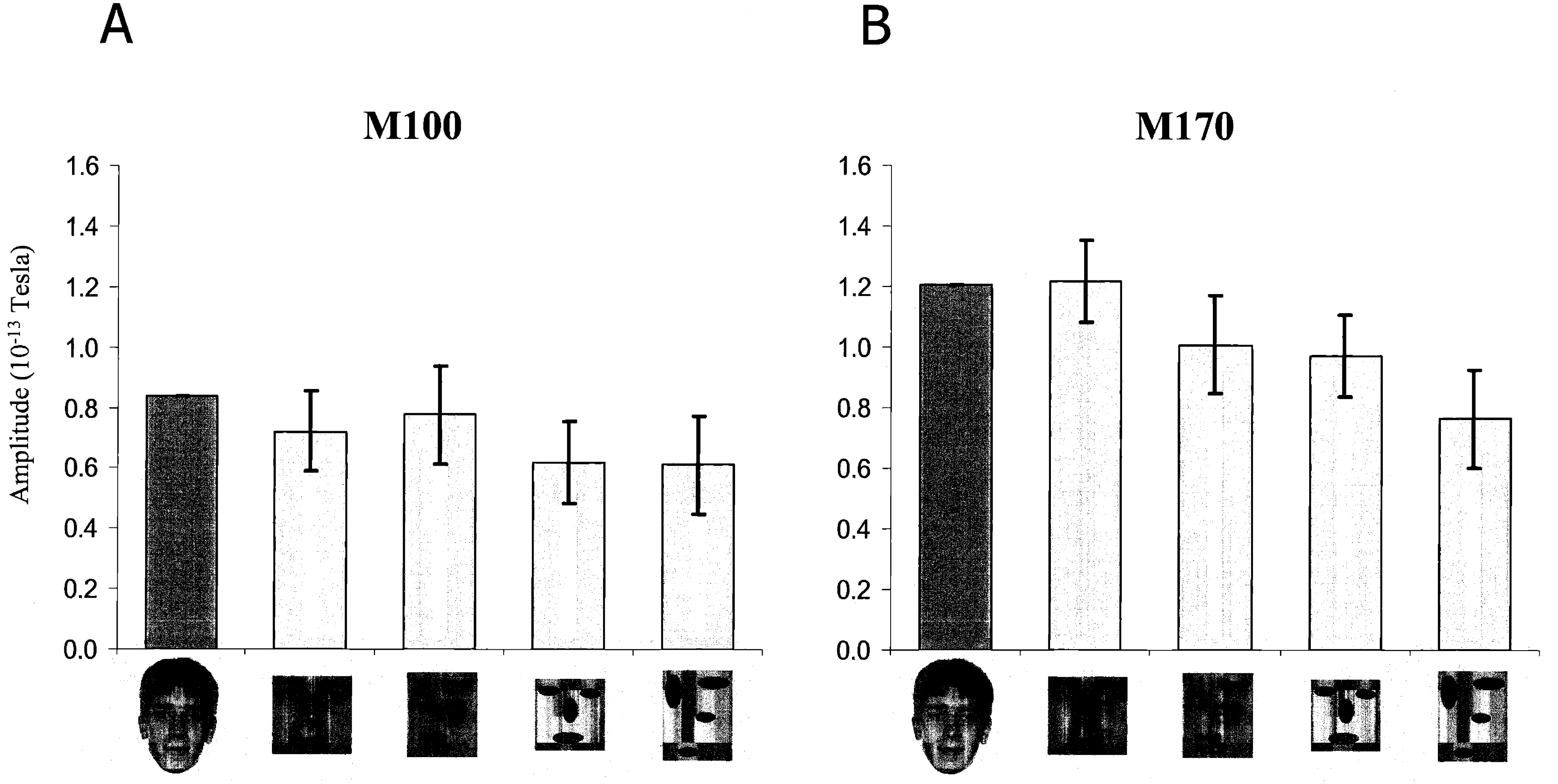

A
M100
M170
2
1
0
-1
-2
-3
-4
-5
14
12
10
8
6
4
2
0
-100
0
100
200
300
-100
0
100
200
300

Figure 3



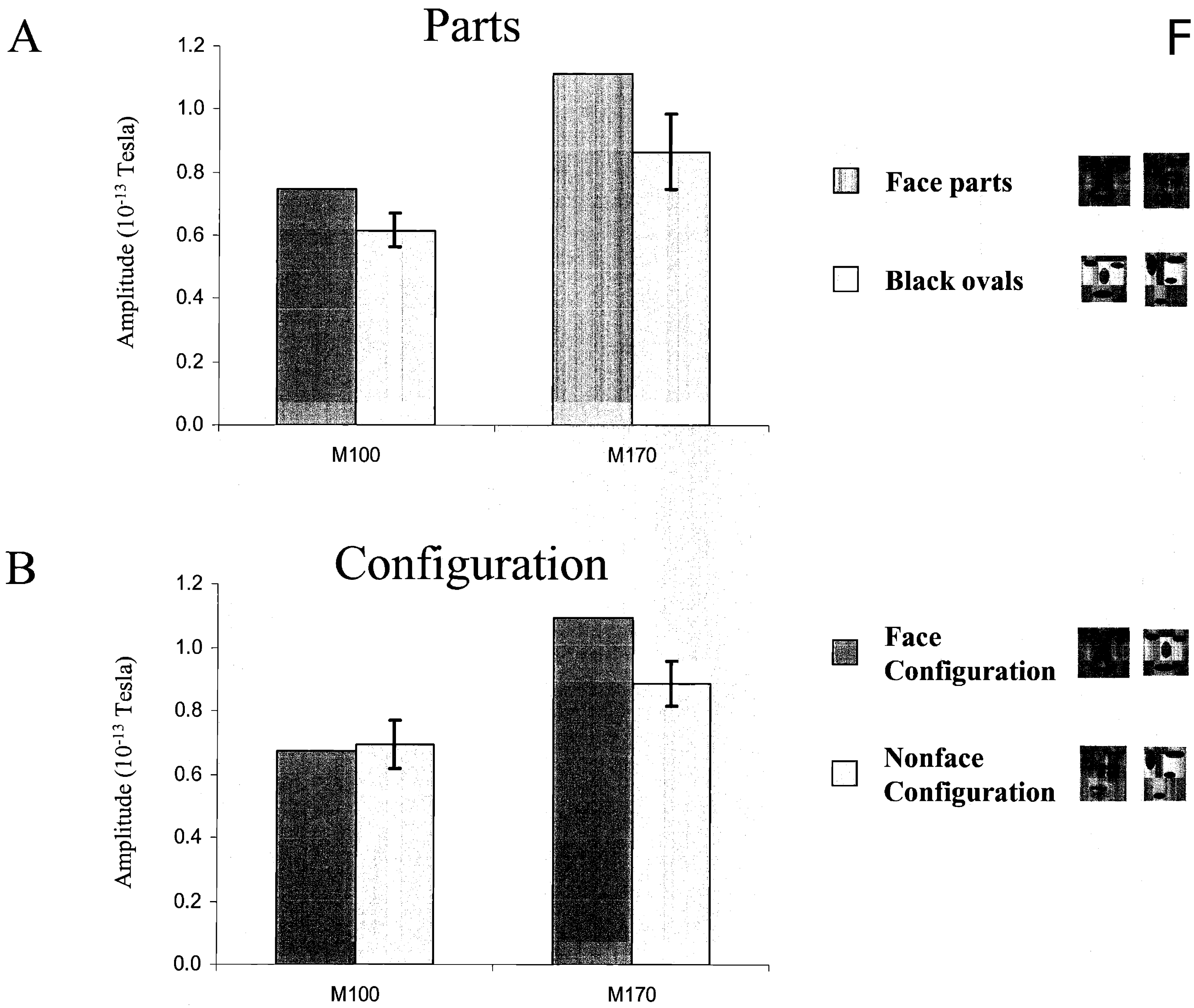

A
Parts
Amplitude (10^{-13} Tesla)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
M100
M170
Face parts
Black ovals
B
Configuration
Amplitude (10^{-13} Tesla)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
M100
M170
Face Configuration
Nonface Configuration

Figure 4

Figure 5

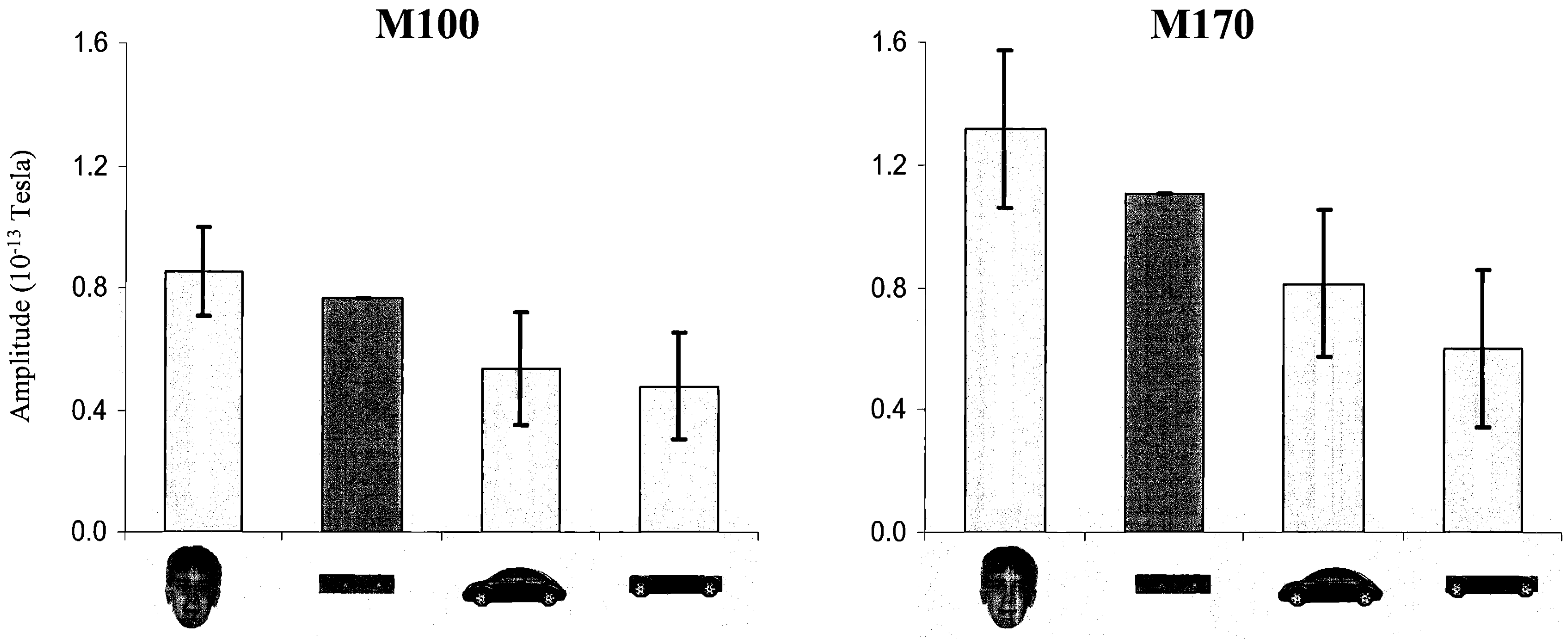
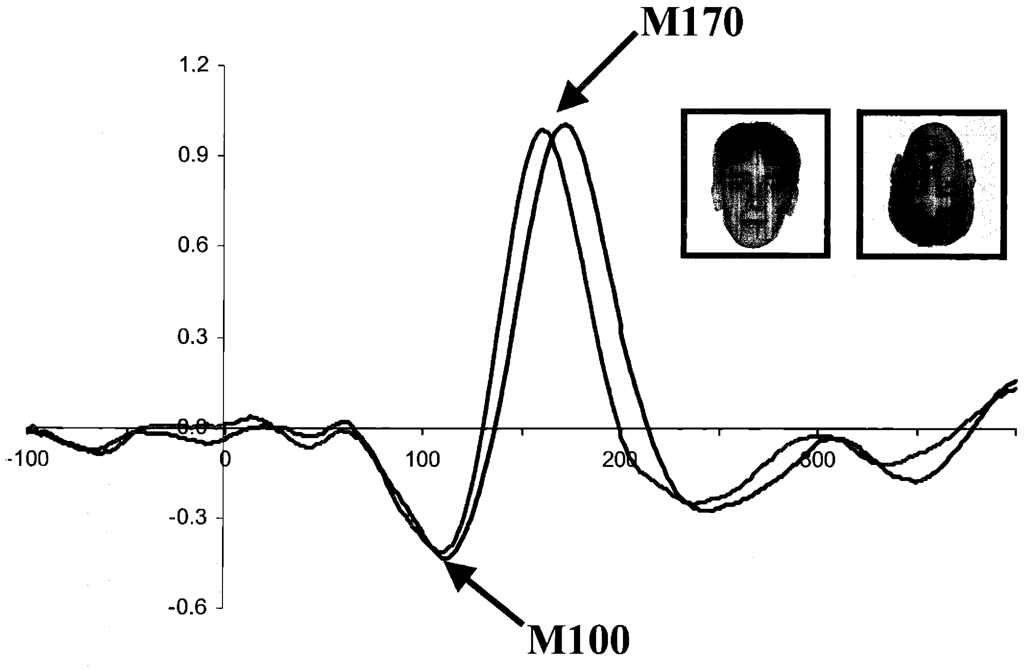


Figure 6



Chapter 5

General Conclusions and Future Work

This thesis used MEG to discover three new facts about face perception in humans. I synopsize each finding before speculating about the broader significance of the work.

5.1 Early stages of face recognition rely on feedforward network

MEG was used to investigate the speed of face recognition in humans. I found that face recognition begins substantially earlier than previously reported: by 100ms, MEG responses over occipitotemporal sensors are significantly higher to faces than to a variety of nonface stimuli. Further, this M100 response is higher when a degraded face stimulus is correctly categorized as a face than when the *same stimulus* is incorrectly categorized as a nonface, showing that the M100 reflects the subject's percept, not simply the low-level properties of visual stimulus.

Given that at least 60-80 ms are thought to be necessary for visual information to reach primary visual cortex in humans (Gomez Gonzalez, Clark, Fan, Luck, & Hillyard, 1994), this leaves only an additional 20–40 ms for the face–selective M100 to be generated. Such latencies are hard to reconcile with models of face categorization that rely on extensive iterative feedback loops and/or recurrent processing, and strengthen the evidence that initial stimulus categorization is accomplished by largely feedforward mechanisms (Thorpe & Imbert, 1989).

In sum, by sharply reducing current estimates of the earliest latency of stimulus categorization, this finding places empirical constraints on theories of human face recognition.

5.2 Two stages of face processing: Categorization begins prior to identification

In this work, I investigated the sequence of cognitive and neural processes underlying our ability to recognize faces. Subjects were instructed to make two judgments about each stimulus, determining both its category (face or house) and its individual identity, while the perceptibility of images was systematically manipulated. By concurrently monitoring subjects' behavioral performance and the associated changes in brain activity measured by MEG, I found that both the M100 and the M170 are correlated with successful categorization of a stimulus as a face versus nonface, but only the later M170 component is correlated with successful identification of individual faces. This finding indicates that the categorization of a stimulus as a face begins substantially earlier than the identification of the particular face.

Given that most people spend more time looking at faces compared to any other objects across their lifespan and given that the primary goal of face perception is to identify individuals, Rosch et al. (1976) and Tanaka (2001) argued that visual expertise in discriminating individual faces may shift the point of initial contact with memory representations from the basic level to the subordinate level. Thus, on their view, identification of individual faces should occur as fast as the categorization of a stimulus as a face. However, the MEG data argue against this view: processes critical for the categorization of a stimulus as a face. Thus, our expertise with faces has apparently not led us to be able to identify individual faces as fast as we can tell they are faces at all.

5.3 Multiple levels of face representations: parts before configurations

The first two findings characterize two consecutive face-selective MEG response components and their functional importance. Here I investigated what kind of face information is extracted at each of these stages, by examining the extent to which face parts (eyes, nose and mouth) and face configurations (the placement of these parts in a standard face arrangement) drive these MEG response components. I found that both the

M100 and the M170 are sensitive to face parts whereas only the M170 is sensitive to face configurations. This finding indicates that the extraction of information about face parts starts 70 msec earlier than the extraction of information about face configurations.

Combined with our prior finding that the M100 is correlated with successful categorization while the M170 is correlated with successful face identification, these results also suggest that face parts are more informative in face categorization (consistent with a recent computational simulation, (Ullman, Vidal-Naquet, & Sali, 2002)), whereas face configurations play a special role in face identification (consistent with a large behavioral literature (e.g. McKone, Martini, & Nakayama, 2001; Tanaka & Farah, 1993; Young, Hellawell, & Hay, 1987)).

A viable interpretation of our data is that the M100 and M170 reflect qualitatively distinct stages of processing. However, one might argue that the results are consistent with a different and widely held view according to which the complexity of feature selectivities increases as one ascends along the ventral visual pathway. On this view, complex features of an object are represented by the activity of neurons that receive convergent inputs from populations of neurons that encode relatively simple features of the object at lower levels in the hierarchy (Barlow, 1972; Tanaka, 1996). Fox example, most neurons in monkey V4 are maximally activated by simple texture, whereas a large portion of neurons in the inferior-temporal cortex (TE) is sensitive to moderately complex features (Kobatake & Tanaka, 1994; Tanaka, 1996; Vogels, 1999). In agreement with these findings, a recent fMRI study in the lateral occipital complex (LOC), a probable human homologue of monkey TE, has demonstrated that anterior regions of the

LOC are more sensitive to image scrambling than the posterior regions, suggesting the sensitivity of neurons in the LOC gradually shifts from object parts to a more global representation along the ventral pathway (Lerner, Hendler, Ben Bashat, Harel, & Malach, 2001).

To apply this view to our data, we have to assume that the representation of face parts is simpler than the representation of face configurations. However, this seems unlikely for three reasons. First, neurons selective for face parts (Perrett, Rolls, & Caan, 1982; Perrett, Hietanen, Oram, & Benson, 1992) and neurons selective for face configurations (Perrett et al., 1992; Wang, Tanaka, & Tanifuji, 1996) can both be found in monkey TE, and there is no evidence of hierarchical organization among them. Second, the MEG recordings from humans reported here indicate that the face-selective M170 is sensitive to the abstract form of face configurations, independent of whether they are formed by real face parts or black ovals. Therefore, the representation of face parts is not a necessary intermediate step for the representation of face configurations. Finally, numerous behavioral studies indicate that the processing of face configurations is qualitatively different from the processing of face parts, and these two processes are unlikely to be carried out in a single system (see Chapter 1). Therefore, the fact that face parts are extracted before face configuration apparently cannot be explained in term of increasing complexity with hierarchical and convergent connectivity.

5.4 Remaining Questions

5.4.1 The relation between neural substrates underlying the M100 and the M170

The M170 is thought to arise from the fusiform gyrus (Halgren, Raij, Marinkovic, Jousmaki, & Hari, 2000), where the FFA (fusiform face area) is found with fMRI(Kanwisher, McDermott, & Chun, 1997), but the source location of the M100 is unclear. However, the earlier latency and somewhat more posterior scalp distribution of the M100 compared to the M170 suggest that the two components may not originate from the same anatomical source. If so, does the neural population underlying the M170 receive its inputs directly from that underlying the M100, or do the M100 and M170 reflect parallel processing paths?

The serial processing model would provide a natural explanation of the fact that the M170 seems to "inherit" the response properties of the M100 (i.e., its sensitivity to face parts, and its correlation with successful face categorization), while adding new properties (i.e., a sensitivity to face configurations and face inversion, and a correlation with successful face identification). Further experiments including source localization and other methods will be critical to test this hypothesis.

5.4.2 Face parts versus face configurations

The fact that information about face parts is extracted from faces earlier than information about face configurations suggests that the human visual system does not use either undifferentiated templates of whole faces (e.g. Farah, Wilson, Drain, & Tanaka, 1998) or relational information (e.g. Leder & Bruce, 2000) at the early stage of face processing indexed by the M100. Instead, face parts are apparently encoded independently from face configurations at this stage. In addition, even when the neural mechanisms underlying the M170 encode both face parts and face configurations 70 msec later, no interaction between face parts and face configurations was found, suggesting that face parts and face configurations are processed independently even at the M170 stage. Because a unified face percept presumably requires the eventual integration of information about face parts and face configurations, it is unlikely these two kinds of information will remain separate forever. Further experiments will be conducted to investigate when and how information about face parts and face configurations are integrated.

5.4.3 Faces versus Objects

Which of the findings on face recognition reported in the thesis will hold for the recognition of nonface objects? Given the numerous sources of evidence that faces are "special" (Farah, 1995), one might think some of these findings will not generalize to nonface objects. However, a recent behavioral study on object recognition has demonstrated that the stages of processing for face recognition do not appear to be different from those for the recognition of stimulus classes other than faces (Grill-Spector & Kanwisher, 2001).

5.5 Bruce & Young’s model (1986) revised

Extensive research on face perception over the last few decades has led to several proposals about the organization of the face processing system. For example, based largely on behavioral and neuropsychological findings, Bruce and Young (1986) have proposed a multi–stage cognitive model of face processing. This model includes a structural encoding stage followed by parallel processing streams for the analysis of face identity, emotional expression, and speech-related mouth movements (Figure 1). Recent fMRI studies have confirmed the proposition of separate regions for the processing of identity and emotion (Haxby, Hoffman, & Gobbini, 2000): as face-selective regions in the lateral fusiform gyrus (Fusiform Face Area, FFA) are more sensitive to face identity (George et al., 1999; Tong, Nakayama, Moscovitch, Weinrib, & Kanwisher, 2000) whereas regions in the superior temporal sulcus (STS) are more involved in processing social information present in faces, such as emotional expression, eye gaze and mouth movements (Puce, Allison, Bentin, Gore, & McCarthy, 1998; Hoffman & Haxby, 2000).

However, less is known about the process of structural encoding of faces. In Bruce & Young's model, this stage serves as an entry point of face processing, providing a bottom-up perceptual representation of a face for the analysis of identity and emotional expression at later stages. Although Bruce & Young have noted that face categorization may be an essential first step for face perception, they did not included a distinct face categorization stage in their model. Instead, they assumed that either face categorization may be carried out by a general object categorization system or it may occur simultaneously with face identification (Bruce & Young, 1986). The data presented here argue against both assumptions, suggesting that face categorization may instead constitute a distinct stage of processing in face perception.

In addition to the FFA and the STS, previous fMRI studies have identified another face-selective cortical region in the lateral inferior occipital gyri (Occipital Face Area, OFA) (Kanwisher et al., 1997; Haxby et al., 2000). The OFA is just posterior to the FFA and both posterior and inferior to the STS, as might be expected of it provides inputs to both of these face-selective regions. Given that the M170 presumably arises from the FFA and the scalp distribution of the M100 is more posterior compared to that of the M170, it is possible that the M100 is generated from the OFA.

The fact that information relevant for discriminating face identity is extracted at the M170 stage does not necessarily mean that the perceptual representation of the face has already been matched to faces stored in memory, i.e., recognized, at this stage. Indeed, scalp ERP studies have reported that the N170 (an analogy to the M170 in MEG) is not modulated by the familiarity of faces. Instead, a later negative component (N400) is found to be larger in response to familiar compared to unfamiliar faces (Eimer, 2000a; Eimer, 2000b; Bentin & Deouell, 2000). Therefore, the M100 and the M170 apparently only reflect bottom-up processing, occurring before the stage in which the perceptual representations are matched to memory.

In sum, I suggest that Bruce & Young's model be revised by splitting structural encoding into two stages: an early stage at which face parts are extracted, which may be most important for face categorization, and a later stage at which face configurations are extracted, which may be most important for the identification of individuals (Figure 2).

Both stages would then provide the input to a later recognition stage at which these representations are compared to stored representations in memory in the process of recognition. In addition, the structural encoding stage also provides the inputs to other face processes such as lip–reading and perceiving emotional expression.

Reference List

Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology? *Perception, 1,* 371-394.

Bentin, S. & Deouell, L. Y. (2000). Structural encoding and identificaiton in face processing: ERP evidence for separate mechanisms. *Cognit.Neuropsychol., 17,* 35-54.

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

Eimer, M. (2000a). Effects of face inversion on the structural encoding and recognition of faces. Evidence from event-related brain potentials. *Brain Res.Cogn Brain Res., 10,* 145-158.

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

Farah, M. J. (1995). Dissociable system for visual recognition: A cognitive neuropsychology approach. In S.M.Kosslyn & D. N. Osherson (Eds.), *Visual Cognition* (pp. 101-119). Cambridge: MIT Press.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychol.Rev., 105,* 482-498.

George, N., Dolan, R. J., Fink, G. R., Baylis, G. C., Russell, C., & Driver, J. (1999). Contrast polarity and face recognition in the human fusiform gyrus. *Nat.Neurosci., 2,* 574-580.

Gomez Gonzalez, C. M., Clark, V. P., Fan, S., Luck, S. J., & Hillyard, S. A. (1994). Sources of attention-sensitive visual event-related potentials. *Brain Topogr., 7,* 41-51.

Grill-Spector, K. & Kanwisher, N. (2001). Common cortical mechanisms for different components of visual object recognition: A combined behavioral and fMRI Study [abstract]. *Journal of Vision, 1,* 474a.

Halgren, E., Raij, T., Marinkovic, K., Jousmaki, V., & Hari, R. (2000). Cognitive response profile of the human fusiform face area as determined by MEG. *Cereb.Cortex, 10,* 69-81.

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

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

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

Kobatake, E. & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J.Neurophysiol., 71,* 856-867.

Leder, H. & Bruce, V. (2000). When inverted faces are recognized: the role of configural information in face recognition. *Q.J.Exp.Psychol.A, 53,* 513-536.

Lerner, Y., Hendler, T., Ben Bashat, D., Harel, M., & Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cereb.Cortex, 11,* 287-297.

McKone, E., Martini, P., & Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *J.Exp.Psychol.Hum.Percept.Perform., 27,* 573-599.

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philos.Trans.R.Soc.Lond B Biol.Sci., 335,* 23-30.

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Exp.Brain Res., 47,* 329-342.

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

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382-349.

Tanaka, J. W. (2001). The entry point of face recognition: evidence for face expertise. *J.Exp.Psychol.Gen., 130,* 534-543.

Tanaka, J. W. & Farah, M. J. (1993). Parts and wholes in face recognition. *Q.J.Exp.Psychol.A, 46,* 225-245.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annu.Rev.Neurosci., 19,* 109-139.

Thorpe, S. & Imbert, M. (1989). Biological Constraints on Connectionist Modeling. In R.Pfeifer, Z. Schreter, F. Fogelman-Soulie, & L. Steels (Eds.), *Connectionism in Perspective* (pp. 63-92). Amsterdam: Elsevier.

Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., & Kanwisher, N. (2000). Response Properties of the Human Fusiform Face Area. *Cognitive Neuropsychology, 17,* 257-279.

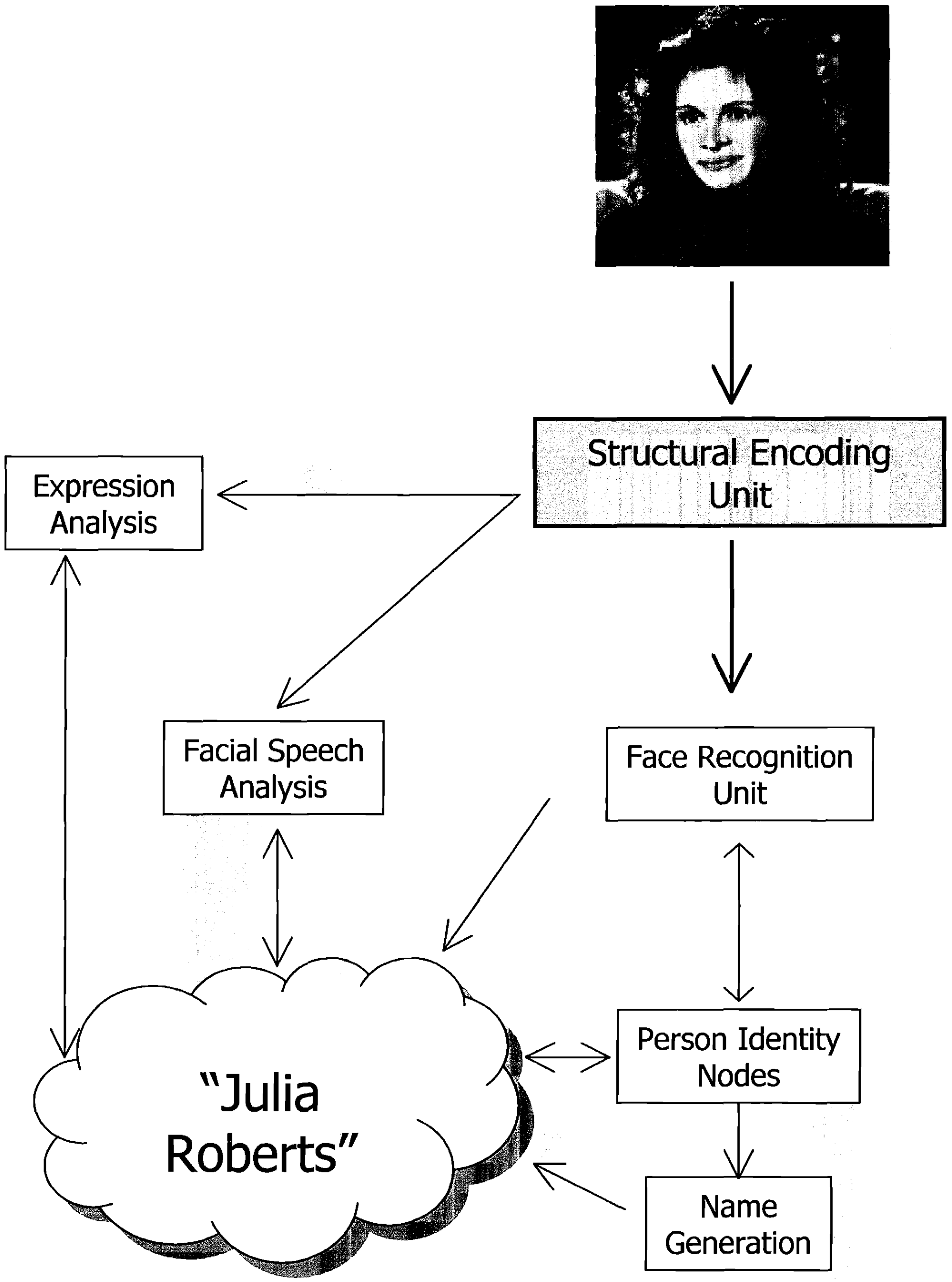
Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nat.Neurosci.*.

Vogels, R. (1999). Effect of image scrambling on inferior temporal cortical responses. *Neuroreport, 10,* 1811-1816.

Wang, G., Tanaka, K., & Tanifuji, M. (1996). Optical imaging of functional organization in the monkey inferotemporal cortex. *Science, 272,* 1665-1668.

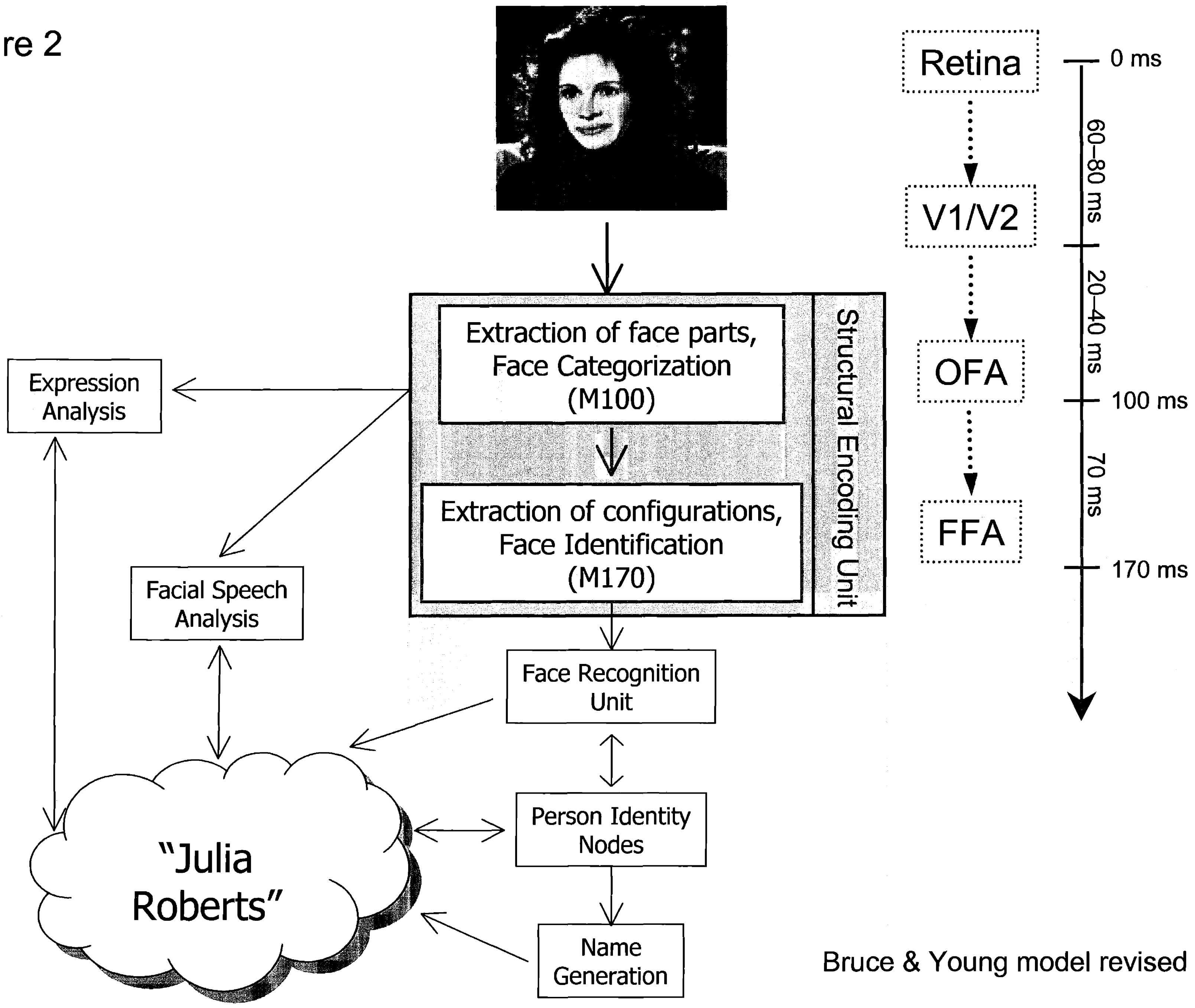
Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16,* 747-759.

Figure 1



Bruce & Young (1986)

Figure 2



Bruce & Young model revised