Aus der Neurologischen Universitätsklinik/

Hertie-Institut für klinische Hirnforschung

Abteilung Kognitive Neurologie

Ärztlicher Direktor: Professor Dr. H.-P. Thier

Sektion Neuropsychologie

Leiter: Professor Dr. Dr. H.-O. Karnath

Human visual processing of natural facial motion analyzed by fMRI

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Medizin

der Medizinischen Fakultät
der Eberhard Karls Universität
zu Tübingen

vorgelegt von

Matthias Brockhaus

aus

Tübingen

2013

Dekan:	**Professor Dr. I. B. Autenrieth**
1. Berichterstatter:	**Professor Dr. Dr. H.-O. Karnath**
2. Berichterstatter:	**Professor Dr. H. Bülthoff**

Contents

Preamble

Nicht das Gehirn denkt, sondern wir denken das Gehirn.
Nietzsche

Dies ist eine wissenschaftliche Arbeit über die visuelle Verarbeitung von Gesichtsbewegungen. Sie soll allen Maßstäben einer wissenschaftlichen Arbeit genügen und zu einem objektiven Ergebnis führen. Dabei versuche ich gegenüber allen Ideen und Hypothesen, denen ich begegne, kritisch zu sein und zu einem unverfälschten Ergebnis zu kommen.

Trotzdem finde ich es wichtig, auf folgenden Punkt hinzuweisen, den schon Jürgen Habermas (Habermas 1968) beschrieb: Meine Persönlichkeit fließt in die Arbeit ein. Ich besitze unreflektierte Grundeinstellungen, die als Ausgangsbasis für meine Tätigkeit dienen. Alle meine Gedanken sind durch diese Grundmeinungen geprägt und dem zufolge ist auch die folgende Arbeit nur die Spitze eines Eisberges meiner Weltsicht, die sich hier in einem spezialisierten, beispielhaften Thema offenbart. Dies alles wird mich von dem Ziel der Objektivität entfernen.

Ich denke, dass es an die Unmöglichkeit grenzt, meine Grundeinstellung zu erfassen. Beispielhaft möchte ich aber aus dem Buch „Auf der Suche nach dem Gedächtnis“ von Eric Kandel zitieren, welche Weltsicht hinter der heutigen Gehirnforschung steht:

Diese neue Wissenschaft beruht auf fünf Prinzipien. Erstens: Gehirn und Geist sind untrennbar. Das Gehirn ist ein komplexes biologisches Organ mit großer Rechenkapazität, das unsere Sinneserfahrungen konstruiert, unsere Gedanken und Emotionen reguliert und unsere Handlungen steuert. Das Gehirn ist nicht nur für relativ einfache Verhaltensweisen wie Laufen und Essen verantwortlich, sondern auch für komplexe Handlungen, die wir für spezifisch

menschlich halten – unter anderem Denken, Sprechen und künstlerisches Schaffen. So gesehen, setzt sich der Geist aus Operationen zusammen, die das Gehirn ausführt, so wie das Gehen sich aus Operationen zusammensetzt, die von den Beinen ausgeführt werden – nur dass die geistigen Operationen unendlich viel komplexer sind.

Zweitens: Jede geistige Funktion im Gehirn – von den einfachsten Reflexen bis zu den kreativsten Akten in Sprache, Musik und bildender Kunst – wird von spezialisierten neuronalen Schaltkreisen in verschiedenen Hirnregionen durchgeführt. Daher sollten wir eigentlich von einer „Biologie der geistigen Prozesse“ sprechen, also jener geistigen Operationen, die von diesen spezialisierten neuronalen Schaltkreisen ausgeführt werden, statt – wie es hier aus Gründen der Einfachheit geschieht – von der „Biologie des Geistes“, was eher ein einziges Hirnzentrum suggeriert, das alle geistigen Operationen vornimmt.

Drittens: Alle diese Schaltkreise bestehen aus den gleichen elementaren Signaleinheiten, den Nervenzellen.

Viertens: Die neuronalen Schaltkreise verwenden spezifische Moleküle, um Signale in und zwischen Nervenzellen zu erzeugen.

Fünftens und letztens: Diese spezifischen Signalmoleküle sind über Millionen Jahre Evolution erhalten geblieben, gewissermaßen „beibehalten“ worden. Einige von ihnen waren in den Zellen unserer frühesten Vorfahren zugegen und sind heute in unseren fernsten und primitivsten evolutionären Verwandten anzutreffen: einzelligen Organismen wie Bakterien und Hefe und Mehrzellern wie Würmern, Fliegen und Schnecken. Um die Bewegungen durch ihre Umwelt zu organisieren, verwenden diese Geschöpfe die gleichen Moleküle, die wir benutzen, um unseren Alltag zu bewältigen und uns an unsere Umgebung anzupassen.

Aus: Eric Kandel „Auf der Suche nach dem Gedächtnis – Die Entstehung einer neuen Wissenschaft des Geistes“, Goldmann-Verlag, S. 12.

Chapter 1

Introduction

1.1 Functional magnetic resonance imaging principles

The first experiments with magnetic resonance imaging (MRI) were realized by F. Bloch and E. M. Purcell 1946 and therefore both were awarded with the Nobel Prize for Physics 1952. After that P. C. Lauterbur and J. M. S. Hutchison displayed 1973 the first anatomical image of a mouse, 1977 R. Damadian researched the first human thorax and thus in the beginnings of the eighties the first patients were examined.

Many nuclei of atomic particles possess a quantum mechanical property called spin that is normally randomly oriented. When an external magnetic field (B_0) is applied, the spins orient themselves according to the magnetic field, i.e. they start to precess around the axis of the magnetic field. This spin system can be represented by a magnetization vector in longitudinal axis (z-direction). The spins precess according to the Larmor frequency (ω), which is related to the field through the gyromagnetic ratio (γ) (i.e. $\omega_0 = \gamma * B_0$).

When a radiofrequency (RF) field of amplitude B_1 rotating synchronously with the precessing spins is applied, the magnetization vector rotates away from its initial equilibrium position by 90 degrees into the transverse (i.e. xy-direction) plane, i.e. the longitudinal magnetization is converted to transverse magnetization. This happens only when the carrier frequency of the FR pulse is equal to the Larmor frequency, hence the term magnetic resonance. While on the transverse plane, the magnetization can be detected by an FR receiver coil.

The application of B_1 not only equalizes the populations of spins in the two energy levels, but also introduces phase coherence among the spins. Coherence decreases quickly as the magnetic moments move out of phase as a result of their mutual interaction. The transverse magnetization, in other

words, is short-lived; it decays exponentially as a result of a process known as relaxation. There are different kinds of relaxation processes known as T1, T2 or T2*, each reflecting different interactions of the spins with their environment or with other spins, and each specified by its time constant (T_i) or its inverse, the relaxation rate ($R_i = 1/ T_i$). These relaxation rates differ depending on the properties of the tissue, and these differences are the basis of image contrast.

In biological sciences, most MRI signals are derived from the hydrogen nuclei of water, as the latter is the most abundant component (80%) of living tissues. The gyromagnetic ratio of protons is 42.58 MHz/Tesla; given that the scanner used in the conducted experiment has a field strength of 3 Tesla, its resonance frequency is of approximately 127 MHz.

Spatial localization is achieved with the use of smaller magnetic field gradients that are superimposed on the homogenous magnetic field of the scanner and by subsequently exploiting the aforementioned Larmor relationship. According to the latter, the positions of protons are encoded by their difference in resonance frequency along the gradient field spanning the scanned volume, with a resolution described in voxels (volumetric pixel).

The functional magnetic resonance imaging (fMRI) utilizes the blood-oxygen-level-dependent (BOLD) contrast, which depicts differences in blood oxygenation. This phenomenon was discovered by Seiji Ogawa 1990 (Ogawa et al. 1990a, 1990b). The usual signal increases reported in BOLD fMRI experiments are due to the fact that neural activation induces a regional increase in cerebral blood flow and glucose utilization that is always larger than the oxygen consumption rate, since oxygen uptake is diffusion limited.
The elevated relation between oxy- and desoxyhemoglobin can be detected because of different magnetic properties: Desoxyhemoglobin is paramagnetic and introduces an inhomogeneity into nearby magnetic field, whereas oxyhemoglobin is weakly diamagnetic and has little effect.
Oxygenated blood leads to a decreasing of phase difference and to a change of

the transverse relaxation time (T2*), thus oxygenation of blood leads to a rising signal of T2*-weighted measurements.
Recent studies using simultaneous recordings of electric and BOLD data provided strong evidence for a high correlation between BOLD signal and electric activity, with a stronger correlation between BOLD and local field potentials than between BOLD and spiking activity (Logothetis et al. 2001).

The time course of the BOLD signal is described by the hemodynamic response function (HRF), shown in Figure 1. The peak response is reached after approximately 4-6 s and the signal returns to baseline after about 25 s. These characteristics constitute the natural restriction of temporal solution of fMRI. However, the stability of the signal across brain regions, subjects and repetitions coupled with the linear additivity of the response to multiple stimuli allow simultaneous recording of BOLD in slices of brain tissue spanning the whole brain and using inter-trial intervals of a few seconds.

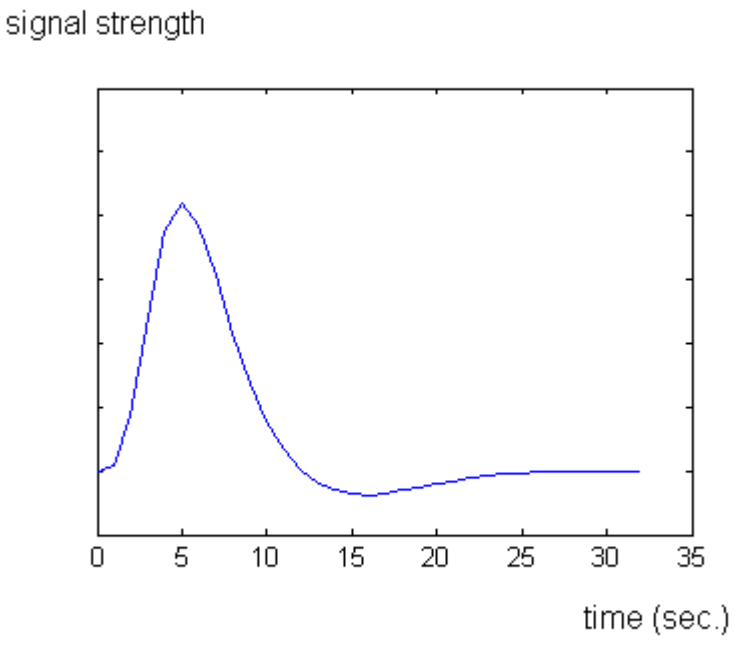


Figure 1:
Schematic display of the time course of the hemodynamic response function (HRF).

1.2 Neural substrates of facial motion perception

1.2.1 The visual system

The visual system is the largest developed perceptive system in the primate brain and its particular importance is readable from its size: Next to the primary visual cortex, which accounts for 30% of the complete monkey cortex, more than 30 other functional areas are described (De Yoe and Van Essen 1988). All together, about 60% of monkey cortex is involved in the perception and interpretation of visual stimulation.

The visual input is preprocessed in the retina, and the information is then transmitted via the optic nerves to the optic chiasm. Here, the axons of the nasal part of the retina cross to the opposite side and join the axons of the temporal retinal part that remain on their original side to form the optic tracts. Therefore, visual information of the left half of our visual field is processed in the right hemisphere and vice versa. Through the lateral geniculate nucleus the information arrives at the primary visual cortex. The primary visual cortex (Area V1 or Brodmann Area 17) is located around the calcarine fissure in the occipital lobe. V1 contains a very well-defined spatial map of the visual information, also called retinotopic organization (Tiao and Blakemore 1978).

Emanating from the primary visual cortex, the further processing is classically assumed to split into two pathways (Ungerleider and Mishkin 1982). These are consistent with the retinal channeling of information consisting of a high-resolution, color-sensitive parvocellular stream and the lower resolution, motion-sensitive magnocellular stream. These pathways are the ”ventral stream” (also known as ”What ?” - system), leading to the ventro-lateral part of the occipital cortex and the ventral part of the temporal lobe, and the ”dorsal stream” (or ”Where?” - system) sending information to centers in lateral temporal lobe and the parietal lobe. The ventral stream goes through area V2, than V4 and leads to the ventral temporal cortex. It is associated with form recognition and object

representation. The dorsal stream goes through V2, then to areas V3a, V7 in the dorsomedial part of the occipital lobe, as well as to hMT/V5 (human middle temporal) and to the posterior parietal cortex. It is associated with motion and representation of object location (“Where” - system). More recent findings show that the dorsal stream is involved in control of eye, arm and hand movements, especially when visual information is used to guide saccades or reaching. The dorsal stream can thus also be considered as the “How” - system (Milner and Goodale 1995). This differentiation is of course a simplification of reality, and being contentious among vision scientists, it is currently under intense scientific scrutiny.

1.2.2 Face perception

Face perception involves many areas of the brain, however some areas have been shown to be particularly important. The fusiform face area (FFA) (Kanwisher et al. 1997), a cortical region in the fusiform gyrus (FG), responds more strongly to faces than to objects like flowers (McCarthy et al. 1997), hands or houses (Kanwisher et al. 1997). Situated at the lateral side of the mid-fusiform gyrus, the FFA shows the most consistent activation in response to faces. Many studies support the idea that FFA activation is triggered by faces and not by control stimuli made of the same low level stimulus features that are present in faces (Kanwisher et al. 1998). Although the FFA shows the strongest increase of blood flow in response to faces, some authors show that the same response magnitude can be obtained in highly-trained subjects that have become “experts” in non-face stimuli (Tarr and Gauthier 2000). Whether the FFA responds to all stimuli in which we are experts, which includes faces, is an ongoing debate. For further information see the review by Posamentier and Abdi (2003) and Tsao and Livingstone (2008).

Furthermore, a face-specific increase in blood flow is also noticed in the superior temporal sulcus (STS) and in the occipital face area (OFA). The OFA is supposed to be sensitive to physical change (Rotshtein et al. 2005) while the

FFA is sensitive to the identity of the faces. In agreement with the current literature on the anatomical basis of prosopagnosia, it is suggested that the FFA and OFA in the right hemisphere and the passing back and forth of information between them are necessary for normal face processing (Rossion et al. 2003). In addition, clinical cases show that the OFA is necessary to perceive faces (Steeves et al. 2006).

1.2.3 Perception of Human Motion

Human motion perception is instantiated in the dorsal stream, passing hMT/V5 and leading to the superior temporal sulcus (STS). The STS also represents a point of convergence for the dorsal and ventral visual stream (Felleman and Van Essen 1991) as it integrates form and motion information arising from the same person (Oram and Perrett 1996; Shiffrar 1994). The processing of relevant or familiar types of biological motion in the STS has been shown repeatedly, e.g. in response to human motion of body (Bonda et al. 1996; Grossman et al. 2000), hands, eye or mouth (Allison et al. 2001). For further information see the review by Blake and Shiffrar (2006).

1.2.4 Perception of facial motion

Humans are highly social creatures relying on the ability to perceive in facial motion physiological status, action, emotion, intention and identity (Bassili 1976; Kamachi et al. 2001; Knight et al. 1997). As mimic is usually not controlled consciously, it can provide a richer source for the state of our partner than verbal communication.

In an influential paper, Haxby and colleagues (2000) proposed a distributed neural system for face processing that distinguishes between the representations of the invariant vs. changeable aspects of faces. The model

proposes a hierarchical processing, beginning with three “core areas” and leading to “extended regions” for different aspects of face perception.

The processing begins in the first core area located in the inferior occipital gyrus, also known as occipital face area (OFA), which provides input to both of the other core areas: (1) the lateral fusiform and (2) the superior temporal sulcus (STS) regions. The former underlies the recognition of identity and the latter the recognition of changeable aspects of faces. For example, the STS (most strongly its posterior part) is important for detecting facial motion due to speech production (Campbell et al. 2001) or natural images of apparent facial motion (Puce et al. 1998; Puce et al. 2003). Additional neural systems are considered as continuing the three core regions, each in charge of special tasks. For example, the STS is connected with areas involved in speech perception from mouth movements and the perception of emotion (Haxby et al. 2000). A separate connection extends from the inferior temporal (IT) cortex to brain areas involved in the retrieval of personal identity, name and biographical information.

Psychological experiments have shown strong evidence that facial motion supports identification of faces (Bruce and Valentine 1988; Hill and Johnston 2001; Thornton and Kourtzi 2002). Two hypotheses exist in the current literature that are not mutually exclusive: (1) The supplemental information hypothesis posits that we represent characteristic facial motion or gestures of an individual face in addition to the invariant structure of the face (Bruce et al. 1988; Knight et al. 1997). These idiosyncratic movements appear to represent dynamic facial signatures (Lander 1999). (2) The representation enhancement hypothesis posits that facial motion contributes to recognition by facilitating the perception of the three-dimensional structure of a face. The rationale behind this hypothesis draws implicitly on structure-from-motion phenomena (e.g. Ullman 1979).

Thus, facial motion does not only provide information about changeable aspects of faces (social communication signals such as speech or expressions) but also about unchangeable aspects of faces, namely identity, through structure-from-motion and idiosyncratic face motion (“dynamic facial signature”). If facial motion supports identification, there must be an interaction between face motion processing areas and identity processing areas, or even directly extraction of identity information from face motion in regions processing face identity.

Two propositions of modification that establish interaction between the two pathways of Haxby’s model exist in the recent literature: (1) O’Toole and colleagues have extended and refined Haxby´s model of face perception (Haxby et al. 2000) to include two ideas (see Figure 2): First, they implemented a second system of identity processing in the dorsal stream that uses dynamic idiosyncratic signatures to recognize identity. Further, they hypothesized that developed structure-from-motion information can be transmitted from the dorsal stream through hMT/V5 to the ventral stream to serve as static form information. (2) Ganel and colleagues (2005) consider facial expression as dynamic variations of the face structure. This would require extracting face expressions using the invariant aspects of the face as a reference. Therefore the processing of facial expression should engage the same anatomical areas traditionally associated with the processing of identity, such as the FFA. They suggest that the involvement could be direct, occurring in the FFA, or that the systems that process identity and expression are interconnected.

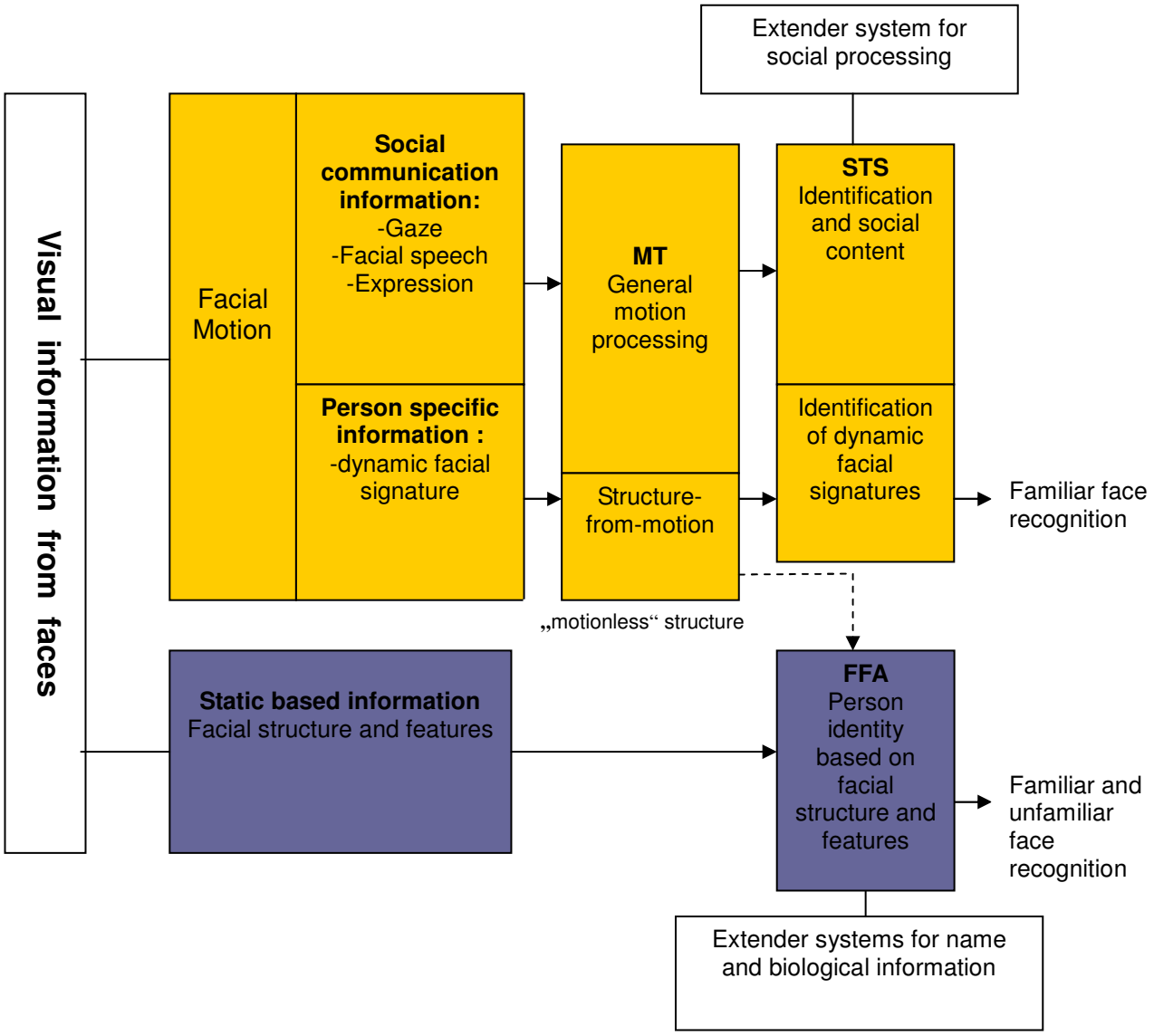


Figure 2:

A reproduction of O'Toole et al.'s (2002) model concerning human recognition of moving faces. It is based on Haxby et al.'s (2000) framework of distributed neural system for face perception. The ventral stream (blue) processes the static structure of a face and the dorsal stream (yellow) processes facial motion. Following Haxby et al.'s model, facial motion contains two different types of information: social communication signs (gaze, expression and facial speech) and person-specific dynamic facial signatures. The social communication information is forwarded to the superior temporal sulcus (STS) and then to the extender systems responsible for specific social tasks.

O'Toole et al. implemented two modifications: (1) Dynamic facial signatures are processed in the STS and can provide a secondary route to face recognition for familiar faces (supplemental information hypothesis). Especially when static information is poor and insufficient for identity recognition we might rely on this secondary system for recognition.

(2) Structure-from-motion could benefit face recognition by communication between the dorsal

and ventral streams. Information from the middle temporal (hMT/V5) visual area could contribute to the structural representation of a face in ventral temporal areas. This added input could benefit recognition for either familiar of unfamiliar faces (representation enhancement hypothesis).

1.3 Purpose of investigation

The present experiment aims to investigate human visual processing of natural, fluid, non-rigid facial motion and the neural basis of their perception.

Until now, facial motion studies mostly utilized non-fluid, unnatural stimuli like implied motion from static images (Puce et al. 1998; Puce et al. 2003), moving avatars, i.e. cartoon faces (Pelphrey et al. 2005; Thompson et al. 2007) or motion stimuli that were produced by morphing a static towards an emotional face (LaBar et al. 2003; Sato et al. 2004; Pelphrey et al. 2007). Using such “unnaturally” moving stimuli might not fully capture the mechanisms underlying the processing of natural facial motion.

The controlled fMRI studies of facial motion that used video sequences of natural facial motion focused on differences between types of face motions and thus did not use non-face control stimuli (Campbell et al. 2001; Hall et al. 2005). A study realized with dynamic face stimuli and non-face stimuli controls (Fox et al. 2009) was not able to directly compare brain activation towards static and dynamic stimuli, because those stimuli were used in different scanning sessions. The authors of this latter study applied two localizer scans, one contrasting static images of faces and objects, the other contrasting dynamic videos of faces and objects, and found that dynamic face localizers are more reliable and more selective than static face localizers.

The first study that directly compared natural, non-rigid face stimuli with static faces and objects were Schultz and Pilz (2009). They found that natural facial motion yielded higher activation than static faces not just in the posterior part of

STS but also in the face-responsive regions FFA and OFA. While this experiment suggested that face motion per se increases face-related activation, the activation increases could have been caused by two potential confounding factors: First, the higher number of different static frames constituting the dynamic stimuli could have increased the activation by stimulating many different cell populations, each sensitive to a different frame of the stimuli. Second, moving faces might attract more attention of the observer as they are more interesting compared to static faces, and this increased attention could have resulted in increased activation.

1.4 Logic

The following work aims to exclude the potential confounding factors of the described experiment from Schultz and Pilz (2009).
To exclude the first possible confounding factor of different frame rates, the following ideas of experimental changes came up: First, moving faces should be compared with frame-scrambled stimuli that contain the same number and rate of frames but lose the frame order of the natural movements. Observing scrambled videos in a pilot test, we wondered if scrambling the frames of the stimuli would not only destroy the percept of natural motion but also create a new percept, with new meaning or content, similar to an “accelerated movie”. This possibility was therefore addressed in a preliminary behavioral experiment in which we tested both normal and scrambled videos at different frame rates. Based on the results of this behavioral experiment, we decided which stimuli to use for the fMRI-Experiment. To exclude potential differences in attentional modulation, we used a task that distracts and controls observers’ attention. A stream of letters was presented in front of the stimuli (rapid serial visual presentation, RSVP) and observers were asked to detect letter repetitions.
In order to ease comparison between our results and the fMRI literature on faces, we decided to study the “classic” face-sensitive regions FFA, OFA and STS, defined using a standardized “face localizer” experiment (e.g. Kanwisher et al. 1997).

Chapter 2

Material and Methods

2.1 Experiment 1: Behavioral experiment

2.1.1 Introduction

The aim of this behavioral experiment is to objectify human perception of stimuli showing fluid, non-rigid facial motion induced by movies made of different numbers of frames, presented in normal and scrambled orders. We used two questions: First, the perceived fluidity of the video. Second, we also wanted to assess whether another movement or another kind of meaning appears by coincidence in the scrambled videos. To assess this possibility, we also asked about the perceived meaningfulness of the stimuli. For example, we hypothesized that the likelihood of emergence of new content from the scrambled stimuli would increase with the frame rate.

2.1.2 Material and Method

To produce the stimuli, we used as a source the video recordings of four male and five female human actors, taken from the Max-Planck-Institute database of moving faces (Pilz et al. 2006; see Figure 3). For these recordings, each face made two expressive gestures in separate videos: surprise and anger. These movie clips consisted of 26 frames recorded at a frame rate of 25 frames per second, for a total duration of 1,040 ms. On the basis of these videos a wide range of stimulus types was generated: 3, 4, 5, 6, 9, 13- and 26-frames, all in natural and scrambled order. While the number of frames was varied, the presentation duration of each frame was adapted so that the total stimulus duration was kept at 1,040 ms for all stimuli.

Ten healthy volunteers from the Tübingen community (24-42 years, mean = 27.7, 4 male) performed two blocks of trials. In each block, the different stimulus types (3, 4, 5, 6, 9, 13- and 26-frames, all ordered and scrambled) were presented 10 times in random order. In one block of trials, subjects had to judge the meaning of the stimuli on a scale from 1 (meaningless) to 8 (very meaningful). In the other block of trials, subjects had to judge the fluidity of the stimuli on a scale from 1 (not fluid) to 8 (very fluid). Block order was randomized across subjects.

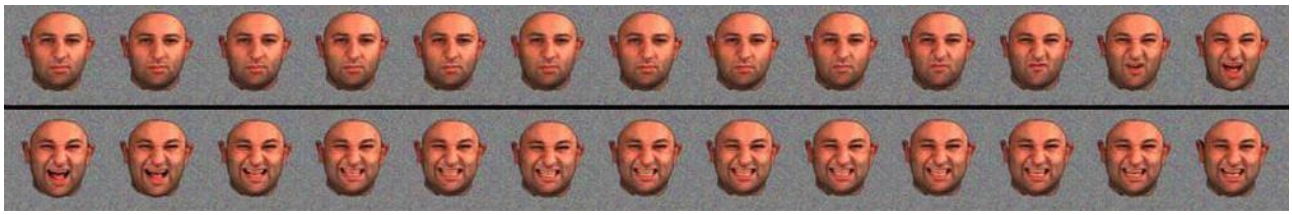


Figure 3:
Example of a video recording from the Max-Planck-Institute database that served as source for our experimental stimuli.

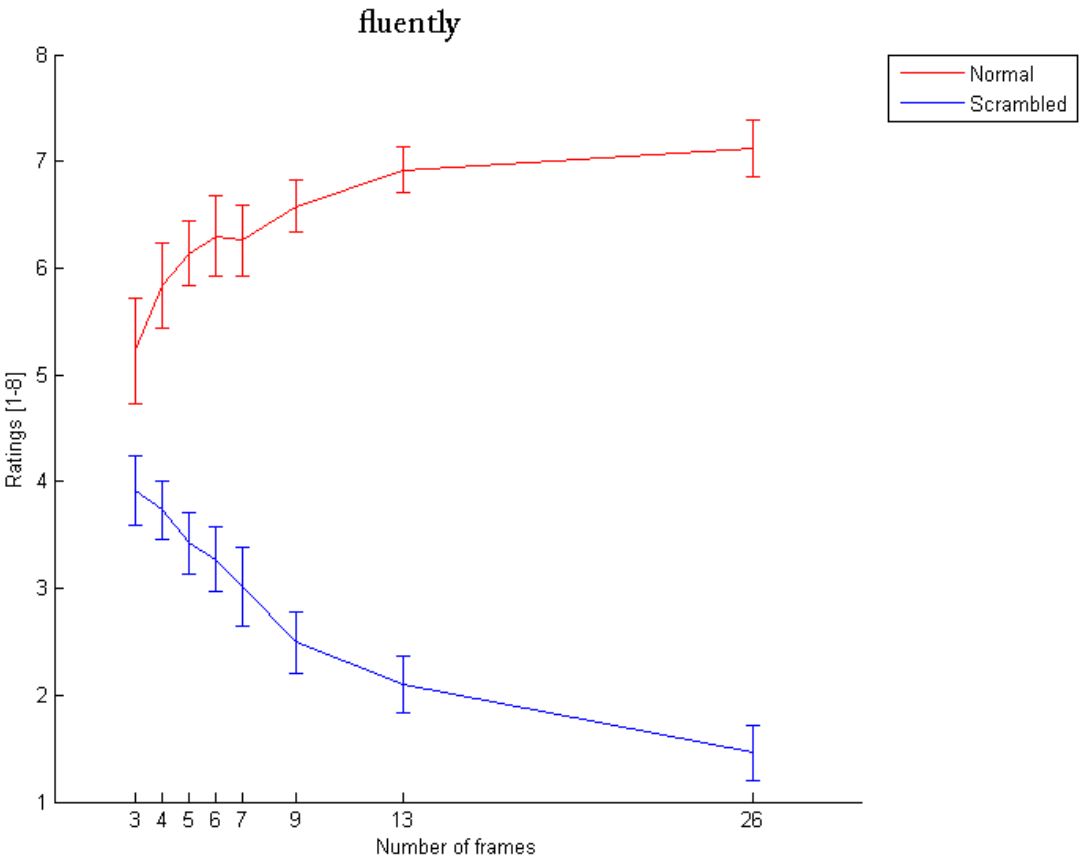
2.1.3 Results

Figure 4 shows subjects' ratings of fluidity and meaningfulness of the stimuli. Subjects reported an increase in perceived fluidity and meaningfulness as a function of frame number for the frame-ordered stimuli, but a decrease for frame-scrambled stimuli (Fluidity ratings: Effect of frame number: $F(7,63)=1.33$, $p>0.2$; effect of frame order: $F(1,9)=107.77$, $p<<0.001$; interaction: $F(7,63)=27.42$, $p<0.001$. Meaningfulness ratings: Effect of frame number: $F(7,63)=5.45$, $p<0.001$, effect of frame order: $F(1,9)=464.84$, $p<<0.001$, interaction: $F(7,63)=31.77$, $p<0.001$. 2-way repeated-measures analysis of variance (ANOVA) performed separately on each type of ratings). The biggest effect of frame-scrambling occurred for 13- and 26-frame stimuli.

Post-hoc tests revealed that the frame-ordered 26-frame stimulus was perceived as significantly more fluid than all the other frame-ordered stimuli except for the 13-frame ordered stimulus (26 ordered vs. 13 ordered:

t(19)=1.22, p>0.2, 26 ordered vs. each of the other ordered stimuli: all t values greater than 3.7, all p values less than 0.003; paired t-tests, bonferroni corrected for N=7 tests, threshold p value = 0.05 / 7 = 0.007).

Post-hoc tests on the meaningfulness ratings revealed that the frame-ordered 26-frame stimulus was perceived significantly more meaningful than 3, 4- and 5-frame ordered stimuli (all t values greater than 3.48, all p values less than or equal to 0.007). All the other frame-ordered stimuli were not significantly less meaningful than the frame-ordered 26-frame stimulus (all t values less than 3.48, all p values greater than 0.014).



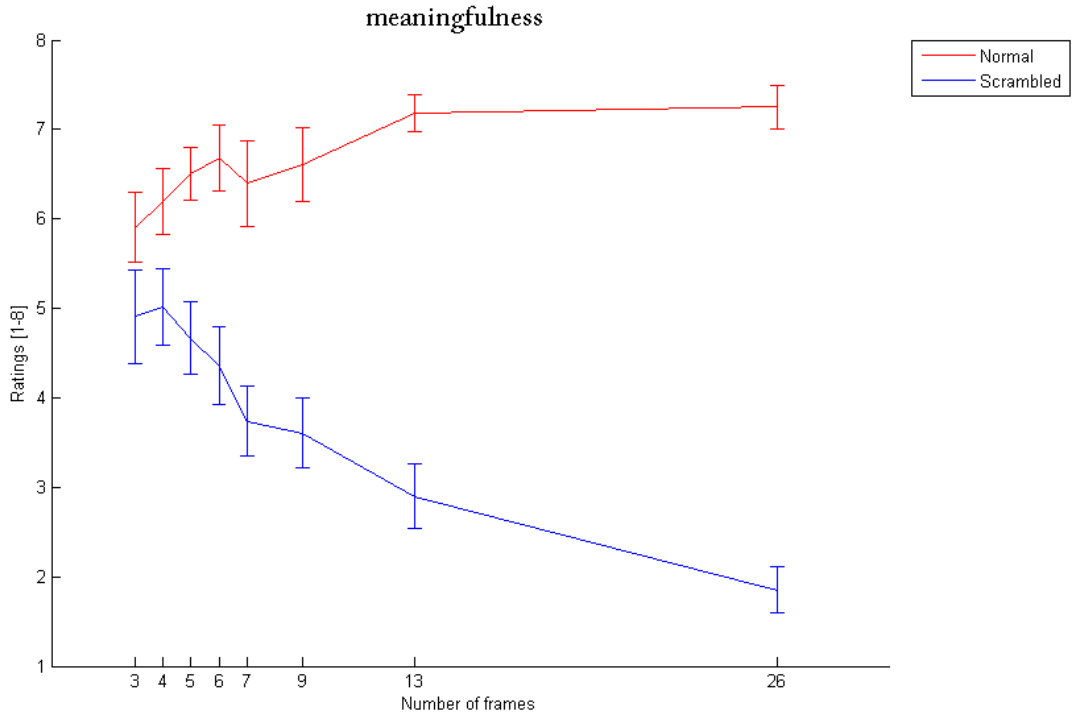


Figure 4:
Results of subjects' ratings on fluidity and meaningfulness of 3, 4, 5, 6, 7, 9, 13- and 26-frame ordered (blue) and scrambled (red) sequences. Error bars represent SEM.

2.1.4 Discussion

The results illustrate a clear, systematic effect: In frame-ordered stimuli, the fluidity and meaning increase with the number of frames, while scrambled videos are evaluated generally as less fluent and meaningful with increasing number of frames. Thus, our hypothesis of new emerging content in scrambled order videos can be excluded.

On the basis of these results, and in order to dissociate the effects on the BOLD signal of static information (number of frames) from the perception of the stimulus, we selected for the subsequent fMRI experiment (1) the original movie stimulus, (2) a stimulus perceived as similar in terms of fluidity and meaning but with a reduced number of frames (13 frames ordered), and (3) a stimulus perceived as less meaningful and fluid with also a reduced number of frames (5

frames ordered). The size of the effect of scrambling on the perceived fluidity and meaningfulness of these stimuli increased as a function of frame numbers (interaction between frame number and frame order), and thus we expected a similar interaction in brain regions processing these stimuli.

2.2 Experiment 2: fMRI experiment

2.2.1 Observers

26 subjects (22-39 years, mean = 26.6, 14 male) from the Tübingen community volunteered for 12 € per hour. All subjects were naive as to the purpose of the experiment, all had normal or corrected-to-normal visual acuity, and had no history of neurological or psychiatric illnesses. All participants provided informed consent and filled out a standard questionnaire approved by the local ethics committee for experiments involving a high field MR scanner, and were informed of the necessary safety precautions.

2.2.2 Face localizer

2.2.2.1 Visual Stimuli

Images for the creation of a localizer of face-sensitive areas where taken from a library of 160 faces and 76 everyday objects from the lab of B. Rossion (http://www.nefy.ucl.ac.be/Face_Categorisation_Lab.htm).
Scrambled faces and objects were generated by Fourier-scrambling as in the previous study (Schultz and Pilz 2009) as follows: Each RGB channel of the static image was Fourier transformed into phase and frequency spectrum. To create the phase-scrambled image, an inverse Fourier transform was performed using the original frequency spectrum and the phase spectrum of an image consisting of pure noise. This generated images in which the frequency spectrum was kept but the phase information was scrambled.

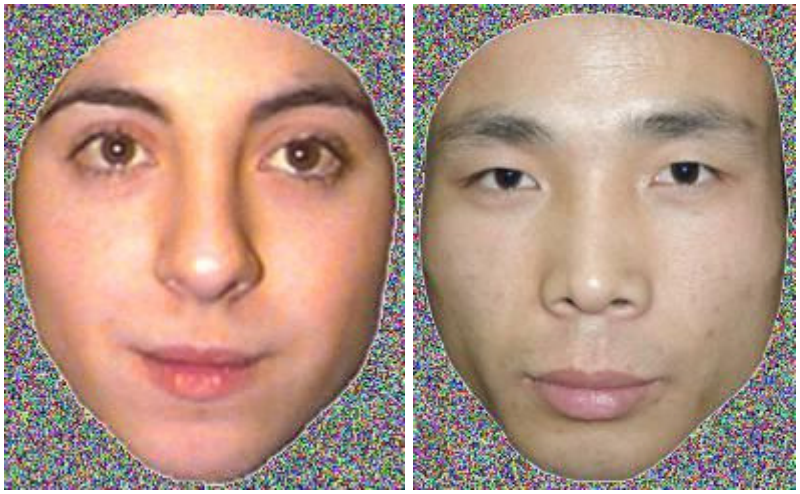


Figure 5:

Two examples of faces used in the face localizer.

Figure 6:

Two examples of everyday objects used in the face localizer.

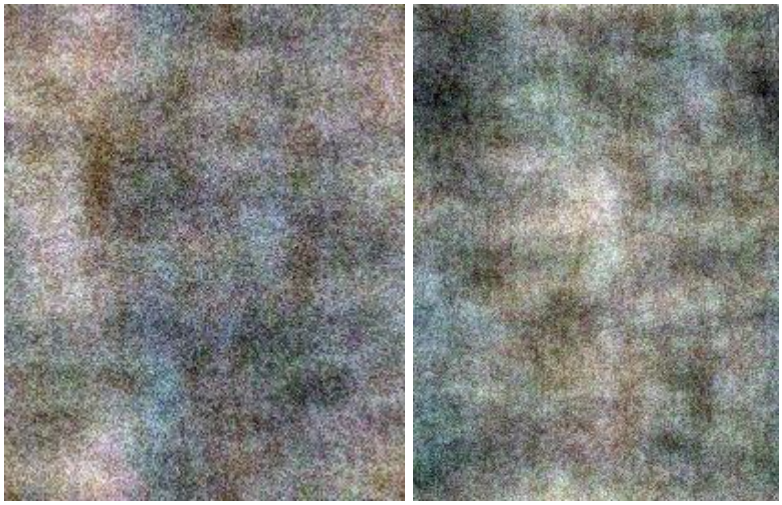


Figure 7:

Two examples of Fourier-scrambled stimuli used in the face localizer.

2.2.2.2 Design and task

To localize the fusiform face area (FFA, Kanwisher et al. 1997) and the occipital face area (OFA, Gauthier et al. 2000) we ran a separate functional localizer experiment (Rossion et al. 2003) as follows. Faces, objects and Fourier-transformed versions of those where presented in a block design, using five blocks for each condition and four blocks of fixation. Each block consisted of 6 stimuli; each presented for 1 s and followed by 2 s of fixation, for a total block length of 18 s. The block order was pseudo-randomized so that the immediate history of all conditions was matched (one-back history matching) in order to counterbalance their influence upon each other (Buracas and Boynton 2002). Subjects' task was to detect image repetitions of any image (one-back repetition detection task, see below) by pressing a button.

2.2.3 Main Experiment

2.2.3.1 Visual Stimuli

For the fMRI experiment we generated 8 different stimulus types (see Figure 8): (a) static: one frame from the video sequence presented for the full duration of the stimulus (1,040 ms), (b) static phase-scrambled: a phase-scrambled version of a), see chapter 2.2.2.1, (c) 5-frame ordered: 5 frames equally spaced in time from the original video (frame numbers 2, 8, 14, 20, 26, each shown for 208 ms), (d) 5-frame scrambled: the frame-scrambled version of c), i.e. the presentation order of the frames was randomized, (e) 13-frame ordered: 13 frames equally spaced in time from the original video (frame numbers 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26) and shown for 80 ms, (f) 13-frame scrambled: frame-scrambled version of e), (g) 26-frame ordered: the original movie clips, (h) 26-frame scrambled: the frame-scrambled version of the original movie clips.

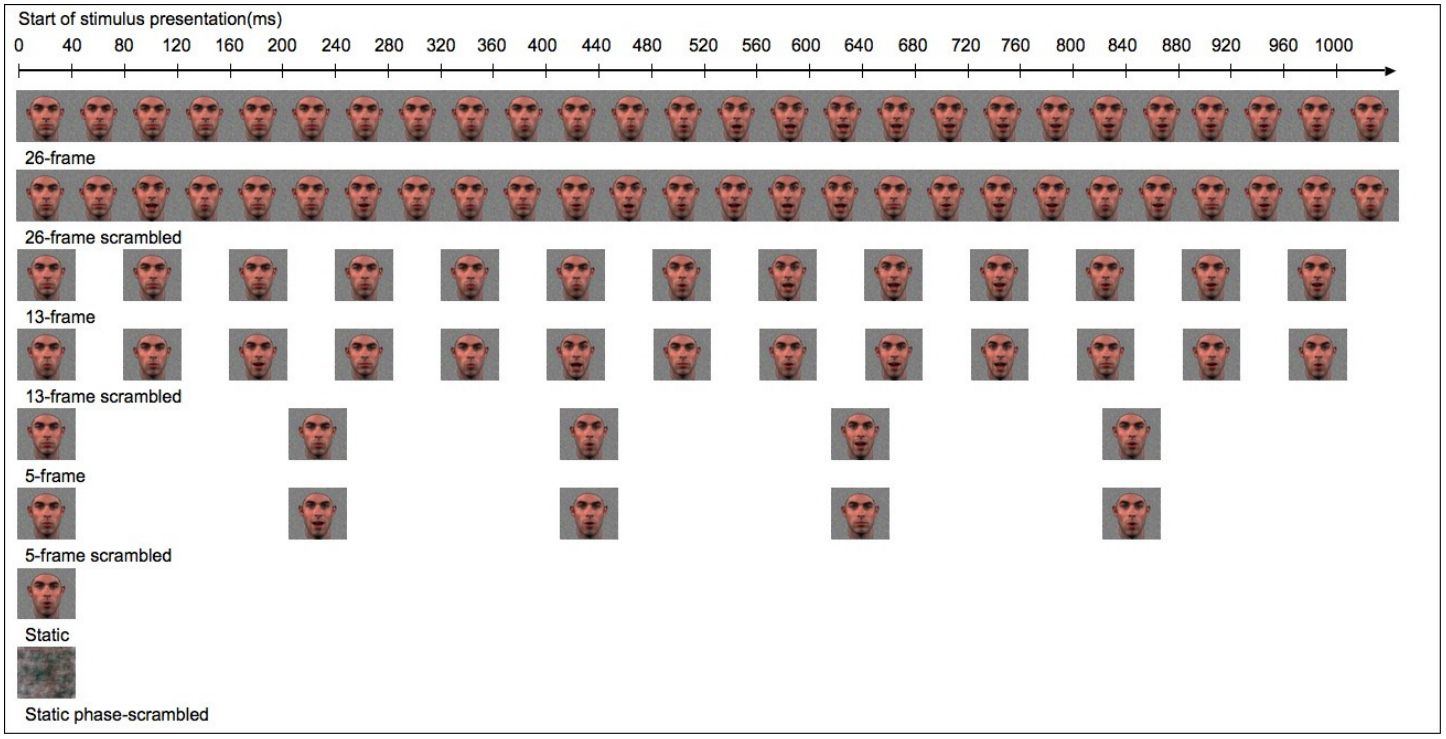


Figure 8:
Stimuli used in the fMRI experiment: 26-frames, 26-frame scrambled, 13-frames, 13-frame scrambled, 5-frames, 5-frame scrambled, static, static phase-scrambled (from top to bottom). Each stimulus was presented for 1,040 ms. The blanks shown between frames in the Figure do not represent interruptions of the stimuli but rather indicate that the same frame as before the blank was still present on the screen.

2.2.3.2 Design and task

In our experiment 9 conditions were used: the 8 types of face stimuli described in section 2.2.3.1 plus a fixation (=rest, no face stimulus) condition. Each trail lasted 2.1 s, in each trial the stimulus was presented for 1,040 ms and followed by an inter-stimulus-interval of 1,060 ms. An event-related design was used, with a pseudo-randomized trial order to increase contrast detection efficiency ("clustered events"; Liu 2004). The experiment was divided into two runs, each containing 25 trials per condition and lasting 8 min.

During stimulus presentation, subject had to perform a one-back repetition detection task on a series of letters that were serially presented at the center of the screen. Letters (in capital Courier font, about 0.15 by 0.2 degrees of visual

angle in size) were changed every 600 ms (presented for 300 ms and followed by a blank for 300 ms). Targets, consisting of a repetition of the previous letter, appeared on average every 25 letters and were spaced between 1 s and 30 s. The reason for the use of this task is as follows: Stimuli with rapid visual changes (e.g. our conditions with stimuli made of 26 frames) might be more salient than stimuli with few changes (e.g. our static stimuli), and these changes in attentional demands might influence underlying neuronal processes (e.g. Bahrami et al. 2007). Our task forces subjects to continuously maintain attention on the centre of the display and not directly on the stimuli, thus keeping the level of attention constant across all stimulus conditions. We purposefully made the task relatively difficult in order to avoid ceiling performance so as to capture eventual differences in performance between conditions.

2.2.4 Technical setup

Observers lay supine on the scanner bed. The stimuli were back projected onto a projection screen situated behind the observers' head and reflected into their eyes via a mirror mounted on the head coil. The projection screen was 140.5 cm from the mirror, and the stimuli subtended a maximum visual angle of approximately 9.0° (horizontal) x 8.3° (vertical). A JVC LCD projector with custom Schneider-Kreuznach long-range optics, a screen resolution of 1,280 pixels x 1,024 pixels and a 60 Hz refresh rate were used. The experiment was run on a 3.2 GHz Pentium 4 Windows PC with 2 GB RAM and an NVIDIA GeForce 7800 GTX graphics card with 256 MB video RAM. The program to present the stimuli and collect responses was written in Matlab using the Psychtoolbox extensions (http://www.psychtoolbox.org) (Brainard 1997; Pelli 1997). We used a magnet-compatible button box to collect subjects' responses (The Rowland Institute at Harvard, Cambridge, USA).

2.2.5 Image acquisition

All participants were scanned at the MR Centre of the Max-Planck-Institute for Biological Cybernetics, Tübingen, Germany. All anatomical T1-weighted images and functional gradient-echo echo-planar T2*-weighted images (EPI) with BOLD contrast were acquired on a Siemens TIM-Trio 3T scanner with an eight-channel phased-array head coil (Siemens, Erlangen, Germany). The imaging sequence for functional images had a repetition time of 1,920 ms, an echo time of 40 ms, a flip angle of 90°, a field of view of 256 x 256 mm and a matrix size of 64 x 64 pixels. Each functional image consisted of 27 axial slices. Each slice had an in-plane resolution of 3.0 x 3.0 mm and a thickness of 3.0 mm, with a 1 mm gap between slices. Volumes were positioned to cover the whole brain based on the information from a 13-slice parasagittal anatomical localizer scan acquired at the start of each scanning session. For each observer, 237 functional images were acquired in a single session lasting approximately for 7.6 min, including an 8 s blank period at the beginning of the run, corresponding to the first four acquired volumes, which were discarded to allow for equilibration of T1 signal. A T1-weighted anatomical scan was acquired after the functional runs [MPRAGE; TR = 1,900 ms, TE = 2.26 ms, flip angle = 9°, image matrix = 256 (read direction) x 224 mm (phase), 176 slices, voxel size = 1 x 1 x 1 mm, scan time = 5.59 min).

2.2.6 fMRI data pre-processing

Prior to any statistical analyzes, the functional images were realigned to the first image and resliced to correct for head motion. A slice time correction was applied so that the data from the 27 frames was corrected to the acquisition time of the 14th frame. The aligned images were then normalized into a standard EPI T2* template with a resampled voxel size of 3 x 3 x 3 mm = 27mm^3 (Friston et al. 1995a). Spatial normalization was used to allow group statistics to be performed across the whole brain at the level of voxels (Ashburner and Friston 1997; Ashburner and Friston 1999). Following

normalization, the images were convolved with an 8 mm full width at half maximum Gaussian kernel to spatially smooth the data. Spatial smoothing was used in this study because it enhances the signal-to-noise ratio of the data, permits the application of Gaussian random field theory to provide for corrected statistical inference (Friston et al. 1996) and facilitates comparisons across observers by compensating for residual variability in anatomy after spatial normalization, thus allowing group statistics to be performed. All of these pre-processing steps were performed using the SPM2 software package from the Wellcome Department of Imaging Neuroscience (http://www.fil.ion.ucl.ac.uk/spm).

2.2.7 fMRI statistical analyzes

Pre-processed fMRI data were analyzed using the general linear model (GLM) framework implemented in SPM2. A two-step mixed-effects analysis was used, as is common in SPM for group analyzes (Friston et al. 1999). The first step used a fixed-effects model to analyze individual data sets. The second step used a random-effects model to analyze the group aggregate of individual results, which come in the form of parameter estimates for each condition and each voxel (parameter maps). As these group statistics are performed at the voxel level, the individual parameter maps need to be in the same anatomical format and were thus computed on the normalized data. For each observer, a temporal high-pass filter with a cutoff of 128 s was applied to the pre-processed data to remove low-frequency signal drifts and artefacts, and an autoregressive model (AR 1 + white noise) was applied to estimate serial correlations in the data and adjust degrees of freedom accordingly. Following that, a linear combination of regressors in a design matrix was fitted to the data to produce beta estimates (Friston et al. 1995b) which represent the contribution of a particular regressor to the data.

2.2.8 Whole-brain analysis

The GLM applied to the individual datasets contained separate regressors of interest for the eight experimental conditions (dynamic faces created by 26, 13 and 5 frames, scrambled dynamic faces by 26, 13 and 5 frames, static face, static scrambled) and the fixation condition. The set of these regressors were created in SPM2 for each of these conditions in the following manner: For each condition, we first modeled the onset and duration of each stimulus as a series of delta functions. The canonical hemodynamic response function (HRF) was implemented in SPM2 as a sum of two gamma functions. The series of delta functions was convolved with the HRF to create separate regressors for each condition. In addition, the GLM included a constant term and six realignment parameters (yaw, pitch, roll and three translation terms). These parameters were obtained during motion correction and used to correct for movement related artifacts not eliminated during realignment. Fitting each subject's data to the GLM, 3D parameter estimate maps for each of our conditions of interest for each subject were produced.

In order to identify which voxels respond to moving faces, we computed the following contrast: 26-frame ordered > static phase-scrambled. This contrast allows identifying regions sensitive to faces, motion or both; it is a very general and broad test which we used to reduce the likelihood of "missing" areas of interest. This approach avoids biasing the results of the subsequent region of interest analysis performed on the result of the whole-brain analysis.

Figure 10 (activations rendered on inflated brain) was created using the spm surfrend
toolbox (http://spmsurfrend.sourceforge.net) and displayed using Neurolens software (http://www.neurolens.org) on the inated template brain from the Freesurfer toolbox (http://freesurfer.nmr.mgh.harvard.edu), and shows activation surviving this threshold using the contrast described above.

2.2.9 Regions of interest analysis

In addition to our whole-brain voxel-wise group analysis; we performed analyzes on individually defined face-sensitive regions of interest (ROI) to investigate the effects of frame-scrambling and number of frames on regions sensitive to facial motion. ROIs were identified using the contrast 26-frame ordered > static phase-scrambled, as follows. We searched in each subject's individual whole-brain analysis for clusters whose peak response was located less than 10 mm away from the 8 peak response of the clusters found in the group analysis. The single-subject GLMs were thresholded at the lower $p<0.05$ uncorrected threshold during this ROI search, because (1) we were looking in regions of a-priori interest which had already survived whole-brain correction in the group ANOVA and (2) to increase the likelihood of finding significant clusters in as many of the individual subjects as possible.

Chapter 3

Results

3.1 Behavioural data collected during fMRI experiment

During the fMRI experiment, subjects performed a one-back repetition detection task (stream of letters presented at the center of the screen) unrelated to the manipulation of interest (the different kinds of face stimuli). Results see Figure 9. Average target detection performance was 68% (SEM=3.45), there were no differences between stimulus conditions ($F(8,200)=1.03$, $p>0.4$; 1-way repeated-measures ANOVA). For the conditions with stimuli made of more than one frame (5-, 13- and 26-frame stimuli, ordered and scrambled), there were no effects of number of frames, frame order or interaction between these factors (all $F<2.2$, all $p>0.12$; 2-way repeated-measures ANOVA). Average response time was 554 ms (SEM = 3 ms), and again, there were no differences between stimulus conditions ($F(8,200)=0.59$, $p>0.7$), and for the multi-frame conditions, no effects of number of frames, frame order randomization or interaction between these factors (all $F<0.7$, all $p>0.4$). These results suggest that attentional resources were distributed similarly between the central task and the face stimuli in all conditions. These results thus reduce the likelihood that differences in brain activation between conditions are due to differences in attention.

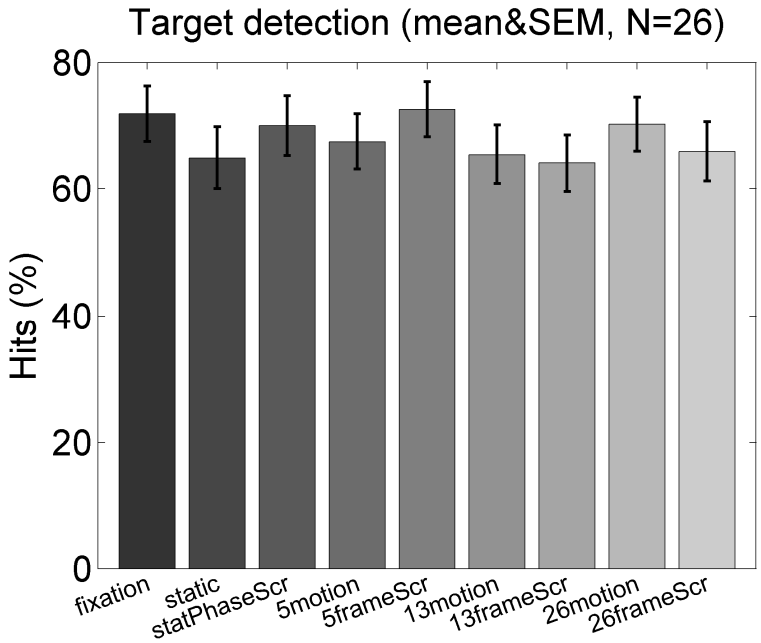


Figure 9:

A rapid serial visual presentation (RSVP) was presented in front of the stimuli and observers were asked to detect random letter repetitions. The figure shows subjects' performance during all type of stimuli, error bars represent SEM.

3.2 Whole brain statistics

Clusters of voxels responding more to 26-frame ordered than to static phase-scrambled were found bilaterally in fusiform gyrus (FG), inferior occipital gyrus (IOG), superior temporal sulcus (STS) and medial temporal lobe (hMT/V5), additional on the right hemisphere in inferior temporal sulcus (ITS) and inferior frontal gyrus (IFG). Figure 10 shows these results and Table 1 the anatomical and statistical details of the peaks of significant activations. Given their anatomical location (see coordinates in Table 1), the clusters in FG and IOG most likely correspond to the fusiform face areas (FFA, Kanwisher et al. 1997) and the occipital face areas (OFA; Gauthier et al. 2000; Hoffman and Haxby 2000). As we did not define these clusters by contrasting faces against objects as was done in the studies defining FFA and OFA, we prefer to use the terms FG and IOG.

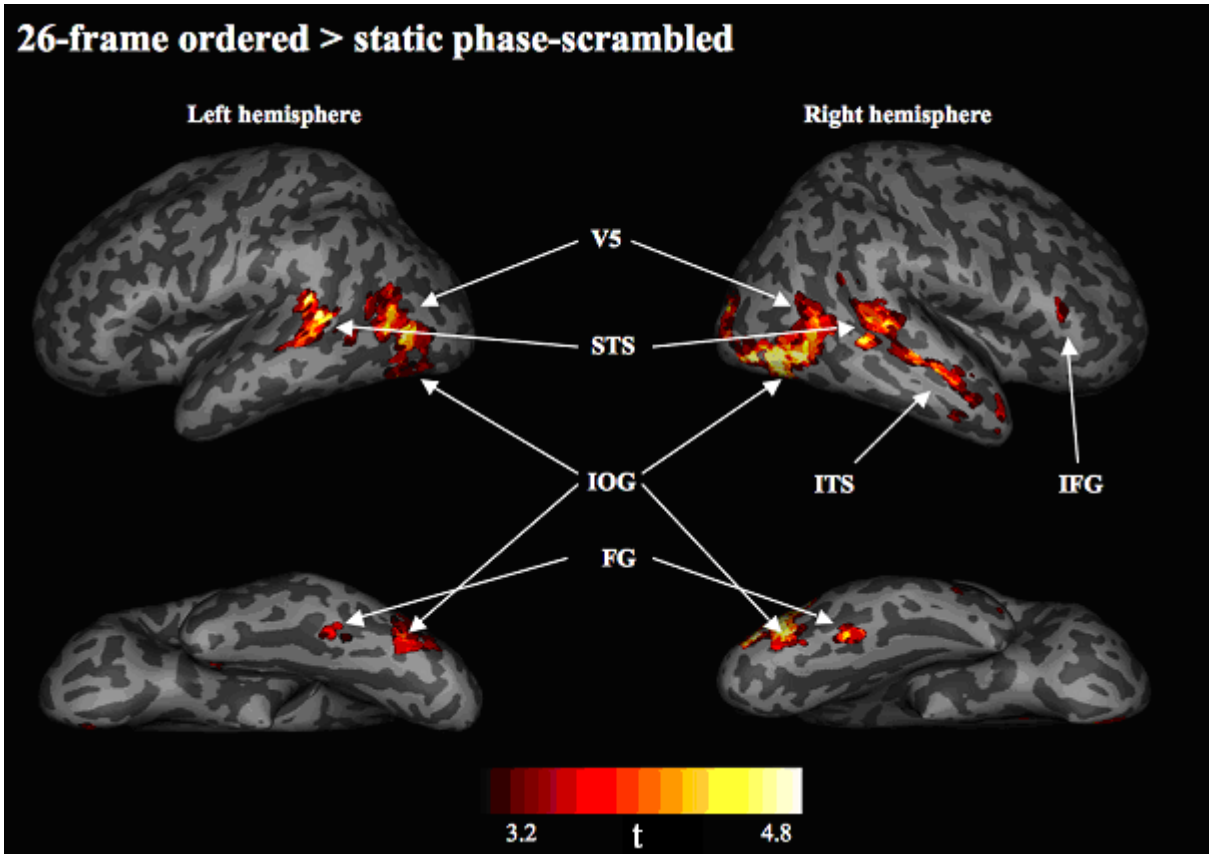


Figure 10:
Results of the whole-brain ANOVA group statistics projected on the surface of an inflated standard structural scan. Parameter estimate maps for 26-frame ordered were computed with static phase-scrambled. This contrast identifies clusters responding to faces, motion or both.

Anatomy	Hemisphere	Coordinates (X,Y,Z)			t	z
fusiform face areas (FFA)	Left	-39	-42	-27	3.29	3.24
	Right	42	-51	-27	5.73	5.51
occipital face area (OFA)	Left	-36	-78	-9	4.17	4.07
	Right	42	-75	-9	6.04	5.78
superior temporal sulcus	Left	-60	-45	9	3.48	3.43
(STS)	Right	51	-63	6	4.36	4.25
middle temporal lope	Left	-57	-60	3	3.29	3.24
(hMT/V5)	Right	42	-51	-27	5.73	5.51
Inferior temporal sulcus (ITS)	Right	57	-3	18	3.57	3.51
Inferior frontal gyrus (IFG)	Right	-3	48	-9	4.25	4.15

Table 1:
Anatomical and statistical details of the peaks of significant activations. All activations survive correction for multiple comparisons across the whole brain. Coordinates indicate local maxima in MNI space. t and z-column, respectively, indicate t-values and z-scores from whole brain ANOVA analysis.

3.3 Individual face-sensitive regions of interest

We located the following ROIs in individual subjects (number of subjects in which we were able to localize the ROIs in brackets): left FG (N=19), right FG (N=20), left IOG (N=25), right IOG (N=24), left STS (N=22), and right STS (N=21). As stated in the previous paragraph, FG and IOG most likely correspond to FFA and OFA, respectively.
BOLD signal changes for the different conditions compared with the control condition (fixation) expressed in percent change of mean activation in each ROI are shown in Figure 11. These data are calculated on the basis of the parameter estimates resulting from the GLM analysis.

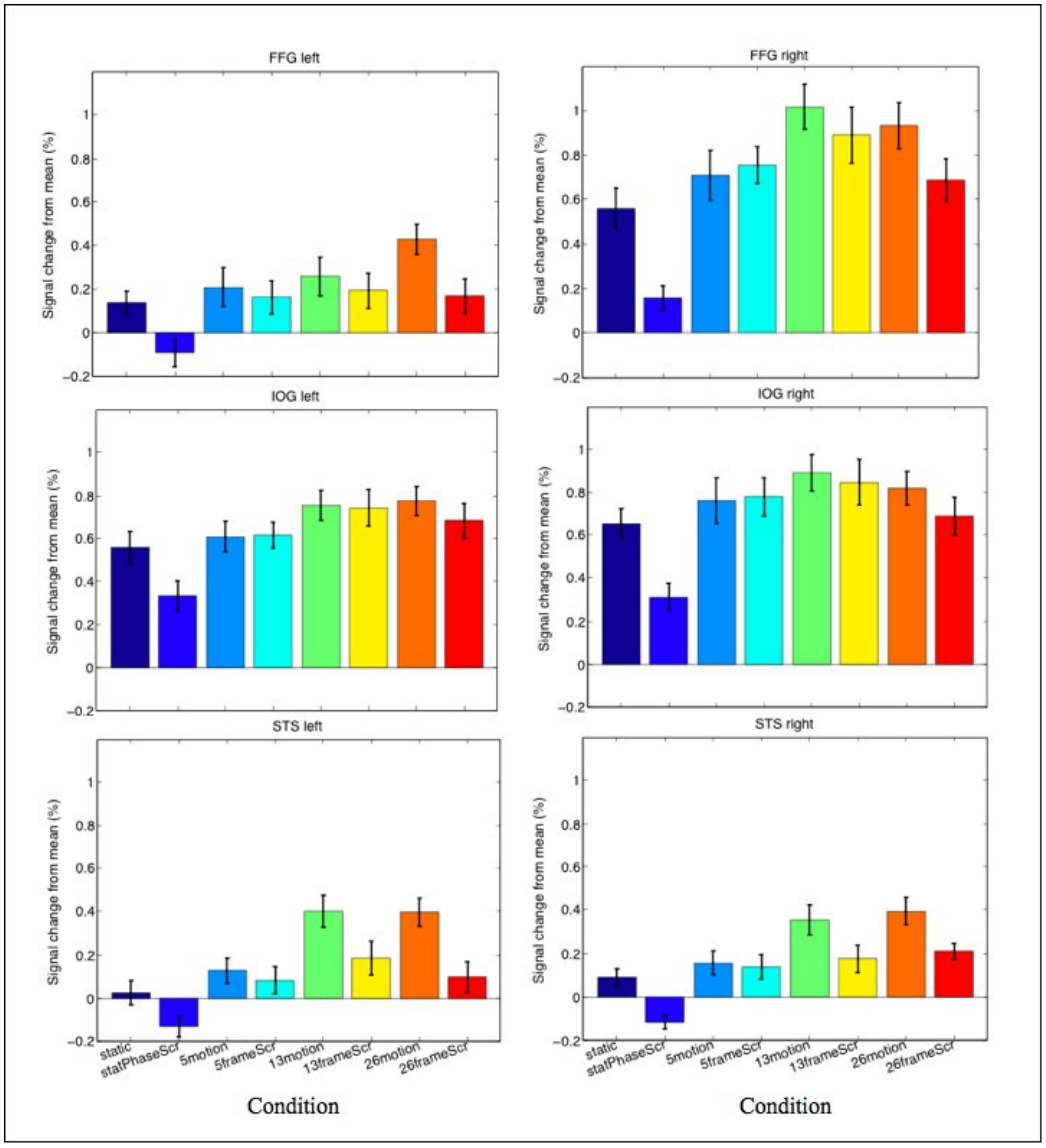


Figure 11:

BOLD signal changes for the different conditions compared with the control condition (fixation) expressed in percent change of mean activation in each ROI. Error bars represent SEM.

We performed separate repeated-measures ANOVAs on each ROI to investigate the effects of frame number and frame scrambling. The design was 3 (frame-number: 5-frame, 13-frame, 26-frame) by 2 (frame-condition: ordered, scrambled). We found significant main effects of frame number in all regions tested (FG, IOG and STS, all bilateral). The main effect of frame order was significant in FG and STS only (all bilateral), but the interaction between frame number and frame order was again significant in all regions tested. Details are provided in Table 2.

2-way repeated measures ANOVA				
ROI	DF	F value		
		Frame number	Frame order	Interaction
FG left	36	3.3*	12.9**	4.8**
FG right	38	3.5*	5.9*	3.3*
IOG left	48	8.1***	3.8	4.1*
IOG right	46	8.5**	3.7	4.9*
STS left	42	6.5**	11.6**	5.1*
STS right	40	6.9**	10.1**	4.5*

Table 2:
Results of the 2-way repeated measures ANOVAs performed on each individual region of interest.
DF describes degrees of freedom for the F-tests used to assess the main effect of frame number and the interaction; values need to be halved for the effect of frame order.
F-values are marked with corresponding p-values: * = $p<0.05$; **=$p<0.01$; ***=$p<0.001$

To investigate the interaction term in further detail we performed post-hoc t-tests on all frame numbers between ordered and frame scrambled condition. We found significant effects for 26-frame ordered vs. 26-frame scrambled in all ROIs. In addition, bilateral STS also showed significant differences between 13-frame ordered and 13-frame scrambled. The difference between 5-frame ordered and scrambled was not significant in any ROI. One-tailed t-tests were bonferroni corrected at a p-value of 0.0033. Results can be seen in Table 3.

Post-hoc t-tests (ordered vs. scrambled)				
ROI	DF	t-value		
		5-frame	13-frame	26-frame
FG left	18	0.2	1.6	4.9***
FG right	19	-0.6	1.9*	3.2**'
IOG left	24	-0.8	-0.3	3.6**'
IOG right	23	-0.6	0.2	3.7**'
STS left	21	-0.5	3.3**'	3.9***
STS right	20	-0.4	3.1**'	3.7***

Table 3:

Pairwise t-test between ordered and frame scrambled stimuli at each frame number in each ROI. t-tests were bonferroni corrected at a p-value of 0.003.

t-values are marked with corresponding p-values: * = p<0.05; **=p<0.01; ***=p<0.001; to mark p-values<0.003 (threshold due to bonferroni correction) we introduced **'=p<0.003

To provide a comparison to the results of the previous study (Schultz and Pilz, 2009), we directly compared the responses to the natural facial motion stimulus (26-frame ordered) and to the static face stimulus (static) with post-hoc t-tests. All regions showed a significant increase in response to the dynamic face stimuli compared with the static face stimuli (see Table 4).

Post-hoc t-tests (26-frame ordered vs. static)		
ROI	DF	t-value
FG left	18	3.8***
FG right	19	4.8***
IOG left	24	4.7***
IOG right	23	3.6***
STS left	21	4.4***
STS right	20	4.1***

Table 4:

Comparison of 26-frame ordered vs. static face stimulus in each ROI.

t-values are marked with corresponding p-values: * = p<0.05; **=p<0.01; ***=p<0.001

Chapter 4

Discussion

4.1 Summary

This study shows increased activation in the core face-sensitive regions of the human brain (STS, FFG and OFA) in response to facial motion. These results thus confirm a previous study (Schultz and Pilz 2009). The present study extends the previous results by controlling for two potentially confounding factors: (1) the increased activation is not due to a higher number of frames or a higher frame rate as this activation increase was not found in response to frame-scrambled control stimuli. (2) To control for an influence of attention on brain activation, we used an accessory task which distracted observer's attention away from the face stimuli. The behavioral results indicate that moving faces did not attract the observer's attention more than static faces or control stimuli.

4.2 Concerning psychological experiments of face perception

Previous studies have shown that facial motion helps to recognize the identity of a face (Bruce and Valentine 1988; Hill and Johnston 2001; Thornton and Kourtzi 2002). Our present study shows that FFA activation increases in response to facial motion. FFA is involved in identity recognition, and although we have not tested if or how subjects learned the identity of our faces, the FFA activation increase we observed could be related to the encoding of face identity information from their motion.

As described in the Introduction, two hypotheses exist about how facial motion helps identity recognition: the representation enhancement hypothesis and the supplemental information hypothesis. The first hypothesis draws implicitly on

structure-from-motion phenomena, which are based on rigid motion information (Ullman 1979). Rigid motion information is almost absent from the stimuli used in the present study: The facial motion provides at most information about only very small face parts, such as the shape of the mouth or the cheeks. This small amount of structure-from-motion information is probably not sufficient to help identity recognition as proposed by the representation enhancement hypothesis.

In contrast, the supplemental information hypothesis is based on the idea that aspects of facial motion are idiosyncratic and thus constitute a kind of dynamic facial signature associated with a particular face identity. This dynamic signature does not necessarily rely on rigid face motion, but can be contained in non-rigid face motion, which constituted most of the motion in our stimuli. Looking at our stimuli, one has the impression that the facial motion could indeed constitute a motion signature associated with the identity of the face. These clues about identity would thus come mostly from non-rigid face information. To conclude, if the activation increase we observed in the FFA is related to encoding of the face identity from the face motion, this would most likely occur through the supplemental information hypothesis.

Unfortunately, it is difficult to distinguish precisely between the two hypotheses in facial motion, as structure-from-motion and dynamic identity signature information are very hard to separate (O'Toole et al. 2002). This is apparent in another kind of stimuli: Studies on the perception of gender from point light walkers (Stevenage and Nixon 1998) have shown that gender judgements are supported both by information about static body structure contained in walker motion (e. g. centre of gravity and shoulder-to-hip ratio) and by characteristically dynamic male and female walking styles (e. g. hip swing). Further work is needed to disentangle the roles of structure-from-motion and dynamic signature information in identity recognition.

The current experiment makes it plausible that a slight increase in FFA activation due to facial motion derives from processing an idiosyncratic gesture as a complement to static identity information. However, both the representation enhancement hypothesis and the supplemental information hypothesis could contribute to explain how information about the face identity is extracted from facial motion. Thus, both hypotheses need to be considered when thinking about the neuronal processing of moving faces.

4.3 Concerning the neuronal basis of face processing

The two main findings of this study are as follows. First, sensitivity of bilateral STS to face motion was confirmed and extended to natural, non-rigid, fluid facial motion. Second, classic static-face-sensitive regions (FFA and OFA) were also shown to be sensitive to face motion, but to a much smaller degree than STS. This second finding shows that even face processing regions of the ventral stream, which are thought to process only static aspects of faces, actually do seem to process facial motion information or receive information about it.

In general, our results are still compatible with separate processing of invariant and changeable aspects of faces (e.g. Haxby et al. 2000) for the following reason: As discussed in the Introduction section 1.2.4, facial motion can convey information about the identity of faces, in addition to facial expressions (Hill and Johnston 2001; Knappmeyer 2001). Thus, the activation increases due to facial motion we observed in the FFA and OFA could be correlates of the identity information conveyed by facial motion, either through structure-from-motion (see O'Toole et al. 2002 for a modification of the Haxby model incorporating a direct input into FFA) or through idiosyncratic facial motion ("dynamic facial signatures").

Results of the current experiment cannot yet address the neural basis of how identity information contained in facial motion serves as identity information in FFA and OFA. In any case, as we observed increased activation in FFA and OFA in response to facial motion, the motion information must somehow arrive in these areas. On the basis of the results of the current study, I propose two possible and not mutually exclusive modifications of O'Toole et al's model to account for this information transfer (see Figure 12).

First, dynamic facial information conveyed in facial motion is sent to STS, which extracts an "idiosyncratic dynamic signature", which is then transmitted to FFA. Second, the motion information itself arrives in FFA (via STS or directly from hMT/V5) and is analyzed there. The current experiment does not allow to test or discriminate these hypotheses.

Both hypotheses are based on O'Toole et al's first modification of Haxby et al's core system (O'Toole et al. 2002) already mentioned in chapter 1.2.4. They argue that facial movements, being idiosyncratic and therefore used to recognize identities, are processed in the dorsal stream. These idiosyncratic movements are preferentially used in suboptimal conditions, when static information is poor and insufficient for identity recognition (Lander 1999; Lander and Bruce 2000).

Our hypotheses are not mutually exclusive with O'Toole et al's second modification of Haxby et al's model: additional information about the static structure of the face is extracted from the face motion (structure-from-motion) and then sent to FFA (O'Toole et al. 2002). As our stimuli might also contain structure-from-motion information, this mechanism might also be involved in our study. However, my impression is that in the stimuli used in the current study, the identity information is mostly conveyed by the idiosyncratic movements and not via structure-from-motion, because rigid face movements appear much smaller than non-rigid movements.

Another open question is the following: which motion-sensitive area feeds forward dynamic facial signature information? So far the STS is assumed as interface. This supposition is based on the possibility of identity recognition from biological motion, both from point-light walkers as well as non-rigid face motion (Cutting and Kozlowski 1977; Hill and Johnston 2001) and on findings associating STS with biological motion processing, including point-light walkers and facial motion (Allison et al. 2000). However, the dynamic facial signature information might also be extracted in another motion-sensitive area, e.g. the middle temporal (hMT/V5). Again, our current results are not able to address this question, however, a recent study using diffusion tensor imaging found little anatomical connections between FFA and STS while the OFA is strongly connected to FFA and STS (Gschwind, Pourtois, Van de Ville, Vuilleumier, Society for Neuroscience abstract 2009).

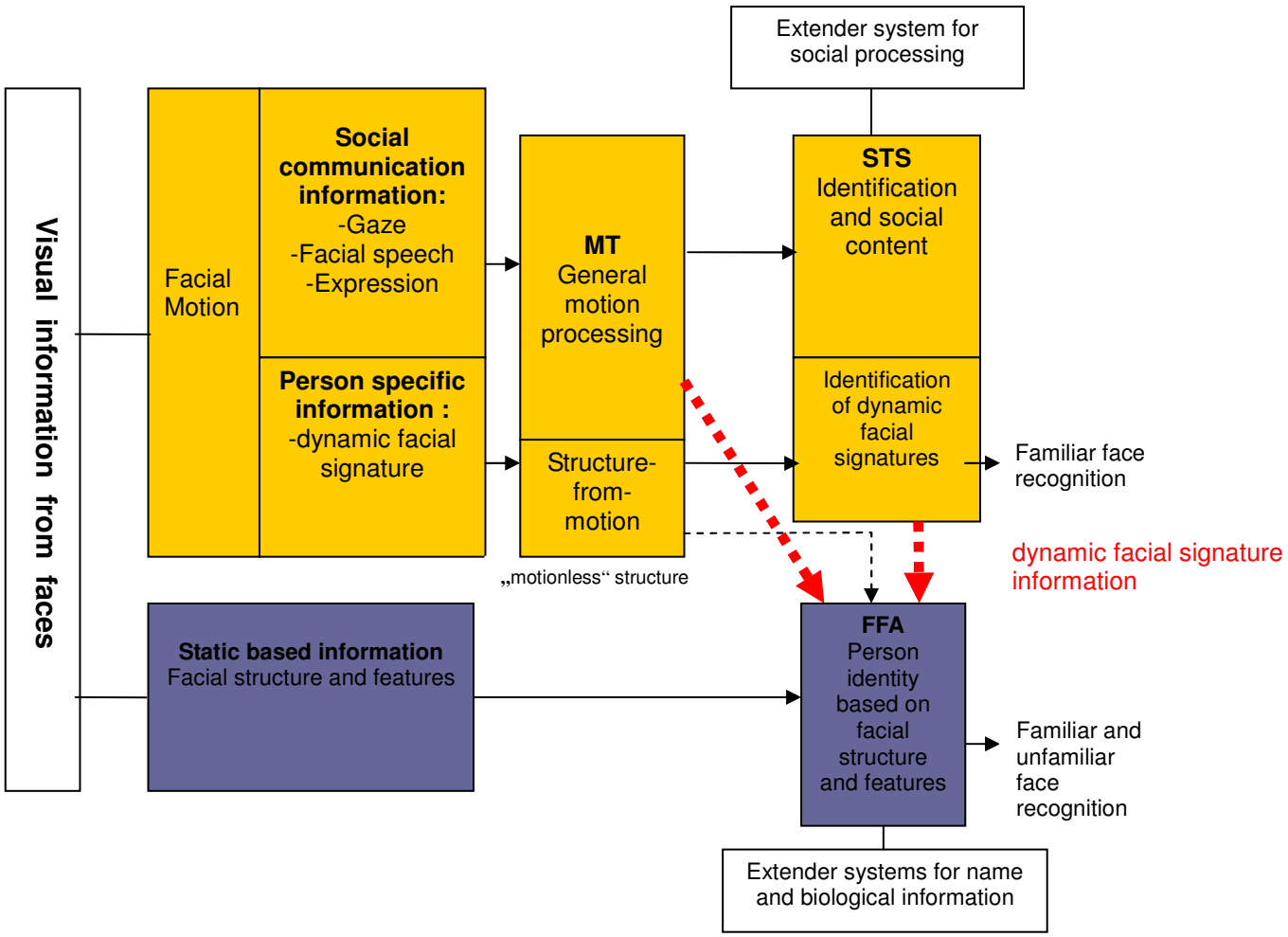


Figure 12:
The model for human recognition of moving faces on the base of O'Toole's review (2002; see also Figure 3). Here, I modified this model by implementing the following hypotheses: dynamic facial information in the sense of an "idiosyncratic dynamic signature" that is conveyed in facial motion and sent to a motion-sensitive area like hMT/V5 or STS, serves as complementary identity information in FFA. This information is either fed back from STS, or another motion-sensitive area like hMT/V5. The dynamic facial signature might be already extracted in the motion-sensitive area, or later in the FFA, which might receive and analyze the raw motion information itself.

4.4 The binding problem in vision

The results of our experiment, current brain-research in general as well as findings of pathologic human brains indicate a deep division of labour and specialized brain regions. But how are information about motion, shape, colour and identity, all communicated by separated nerve tracts, organized into a coherent unitary percept? Semir Zeki points out the problem in his book "A Vision of the brain":

"Yet the common, daily experience of the normal human brain stands forever opposed to the notion of a division of labour and of functional segregation. For that experience is one of wholeness, of a unitary visual image, in which all the visual attributes take their correct place, in which one can register the precise position, shape and colour as well as the direction and speed of motion of a bus simultaneously and instantaneously, as if all the information coming from that bus had been analyzed in one place, in a fraction of a second. Nothing in that integrated visual image suggests that different visual attributes are processed in physically separate parts of our cortex. The task, then, is to enquire into how the brain puts the separate attributes together. An initial step in this enquiry is to study the anatomical opportunities that exist for the specialized visuals areas to "talk" to one another. The task, in brief, is to address the problem of integration anatomically. We shall then see that anatomy, in its usual way, gives powerful clues to how the cerebral cortex might be organized to undertake its integrative functions. Moreover, the strategy used by the visual cortex to achieve integration may give us some insights into the even grander problem of cortical integration in general, for the very same problem has to be addressed when studying the cortex at large: how do the specialized areas of the cerebral cortex interact to provide the integration evident in thought and behaviour."

This author proposes that consciousness, although often described as a unitary percept, is in fact made of several “micro-consciousnesses”, each generated by a different brain region (Zeki 2007). Here, we have the opportunity to compare psychophysics results with fMRI data to be able to compare perception with brain activation. Interestingly, there are differences between the observers’ perception compared with activation in motion-sensitive areas (STS) or face-sensitive areas (FFA and OFA): While FFA/OFA showed only significant differences between 26-frame ordered vs. 26-frame scrambled, STS showed significant differences between 26-frame ordered vs. 26-frame scrambled and also 13-frame ordered vs. 13-frame scrambled. As observers reported 13- and 26-frame stimuli to appear equally fluid and meaningful, perception and brain activation are only similar in motion-sensitive areas. Thus, perhaps unsurprisingly and in concordance with the “micro-consciousness” hypothesis, the neural basis of this motion percept might lie in a region classically related to motion processing (STS).

The activation increase in response to facial motion found in areas sensitive to static faces can likely be ascribed to additional information processing of motion. This activation increase could be a hint for brain connections that allow us to assemble separately processed information about (1) motion and (2) identity of faces into one coherent percept of an identity that moves. This assembling of the percept of an identity evokes a similarity to consciousness, as described above: both could be made of several separate pieces bound together into a unitary experience. An enquiry into this hypothesis could be made in the future.

4.5 Outlook

Experiments from the end of the last decade reveal cues for a more complex understanding of the neuronal substrates of facial motion processing. It emerges that the two classical visual streams could be connected to provide facial motion information for identity recognition. The current work indicates further interactions between invariant and changeable aspects of faces in ventral temporal areas. Future studies will be needed to reveal the exact neural basis of this interaction.

Further scientific investigation of the interaction between facial motion and identity will require an objective measurement of how either structure-from-motion or the supplemental information hypothesis account for identity recognition from facial motion.

Further findings about the neural basis of face perception could help to understand clinical questions like Prosopagnosia.

Lastly, on technical note, our results provide a strong argument for the use of dynamic stimuli to localize areas related to the processing of human faces. Herewith we can support the argument put forward by Fox et al. (2008).

Chapter 5

Conclusion

In summary, the present work shows that dynamic faces elicit more activation than static faces not only in motion-related face-processing areas (STS), but also in form-related face-processing areas (FFA and OFA). Crucially, this increased activation is not due to the number of frames constituting the stimuli. Classic areas responding to the invariant structure of faces are therefore shown to respond more to dynamic than to static faces.

Chapter 6

Summary

Psychological studies have shown strong evidence that facial motion supports identification of faces. Until now, the sensitivity to moving faces in brain regions thought to process identity information was not well known. The present thesis studied how the brain processes natural non-rigid facial motion in direct comparison to static face stimuli. A previous study by Schultz and Pilz (2009) showed that dynamic faces elicit higher responses than static faces in lateral temporal areas (hMT/V5 and STS). Interestingly, that study showed that static-face-sensitive regions FFA and OFA also respond more to dynamic than static faces. The current study pursues this work to exclude potential confounding factors. Previous results were confirmed and specified: In order for this response increase to appear, a correct temporal order of the frames constituting the dynamic face stimuli is required. These results suggest cortical integration of facial motion and identity information and therefore link knowledge about cortical functioning with psychological experiments. For further investigation we suggest to examine a hypothesized transmission of dynamic facial signature information from motion-sensitive areas to the ventral temporal cortex.

Chapter 7

References

Allison T, Puce A, McCarthy G (2000) Social perception from visual cues: role of the STS region. Trends Cogn Sci 4:267–278

Ashburner J, Friston K (1997) The role of registration and spatial normalization in detecting activations in functional imaging. Clin MRI/Dev MR 7:26–28

Ashburner J, Friston KJ (1999) Nonlinear spatial normalisation using basis functions. Human Brain Mapping, vol 7 pp 254–266

Bahrami B, Lavie N, Rees G (2007) Attentional Load Modulates Responses of Human Primary Visual Cortex to Invisible Stimuli. Current Biology 17 509–513

Bassili JN (1976) Temporal and spatial contingencies in the perception of social events. J Pers Soc Psychol 33:680–685

Blake R, Shiffrar M (2007) Perception of Human Motion. Annu. Rev. Psychol. 58:47–73

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

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433– 436

Bruce V, Valentine T (1988) When a nod's as good as a wink: the role of dynamic information in facial recognition. In Practical Aspects of Memory: Current Research and Issues (Vol. 1) (Gruneberg, M.M., eds), pp. 169–174, John Wiley & Sons

Bruyer R, Laterre C, Seron X, Feyereisen P, Strypstein E, Pierrand E, Rectem D (1983) A case of prosopagnosia with some preserved covert remembrance of familiar faces. Brain Cogn 2:257-284

Campbell R, MacSweeney M, Surguladze S, Calvert G, McGuire P, Suckling J, Brammer MJ, David AS (2001) Cortical substrates for the perception of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). Brain Res Cogn Brain Res 12:233–243

Cutting JR, Koslowski LT (1977) Recognizing friends by their walk: Gait perception without familiarity cues. Bulletin of the Psychonomic Society, 9, 353–356.

DeYoe EA, Van Essen DC (1988) Concurrent processing streams in monkey visual cortex. Trends Neurosci 11, 219-26.

Eger E, Henson RNA, Driver J, Dolan RJ (2004) BOLD repetition decreases in object-responsive ventral visual areas depend on spatial attention. J Neurophysiol 92:1241–1247

Felleman DJ, van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex 1:1–47

Fox CJ, Iaria G, Barton JJS (2009) Defining the face processing network: Optimization of the functional localizer in fMRI. Hum Brain Mapp 30:1637–1651

Friston KJ, Ashburner J, Frith CD, Poline JB, Heather JD, Frackowiak RS (1995a) Spatial registration and normalisation of images. Hum Brain Mapp 2:165–189

Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RS (1995b) Statistical parametric mapping in functional imaging: a general linear approach. Hum Brain Mapp 2:189–210

Friston KJ, Holmes A, Poline JB, Price CJ, Frith CD (1996) Detecting activations in PET and fMRI: levels of inference and power. Neuroimage 4:223–235

Friston KJ, Holmes AP, Worsley KJ (1999) How many subjects constitute a study? Neuroimage 10:1–5

Ganel T, Kenneth F, Valyear, Goshen-Gottstein Y, Goodale MA (2005) The involvement of the “fusiform face area” in processing facial expression. Neuropsychologia 43, 1645–1654

Ganel T, Goshen-Gottstein Y (2004) Effects of familiarity on the perceptual integrality of the identity & expression of faces: The parallel-route hypothesis revisited. Journal of Experimental Psychology: Human Perception and Performance 30 pp. 583–597.

Gauthier I, Tarr MJ, MoylanJ, Skudlarski P, Gore JC, Anderson AW (2000) The fusiform “face area” is part of a network that processes faces at the individual level. Journal of Cognitive Neuroscience, 12, 495–504.

Grossman E, Donnelly M, Price R, Pickens D, Morgan V, Neighbor G, Blake R (2000) Brain areas involved in perception of biological motion. J Cogn Neurosci 12:711–720

Habermas J (1968) Erkenntnis und Interesse. Frankfurt a.M, Suhrkamp

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425–30

Haxby JV, HoVman EA, Gobbini MI (2000) The distributed human neural system for face perception. Trends Cogn Sci 4:223–233

Hall DA, Fussell C, Summerfield AQ (2005) Reading fluent speech from talking faces: typical brain networks and individual differences. J Cogn Neurosci 17:939–953

Heeger DJ, Ress D (2002) What does fMRI tell us about neuronal activity? Nat Rev. Neurosci. 3, 142-151.

Hill H, Johnston A (2001) Categorizing sex and identity from the biological motion of faces. Curr. Biol. 11, 880–885

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

Ishai A (2008) Let's face it: It's a cortical network. NeuroImage 40, 415-419.

Jacques C, Rossion B (2007) Early electrophysiological responses to multiple face orientations correlate with individual discrimination performance in humans. Neuroimage 36:863–876

Johansson G (1973) Visual perception of biological motion and a model for its analysis. Percept. Psychophys. 14, 201–211

Kamachi M, Bruce V, Mukaida S, Gyoba J, Yoshikawa S, Akamatsu S (2001) Dynamic properties influence the perception of facial expressions. Perception 30:875–887

Kandel E (2006) Auf der Suche nach dem Gedächtnis – Die Entstehung einer neuen Wissenschaft des Geistes, Goldmann-Verlag, S. 12.

Kanwisher NG, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17:4302–11

Kanwisher N, Tong F, Nakayama K (1998) The effect of face inversion on the human fusiform face area. Cognition 68:B1–11

Knappmeyer B, Thornton I, Bülthoff HH (2001) Facial motion can determine identity. J. Vis. 1, 3 (12), 338a

Knight B, Johnston A (1997) The role of movement in face recognition. Visual

Cogn. 4, 265–273

Kovacs G, Zimmer M, Banko E, Harza I, Antal A, Vidnyanszky Z (2006) Electrophysiological correlates of visual adaptation to faces and body parts in humans. Cereb Cortex 16:742–753

Kurucz J, Feldmar G (1979) Prosopoaffective agnosia as a symptom of cerebral organic brain disease. J. Am. Geriatr. Soc. 27, 91–95

LaBar KS, Crupain MJ, Voyvodic JT, McCarthy G (2003) Dynamic perception of facial aVect and identity in the human brain. Cereb Cortex 13:1023–1033

Lander K, Christie F, Bruce V (1999) The role of movement in the recognition of famous faces. Mem. Cogn. 27, 974–985

Lander K, Bruce V (2000) Recognizing famous faces: exploring the benefits of facial Motion. Ecological Psychology 12, 259 – 272

Liu TT (2004) Efficiency, power, and entropy in event-related fMRI with multiple trial types. Part II: design of experiments. Neuroimage vol. 21 pp. 401-413).

Logothetis N.K., Pauls J., Augath M., Trinath T., Oeltermann A. (2001) Neurophysiological investigation of the basis of the fMRI signal. Nature 412, 150-157

McCarthy G, Luby M, Gore J, Goldman-Rakic P (1997) Infrequent events transiently activate human prefrontal and parietal cortex as measured by functional MRI. J. Neurophysiol. 77:1630–1634

Milner AD, Goodale MA (1995) The visual brain in action. Oxford: Oxford University Press.

Ogawa S, Lee TM, Kay AR, Tank DW (1990a) Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. U S A 87, 9868-9872.

Ogawa S, Lee TM, Nayak AS, Glynn P (1990b) Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. Magn. Reson. Med. 14, 68-78.

Oram MW, Perrett DI (1996) Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. J. Neurophysiol. 76:109–29

O'Toole AJ, Roark DA, Abdi H (2002) Recognizing moving faces: a psychological and neural synthesis, TRENDS in Cognitive Sciences Vol.6 No.6 June 2002

Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437–442

Pelphrey KA, Morris J, Michelich C, Allison T, McCarthy G (2005) Functional anatomy of biological motion perception in posterior temporal cortex: an fMRI study of eye, mouth and hand movements. Cereb Cortex 15:1866–1876

Pelphrey KA, Morris JP, McCarthy G, LaBar KS (2007) Perception of dynamic changes in facial affect and identity in autism. Soc Cogn Affect Neurosci Adv Access

Perrett DI, Oram MW (2003) Neurophysiology of shape processing. Image and Vision Computing 11: 317-333

Pessoa L, Ungerleider LG (2004) Neuroimaging studies of attention and the processing of emotion-laden stimuli. Progress in Brain Research 144, pp. 171–182.

Pilze K, Thornton I, Bülthoff H (2006) A search advantage for faces learned in motion. Exp Brain Res 171: 436–447

Posamentier MT, Abdi H (2003) Processing faces and facial expressions. Neuropsychol Rev vol. 13 (3) pp. 113-43

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

Puce A, Syngeniotis A, Thompson JC, Abbott DF, Wheaton KJ, Castiello U (2003) The human temporal lobe integrates facial form and motion: evidence from fMRI and ERP studies. Neuroimage 19:861–869

Rossion B (2003) A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. Brain, 126:2381-2395

Rotshtein P, Henson RN, Treves A, Driver J, Dolan RJ (2005) Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. Nat. Neurosci. 8:107–13

Rousselet GA, Husk JS, Bennett PJ, Sekuler AB (2007) Single-trial EEG dynamics of object and face visual processing. Neuroimage 36:843–862

Sadr J, Sinha P (2004) Object recognition and random image structure evolution. Cogn Sci 28:259–287

Saenz M, Buracas GT, Boynton GM (2002) Global effects of feature-based attention in human visual cortex. Nat Neurosci, 5, 631-2.

Sary G, Vogels R, Orban GA (1993) Cue invariant shape selectivity of macaque inferior temporal neurons. Science 260, 995–997

Sato W, Kochiyama T, Yoshikawa S, Naito E (2004) Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study. Cogn Brain Res 20:81–91

Schultz J, Pilz KS (2009) Natural facial motion enhances cortical responses to faces. Exp Brain Res 194:465–475

Shiffrar M (1994) When what meets where. Curr. Dir. Psychol. Sci. 3:96–100

Steeves JK (2006) The fusiform face area is not sufficient for face recognition: evidence from a patient with dense prosopagnosia and no occipital face area. Neuropsychologia; 44(4):594-609.

Stevenage S, Nixon MS (1998) Visual analysis of gait as a cue to identity. Appl. Cogn. Psychol. 13, 513–526

Tarr MJ, Gauthier I (2000) FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. Nat. Neurosci. 3:764–69

Tiao YC, Blakemore C (1978) Functional organization of the visual cortex of the golden hamster. J. Comp. Neurol. 168, 459-81.

Thompson JC, Hardee JE, Panayiotou A, Crewther D, Puce A (2007) Common and distinct brain activation to viewing dynamic sequences of face and hand movements. Neuroimage 37:966–973

Thornton IM, Kourtzi Z (2002). A matching advantage for dynamic human faces. Perception, 31(1), 113–132.

Tsao DY, Livingstone MS (2008) Mechanisms of Face Perception. Annu. Rev. Neurosci. 2008. 31:411–37

Ullman S (1979) The interpretation of structure from motion. Proc. R. Soc. Lond. B. 203, 405-426

Ungerleider LG, Mishkin M (1982) Two cortical visual systems. In Ingle, D. J. Ingle, M.A.Goodale, and R.J.W.Mansfield (Eds.), Analysis of Visual Behavior, pp. 549--586.

Zeki S (1993) A Vision of the Brain. Blackwell Scientific Publications, citation: p. 295

Zeki S (2007) A Theory of Micro-consciousness. in Velmans, M. Schneider S. (eds). The Blackwell Companion to Consciousness. Blackwell Publishing

Chapter 8

Appendix

8.1 List of Abbreviations

ANOVA	analysis of variance
BOLD	blood oxygenation level dependent
FFA	fusiform face area
FG	fusiform gyrus
fMRI	functional magnetic resonance imaging
GLM	general linear model
hMT	human middle temporal
HRF	hemodynamic response function
IFG	inferior frontal gyrus
IOG	inferior occipital gyrus
IT	inferior temporal
ITS	inferior temporal sulcus
MNI	Montreal Neurological Institute
MRI	magnetic resonance imaging
OFA	occipital face area
pSTS	posterior superior temporal sulcus
RF	radiofrequency
ROI	region of interest
RSVP	rapid serial visual presentation
SEM	standard error of the mean
STS	superior temporal sulcus
vs.	versus

8.2 List of figures

8.3 List of tables

Chapter 9

Acknowledgements

I owe my deepest gratitude to Johannes Schultz whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject. Thank you for your excellent mentoring, for interesting conversations during image acquisition and for your ability to explain statistics and fMRI-functioning in an understandable manner. Thank you for your great mental balance, your willingness to accept criticism and your collegiality. I will positively remember your enthusiastic way of working. It is a reason that I can imagine staying in science.

I would like to thank Prof. Dr. Dr. Hans-Otto Karnath for his supervision and Prof. Dr. Heinrich H. Bülthoff and the Max-Planck Institute for Biological Cybernetics including many co-workers for making this thesis possible.

I am grateful to Karin Pilz for the helpfulness and gracious cooperation.

I further want to thank Sarah for her professional help with statistics, Jonas for his encouragement to start this work, Christof for his unrestricted personal support and my family for making my whole studies possible.