

What the Human Brain Likes About Facial Motion

Johannes Schultz¹, Matthias Brockhaus², Heinrich H. Bühlhoff^{1,3} and Karin S. Pilz⁴

¹Max Planck Institute for Biological Cybernetics, Tübingen, Germany ²Medical School, University of Tübingen, Tübingen, Germany ³Department of Brain and Cognitive Engineering, Korea University, Seoul, South Korea ⁴School of Psychology, University of Aberdeen, Aberdeen, UK

Address correspondence to Johannes Schultz, Max Planck Institute for Biological Cybernetics, Spemannstraße 38, 72076 Tübingen, Germany.
Email: johannes.schultz@tuebingen.mpg.de.

Facial motion carries essential information about other people's emotions and intentions. Most previous studies have suggested that facial motion is mainly processed in the superior temporal sulcus (STS), but several recent studies have also shown involvement of ventral temporal face-sensitive regions. Up to now, it is not known whether the increased response to facial motion is due to an increased amount of static information in the stimulus, to the deformation of the face over time, or to increased attentional demands. We presented nonrigidly moving faces and control stimuli to participants performing a demanding task unrelated to the face stimuli. We manipulated the amount of static information by using movies with different frame rates. The fluidity of the motion was manipulated by presenting movies with frames either in the order in which they were recorded or in scrambled order. Results confirm higher activation for moving compared with static faces in STS and under certain conditions in ventral temporal face-sensitive regions. Activation was maximal at a frame rate of 12.5 Hz and smaller for scrambled movies. These results indicate that both the amount of static information and the fluid facial motion per se are important factors for the processing of dynamic faces.

Keywords: biological motion, face processing, FFA, fMRI, STS

Introduction

Facial motion is an essential source of information about social interactions between primates. In face-to-face conversations, it conveys information about other people's intentions and emotions (Bassili 1976; Pilz et al. 2006) and it helps deaf people to understand language through lip-reading (Campbell 1992). In addition, facial motion can facilitate the encoding and recognition of facial identity (Lander et al. 1999, 2006; Hill and Johnston 2001; O'Toole et al. 2002; Thornton and Kourtzi 2002; Knappmeyer et al. 2003; Lander and Chuang 2005; Pilz et al. 2006).

It has been suggested that different neural pathways are involved in processing changeable and invariant aspects of faces (Haxby et al. 2000; O'Toole et al. 2002; Kanwisher and Yovel 2006; Ishai 2008; Pitcher et al. 2011). Changeable aspects of faces such as eye gaze, expression, and lip movement are thought to be primarily represented in the posterior part of the superior temporal sulcus (STS) (Allison et al. 2000; Grossman et al. 2000, 2004; Blakemore and Decety 2001; Grossman and Blake 2001, 2002). In contrast, invariant aspects of faces such as facial form and the configuration of facial features are thought to be primarily represented in the ventral temporal cortex in the occipital face area (OFA) (Gauthier et al. 2000) and particularly in the fusiform face area (FFA) (Sergent et al. 1992; Kanwisher et al. 1997; McCarthy

et al. 1997; Gauthier et al. 2000; Haxby et al. 2000; Grill-Spector et al. 2004; Rotshtein et al. 2005; Kanwisher and Yovel 2006). However, recent studies have shown that not only the STS but also the ventral temporal face-sensitive regions respond strongly to moving face stimuli (Fox et al. 2009; Schultz and Pilz 2009; Pitcher et al. 2011). This suggests a stronger interaction between the neural pathways involved in processing changeable and invariant facial information than has been previously assumed.

Despite the importance of facial motion for human behavior and the increased neural response evoked by moving compared with static face stimuli, little is known about what aspects of moving faces actually drive the response to facial motion. First, it could be facial motion itself, that is, fluid deformations of the face occurring over time. For example, it has been suggested that facial motion triggers representations that are directly related to the dynamics of a moving stimulus (Thornton and Kourtzi 2002; Pilz et al. 2006). Secondly, increased response to moving stimuli might be due to an accumulation of evidence, given an increased amount of static information, that is, different frames constituting movies of faces (Perrett et al. 1998). Lastly, as facial motion is very important for human behavior, moving faces might attract more attentional resources than static faces (Franconeri and Simons 2003) and thus evoke an increased blood oxygen level-dependent (BOLD) response (Corbetta and Shulman 2002).

In this study, we investigated the roles of fluid nonrigid facial motion and static information on the BOLD response evoked by moving faces. Participants watched movies of faces with different frame rates, in which the order of frames was either as recorded or scrambled in time. The subjective perception of the fluidity and the meaning of these stimuli were tested in a behavioral study before the functional magnetic resonance imaging (fMRI) experiment. This allowed for testing the following hypotheses: if increased activation in response to moving compared with static faces was due to the greater amount of static information, higher frame rates should evoke a higher neural response and there should be no difference between scrambled and ordered stimuli. In contrast, if the important factor was the fluidity of facial motion, stimuli with correct frame order should evoke a higher activation than stimuli with scrambled frame order. Using this paradigm, we are able to investigate whether the ventral face areas that were previously thought to primarily respond to static facial information, but have recently been shown to respond to moving stimuli, respond more to the fluid motion of a dynamic stimulus or the single static images that are present in a dynamic stimulus. To equate attentional demands across all stimulus types and therefore to exclude possible effects of attention, we

asked participants to perform a difficult task at the center of the screen that was unrelated to the face stimuli.

Materials and Methods

Participants

Ten participants (24–42 years, mean = 27.7, 4 male) from the database of the Max Planck Institute for Biological Cybernetics, Tübingen, Germany took part in the behavioral experiment, and 26 participants (22–39 years, mean = 26.6, 14 male) from the same database participated in the fMRI experiment. None of the participants took part in both experiments. All participants were naive as to the purpose of the experiment, had normal or corrected-to-normal visual acuity, and no history of neurological or psychiatric illnesses. All participants provided informed consent, filled out a standard questionnaire approved by the local Ethics Committee for experiments involving a 3 T MR scanner, and were informed of the necessary safety precautions.

Stimuli

For both behavioral and fMRI experiments, we used video recordings from a database of moving faces, as described in previous studies (Pilz et al. 2006; Schultz and Pilz 2009). All stimuli used in the current studies were derived from that database. Videos were recorded from 4 male and 5 female human actors, each performing 2 expressive gestures, anger and surprise, resulting in a total of 18 videos. Each movie clip consisted of 26 frames recorded at a frame rate of 25 frames per second (i.e. 25 Hz). To study the contribution of static information and fluid nonrigid facial motion on the response evoked by moving faces, we created intermediary stimuli between the movie clips recorded at 25 Hz and a static face. This was done in 2 ways: first, by reducing the frame rate of the movie in several steps (25 Hz = as recorded, 12.5, 9, 6, 5, 4, and 3 Hz). This was implemented by dropping frames from the original stimulus at a regular interval and by increasing the presentation time of the remaining frames. For example, for the 12.5 Hz stimulus, we took every second frame of the original 25 Hz stimulus and presented the frames for twice as long, thus halving the frame rate. Secondly, for each frame rate, we created stimuli with ordered frames as well as frame-scrambled stimuli. All stimuli had a duration of 1040 ms; sample stimuli are shown in Figure 1. Reducing the frame rate reduced the amount of static information in the stimulus and made the motion appear gradually less fluid, so that it appeared like stop-motion. Scrambling the order of the frames also made the motion appear less fluid, but kept the amount of static information constant. How meaningful and how fluid the motion in these stimuli appeared were assessed in a behavioral experiment (discussed subsequently). Based on the results, we selected the stimuli with 25, 12.5, and 5 Hz frame rates in both ordered and frame-scrambled versions for the fMRI experiment. For the fMRI

experiment, we further used a static stimulus, that is, the last frame of each original video, as a reference and a static phase-scrambled stimulus as a low-level control (see method mentioned subsequently).

Phase-scrambled static faces were generated as follows: each RGB channel of the static image was Fourier-transformed into phase and frequency spectrum. An inverse Fourier transform was performed for each channel using the original frequency spectrum and a phase spectrum consisting of white noise: the frequency spectrum was kept but the phase information was scrambled. The same method was used in our previous study (Schultz and Pilz 2009).

Design of Behavioral Experiment

The purpose of this experiment was to compare the meaning and fluidity of the facial motion displayed in the recorded videos and the reduced frame-rate videos, for both the videos with correct and scrambled frame orders. Each participant performed 2 blocks of trials. All stimulus conditions (ordered and frame-scrambled movies at 3, 4, 5, 6, 9, 12.5, and 25 Hz) were presented 10 times, each in a fully randomized order. In 1 block of trials, participants had to judge how meaningful the facial motion in the videos appeared, using a scale from 1 (meaningless) to 8 (very meaningful). In the other block of trials, participants had to judge the fluidity of the facial motion in the videos, using a scale from 1 (not fluid) to 8 (very fluid). Block order was randomized across participants. Based on the results of this experiment, we decided which frame rates to test in the fMRI experiment.

Design of the fMRI Experiment

The fMRI experiment contained 9 conditions: face movies with frame rates of 5, 12.5, and 25 Hz both ordered and frame-scrambled (frame rates were selected based on the results of the behavioral experiment described earlier), the static and static phase-scrambled stimuli, and a fixation condition (no stimulus except a fixation cross). In each trial, a stimulus was presented for 1040 ms, followed by an inter-stimulus interval of 1060 ms, resulting in a total trial length [also known as stimulus onset asynchrony (SOA)] of 2100 ms. An event-related design was used, with a pseudo-randomized trial in order to increase contrast detection efficiency (Liu 2004). We used a pseudo-randomized, event-related design optimized to increase contrast detection efficiency: in each condition, about half the trials were presented in “mini-blocks” of variable duration and half the trials were presented in isolation in order to avoid awareness of the design structure by the participant (Liu 2004). To increase variability of SOAs within each condition, we used 11% of null events or fixation trials (Friston et al. 1999b; Josephs and Henson 1999; Wager and Nichols 2003). We obtained a temporal jitter of SOAs between trials of the same type with a roughly exponential distribution (51% of SOAs were of 2.1 s, 28% between 2.1 and 31.5 s, 17% between 31.5 and 63 s, and 5%

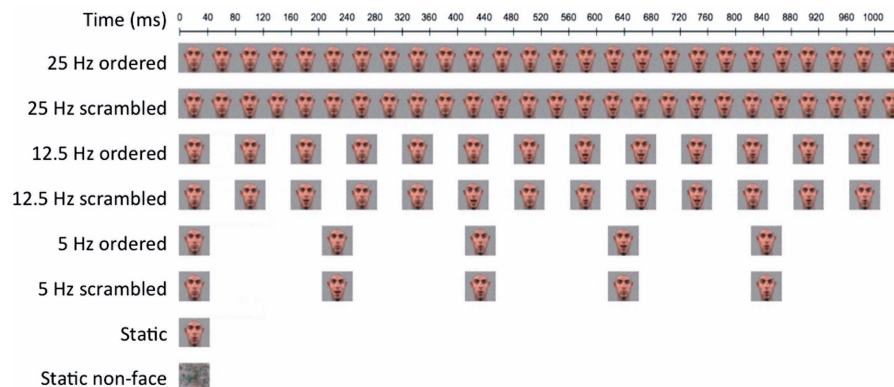


Figure 1. Examples of the stimuli used in the fMRI experiment: moving face stimuli with frame rates of 25, 12.5, and 5 Hz, all presented either with ordered or scrambled frame order, a static face, and a static phase-scrambled face (from top to bottom). Each stimulus was presented for 1040 ms. Images depict the start of the presentation of a frame; stimuli were shown continuously without a blank interval between frames.

between 63 and 168 s). The experiment was divided into 2 runs of 8 min, with each run containing 25 trials per condition.

During stimulus presentation, participants had to perform a 1-back repetition detection task on a series of letters that were presented at the center of the screen. Letters (in capital Courier font, about 0.15° by 0.2° of visual angle in size) were presented for 300 ms, with a blank inter-letter interval of 300 ms. Participants had to press a button whenever the same letter was presented twice in immediate succession. Such letter repetitions appeared on average every 25 letters, that is, every 1–30 s. The presentation of the letters and the occurrence of the targets were completely decoupled from the presentation schedule of the background stimuli. The motivation for using this task was as follows: stimuli with rapid visual changes, such as the facial motion videos used in the current experiment, might be more salient than stimuli with few changes, such as the static stimuli we presented. Hence, attentional demands might be higher for stimuli containing facial motion compared with static stimuli. Such changes in attentional demands might influence underlying neuronal processes (e.g. Bahrami et al. 2007). By using a central task, participants are forced to continuously maintain attention at the center of the display, which keeps the level of task-related attention constant across stimulus conditions. We purposefully made the task relatively difficult in order to avoid ceiling performance. This allowed us to increase the chance of detecting differences in performance between conditions, which might be indicators of differences in attention. Conversely, finding no differences in performance would suggest that participants' attention was relatively similar across stimulus conditions.

Design of Face Localizer fMRI Experiment

To be able to relate the results from the main experiment to classically defined face-sensitive regions, we acquired data in a separate functional localizer experiment. This experiment had 5 stimulus conditions: faces, objects, phase-scrambled faces, phase-scrambled objects, and fixation (no stimulus except a fixation cross). All stimulus images used in the localizer experiment are courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University, <http://www.tarrlab.org/> and were different from the stimuli used in the main experiment. We contrasted static faces with objects since this is the most common approach to identify face-sensitive areas FFA, OFA, and the static face-sensitive parts of STS (Kanwisher et al. 1997; Halgren et al. 1999; Kanwisher and Yovel 2006). Phase-scrambled static faces and objects were intended as alternative control stimuli. We used a block design with 5 blocks per condition. Each block lasted 18 s and was composed of 6 stimuli presented for 1 s every 3 s. Each condition was preceded by every other condition equally often. Participants' task was to detect repetitions of the stimuli, which occurred on average about 10 times per run.

Technical Setup of fMRI Experiment

Participants lay supine on the scanner bed. The stimuli were back-projected onto a projection screen situated behind subjects' head and were reflected into their eyes via a mirror mounted on the head coil. The projection screen was set up 140.5 cm away from the mirror, and the stimuli subtended a maximum of approximately 9.0° (horizontal) \times 8.3° (vertical) visual angle. A JVC LCD projector with custom Schneider-Kreuznach long-range optics, a screen resolution of 1280×1024 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 Psychtoolbox extensions (<http://www.psychtoolbox.org>; Brainard 1997; Pelli 1997; Kleiner et al. 2007). Participants' responses were collected using a custom-made magnet-compatible button box.

MR Data Acquisition

All participants were scanned at the MR Centre of the Max Planck Institute for Biological Cybernetics, Tübingen, Germany, using a

Siemens TIM-Trio 3 T scanner with an 8-channel phased-array head coil (Siemens, Erlangen, Germany). All anatomical T1-weighted and functional images were acquired during the same scanning session. The functional images were gradient-echo, echo-planar T2*-weighted images (EPI) with BOLD contrast, for which the imaging sequence had a repetition time of 1920 ms, an echo time of 40 ms, a flip angle of 90° , a field of view of 256×256 mm, a matrix size of 64×64 pixels, and an in-plane voxel size of 3.0×3.0 mm. Each functional image consisted of 27 axial slices. Each slice had a thickness of 3.0 mm with a 1 mm gap between slices. Volumes were positioned to cover most of the brain (in some participants, data from 1 cm of the most dorsal extent of the parietal and frontal lobes were missing), based on the information from a 13-slice parasagittal anatomical localizer scan acquired at the start of each scanning session. For each subject, 257 functional images were acquired in each of the 2 experimental runs. Each run lasted for about 8 min including an 8 s blank period at the beginning of the run. In addition, we acquired a further 242 functional images per participant in the separate face localizer scan. The first 4 images of each run were discarded to allow for equilibration of the T1 signal. A T1-weighted anatomical scan was acquired after the functional runs [magnetization-prepared rapid gradient echo; TR = 1900 ms, TE = 2.26 ms, flip angle = 9° , image matrix = 256 (read direction) \times 224 mm (phase), 176 slices, voxel size = $1 \times 1 \times 1$ mm, scan time = 5.59 min].

fMRI Data Analysis

Preprocessing

Prior to any statistical analysis, the functional images were realigned to the first image and resliced to correct for head motion. A slice-time correction was applied to the data so that the acquisition time of the 27 slices was synchronized to the acquisition time of the middle (14th) slice. The functional images were normalized to MNI space in the following 3 steps. First, the anatomical T1 image was coregistered with the aligned functional images (little correction was needed as both kinds of images were acquired in the same scanning session). Secondly, the T1 images were normalized to MNI space. Thirdly, the functional images were normalized to a standard EPI T2* using the parameters obtained in step 2 and resampled to a voxel size of $3 \times 3 \times 3$ mm = 27 mm³ (Friston et al. 1995a). This approach yields a better normalization quality than using T2* images alone. Spatial normalization was used to allow group statistics to be performed across the whole brain at the level of voxels (Ashburner and Friston 1997, 1999). Following normalization, the images were convolved with a 6 mm full width at half maximum Gaussian kernel to spatially smooth the data. Spatial smoothing was used because it enhances the signal-to-noise ratio of the data, permits the application of the Gaussian random field theory to provide for corrected statistical inference (Friston et al. 1996; Worsley et al. 1996), and facilitates comparisons across participants by compensating for residual variability in anatomy after spatial normalization, thus allowing group statistics to be performed.

Statistical Analyses

Preprocessed fMRI data were analyzed using the general linear model (GLM) framework implemented in the SPM2 software package from the Wellcome Trust Centre for Neuroimaging (<http://www.fil.ion.ucl.ac.uk/spm>). For the face localizer experiment, we used a fixed-effects model to analyze individual data sets. For the main experiment, we used a 2-step mixed-effects analysis, as is common in SPM for group analyses (Friston et al. 1999a). 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).

Whole-Brain Analysis

First-level, fixed-effects models were created for the face localizer data as well as for the main experiment data for each participant. We applied a temporal high-pass filter with a cutoff of 128 s to the preprocessed data to remove low-frequency signal drifts and artifacts,

and an autoregressive model (AR 1 + white noise) was applied to estimate serial correlations in the data and to 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. The GLM applied to the individual data sets contained separate regressors of interest for each condition. These regressors were created in SPM2 as follows: the onset and duration of each stimulus were modeled as a series of delta functions, and the resulting time series of predicted neural events was then convolved with the canonical hemodynamic response function (HRF) to create the regressors. The HRF was implemented in SPM2 as a sum of 2 gamma functions. In addition, the design matrix included a constant term and 6 realignment parameters (yaw, pitch, and roll and X-, Y-, and Z-axis translation terms). These parameters were obtained during motion correction and were used to correct for movement-related artifacts that were not eliminated during realignment. By fitting each subject's data to the GLM, 3D parameter estimate maps for each of our conditions of interest were produced.

For the main experiment, single-subject parameter maps were imported into SPM2's analysis of variance (ANOVA) model to evaluate group statistics (random effects) for the following contrasts: 25 Hz movies with ordered frame > static face, increase proportional to frame rate (each condition was weighted by the frame rate: 1, 5, 12.5, and 25; ordered and scrambled conditions were weighted similarly, and the contrast was mean-centered), all ordered > all frame-scrambled stimuli (averaging over frame rates), and the interaction between frame rate and frame order testing for regions with a greater scrambling effect at higher frame rates. SPM2 uses the Greenhouse-Geisser correction for nonsphericity in the data. The results were thresholded so that all clusters survived multiple corrections across the whole brain at $P < 0.05$ and all voxels therein survived an uncorrected threshold of $P < 0.001$ (Friston et al. 1994). Results are shown in Figure 3, rendered on an inflated template brain from the Freesurfer toolbox (<http://surfer.nmr.mgh.harvard.edu>) using the `spm_surfrend` toolbox (<http://spmsurfrend.sourceforge.net>) and displayed using the NeuroLens software (<http://www.neurolens.org>).

Regions of Interest (ROIs)

In addition to our whole-brain voxel-wise group analysis, we assessed the response of specific ROIs in detail. First, we were interested in neocortical brain regions responding to moving face stimuli. To avoid biasing our selection toward motion or faces, we selected these ROIs using the contrast 25 Hz ordered frames than static phase-scrambled faces, calculated on the data of the first run of the main experiment. To be able to relate our results to responses in classic face-sensitive areas FFA and OFA, we separately identified those using the contrast static faces than static objects calculated on the data of the separate face localizer scan, as described earlier. We defined the ROIs as functional masks resulting from activated clusters, using the threshold of

$P < 0.005$ uncorrected in order to identify these ROIs in as many participants as possible. The number of participants in which we found these regions and the corresponding average ROI coordinates are reported in Table 2. Within each functionally defined ROI, we collected the BOLD signal data obtained under different conditions of the main experiment, averaged the data over voxels, and calculated the size of the response as percent signal change from fixation. We investigated the effects of frame rate, frame order, and the interaction between these 2 factors using 2-way repeated-measures ANOVA. We corrected for false positives by correcting the P -values for multiple tests using the Holm-Bonferroni method (10 ROIs \times 3 contrasts per ANOVA = 30 tests). Further, we tested whether conditions with stimuli containing more than 1 frame evoked a higher activation than static faces using 3 t -tests per ROI. Again, P -values were corrected for multiple tests using the Holm-Bonferroni method (N t -tests = 30). For the ROIs defined using the data of the first run of the main experiment, we only used the data of the second run of the main experiment to perform the just-mentioned analyses.

Finding the Optimal Frame Rate for Each ROI

To identify the optimal stimulus frame rate for each ROI, we proceeded as follows. For each ROI, we fitted the individual responses measured at different frame rates (1, 5, 12.5, and 25 Hz) with a Gaussian function and then determined the peak of the fitted function. We used a Gaussian function because this is a robust method that uses all data points along a curve and not only those lying near the peak location, which is why Gaussian functions are frequently used to determine the peak of data sets that roughly follow an inverted U-shape, for example, the BOLD response (Krugger and von Cramon 1999). This fitting method allowed a good fit of different response profiles observed in the different participants and ROIs (including inverted U and linear increases or decreases). Peak estimations obtained with the Gaussian agreed well with our own evaluations based on visual inspection. However, our results are not dependent on the fitting function, as almost identical results were obtained when data were fitted with a second-order polynomial function. Both methods have been used successfully to determine the temporal response profile of brain regions in a previous study (McKeeff et al. 2007). We performed the fitting twice for each ROI: once using the stimuli with ordered frames and once using the frame-scrambled stimuli. This allowed us to assess whether the optimal frame rate changed depending on the ordering of the frames.

Results

Behavioral Experiment

In this behavioral experiment, we assessed how fluid and meaningful the facial motion displayed by stimuli with

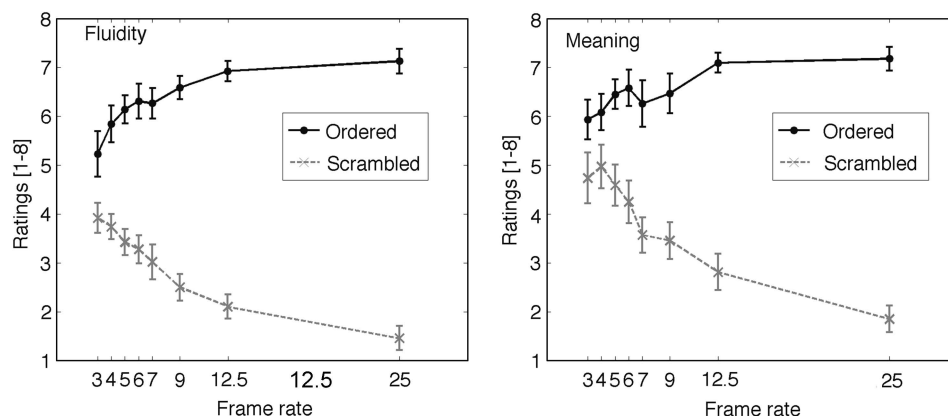


Figure 2. Participants' ratings on the fluidity (left) and meaning (right) of the face motion as a function of the frame rate of the video stimuli (3, 4, 5, 6, 7, 9, 12.5, and 25 Hz) and the frame order (ordered or scrambled). Error bars represent standard errors of the mean across participants.

various reduced frame rates appeared, when frames were shown both in the order in which they were recorded or in the scrambled order. We tested the following frame rates: 3, 4, 5, 6, 7, 9, and 12.5 Hz and the original recordings at 25 Hz. We also compared the ratings obtained for these stimuli with those obtained for the original movie stimuli (25 Hz) in order to select stimuli for the subsequent fMRI experiment. Results are shown in Figure 2. A 2-way, repeated-measures ANOVA [factors: “frame order” (2 levels: ordered and scrambled) and “frame rate” (7 levels)] revealed that, as expected, frame rate and frame order affected the perceived fluidity and meaning of the videos. Meaning increased with frame rate and was higher for stimuli with correct frame order. Fluidity ratings were also higher at higher frame rates. The frame order effect was stronger at higher frame rates on both fluidity and meaning ratings. Detailed results are as follows: (1) fluidity ratings: main effect of frame rate: $F_{6,54} = 1.33$, $P > 0.2$; main effect of frame order: $F_{1,9} = 107.77$, $P \ll 0.001$; interaction: $F_{6,54} = 27.42$, $P < 0.001$ and (2) meaning ratings: main effect of frame rate: $F_{6,54} = 5.45$, $P < 0.001$; main effect of frame order: $F_{1,9} = 464.84$, $P < 0.001$; interaction: $F_{6,54} = 31.77$, $P < 0.001$.

Post hoc tests on the stimuli with ordered frames revealed that stimuli at all frame rates except at 12.5 Hz appeared less fluid than the 25 Hz stimuli, and all stimuli except those at 6, 7, 9, and 12.5 Hz appeared less meaningful than the original 25 Hz stimuli. Results are as follows: (1) fluidity ratings: 25 Hz ordered versus 12.5 Hz ordered: $t(9) = 1.22$, $P > 0.2$; 25 Hz ordered versus each of the other ordered stimuli: all $t > 3.7$, all $P < 0.003$; paired t -tests, Bonferroni corrected for $N = 7$ tests, threshold P -value = $0.05/7 = 0.007$; (2) meaning ratings: 25 Hz ordered versus 3, 4, or 5 Hz: $t > 3.48$, all $P \leq 0.007$; 25 Hz versus 6, 7, 9, and 12.5 Hz: all $t < 3.48$, all $P > 0.014$.

Given these results, we decided to use the following stimuli in the fMRI experiment: (i) the original movie stimuli at 25 Hz, (ii) stimuli with a reduced frame rate that was rated as no different in terms of fluidity and meaning from the original movie stimulus (12.5 Hz), and (iii) stimuli with a reduced frame rate that was rated as less meaningful and fluid than the original movie stimulus, but had nevertheless more frames than a static face stimulus (5 Hz). We used both stimuli with correct and scrambled frame order at each frame rate. These stimuli allowed us to test the roles of frame rate and frame order on the increased response evoked by dynamic compared with static faces.

Behavioral Data Collected During fMRI Experiment

During the fMRI experiment, participants performed a 1-back repetition detection task (stream of letters presented at the center of the screen) unrelated to the face stimuli. Average target detection performance across all participants and conditions was 68% (SEM = 3.45%), with an average response time (RT) of 554 ms (SEM = 34 ms). A 1-way repeated-measures ANOVA showed no differences in target detection performance or RT between the 9 stimulus conditions (detection: $F_{8,200} = 1.03$, $P > 0.4$; RT: $F_{8,200} = 0.59$, $P > 0.7$). For the conditions with stimuli made of more than 1 frame, a 2-way repeated-measures ANOVA did not show any effects of frame rate or frame order and no interaction between these factors (detection: all $F < 2.2$, all $P > 0.12$; RT: 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. This reduces the likelihood that differences in brain activation between conditions are due to differences in attentional demands.

Results of Whole-Brain fMRI Analyses

We tested all voxels in the brain on several 1-tailed contrasts in the group analysis. Figure 3A shows clusters of voxels responding more to 25 Hz stimuli with ordered frames than to static faces, located in STS and medial temporal gyrus (MTG, including hMT+/V5). Regions showing a BOLD response linearly increasing with frame rate (averaged over ordered and scrambled) were found bilaterally in MTG (including hMT+/V5), STS, and inferior frontal gyrus (IFG), which can be seen in Figure 3B. Figure 3C shows that movies with ordered frames evoked a stronger response than movies with scrambled frame order (averaged over frame rates) in STS and right anterior STS, bilaterally. We did not detect regions with a significant interaction between frame rate and frame order, and there were no regions that responded significantly more to frame-scrambled compared with frame-ordered stimuli (Table 1).

Results of the ROI Analyses

We tested 2 types of ROIs. The first type, those responding more to moving faces than to static phase-scrambled faces, localized using the first run of our main experiment, included STS, inferior occipital gyrus (IOG), and fusiform gyrus (FG). The second type, localized with a classic face localizer, were the classic face-sensitive regions FFA and OFA. The number of participants in which each ROI was found and data about the coordinates are reported in Table 2. As can be seen in Table 2, the coordinates for FG and FFA are very similar, as are those for IOG and OFA. However, these ROIs do not contain the same voxels: for example, of the 16 subjects in which both right FG and right FFA were found, 8 of them show no overlap at all between these ROIs, 5 have <10% of the union of the voxels of both ROIs that are common to both ROIs, and the remaining 3 have overlap between 21% and 32%. There was no obvious systematic in the relative locations of these ROIs. Similar results were found for the other ROIs: of the 9 participants with left FG and left FFA, only 2 had more than 10% overlap; for right IOG and right OFA, 3 of the 19 participants had 10% overlap or more; and none of the 14 participants with left IOG and left OFA had more than 10% overlap. We report results from all these ROIs to show the similarity and differences of our results across the different contrasts and data used to define these ROIs. The responses of these ROIs to the different conditions in the experiment are shown in Figure 4. To assess whether response amplitudes in the ROIs reliably varied as a function of frame rate and frame order, we performed 2-way repeated-measures ANOVA on the data of each ROI. Results as listed in Table 3 revealed that bilateral IOG and OFA as well as right FFA showed a reliable response change when the frame rate varied (a trend at $P < 0.1$ was found in left FFA). Further, response in right STS changed as a function of frame order, with a trend at $P < 0.1$ found in left STS. No regions showed a significant interaction between frame order and frame rate.

Next, we assessed whether our ROIs showed a greater response to stimuli with more than 1 frame compared with

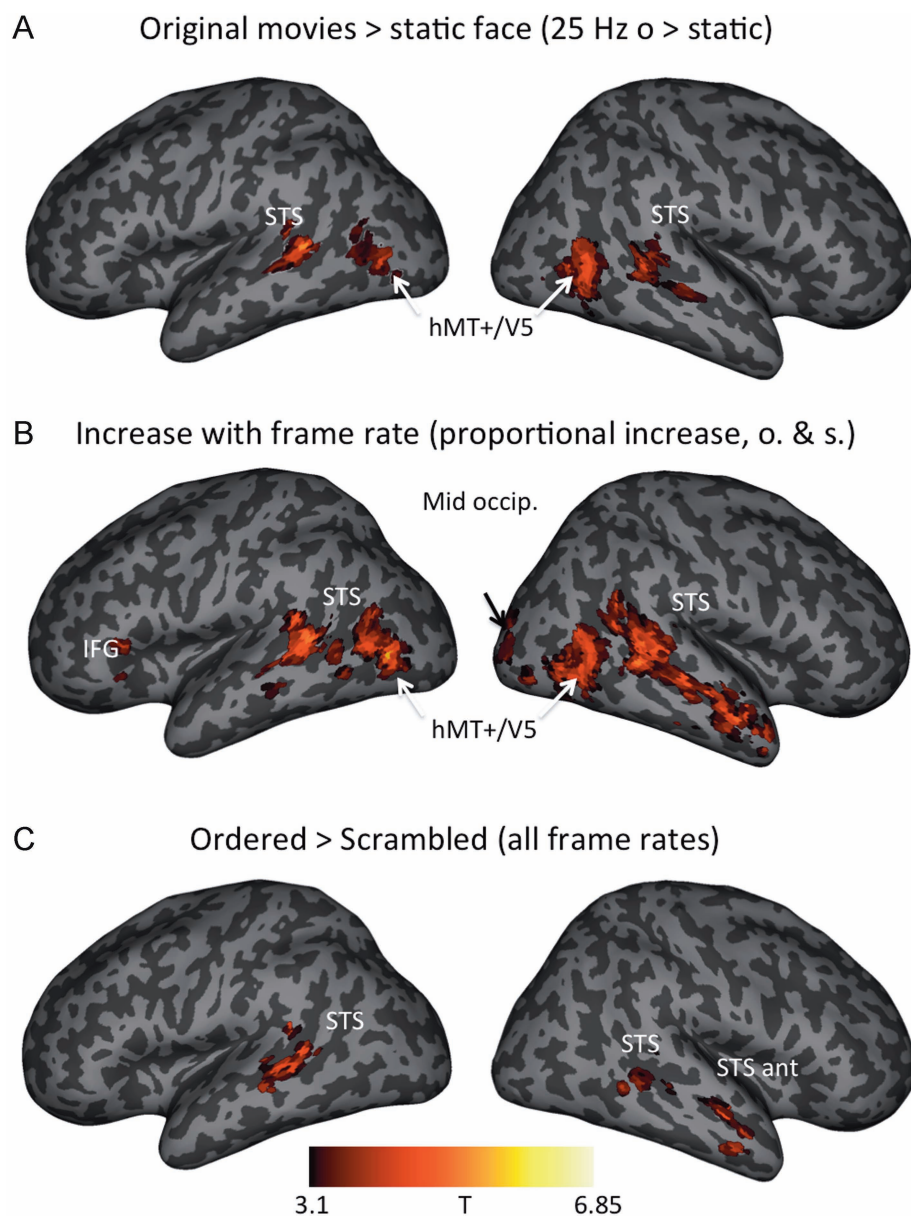


Figure 3. Results of the whole-brain fMRI group analysis: results of 1-tailed *t*-tests projected on the surface of an inflated standard structural scan. (A) Brain regions showing greater activation in response to the recorded movies of facial motion than in response to static faces; the contrast used was 25 Hz ordered frames > static face. (B) Brain regions showing activation proportional to the frame rate, irrespective of frame order. (C) Brain regions with greater activation in response to stimuli with ordered frames than to stimuli with scrambled frames, irrespective of frame rate. O, stimuli with ordered frames; S, stimuli with scrambled frame order; STS, activation near superior temporal sulcus; STS ant, anterior part of the STS; hMT+/V5, human motion complex; IFG, inferior frontal gyrus; Mid occip, middle occipital gyrus.

static face stimuli. This was tested using separate *t*-tests comparing 5, 12.5, and 25 Hz stimuli with ordered frames with the static face stimuli (corrections for multiple tests were made using the Holm–Bonferroni method, see Materials and Methods). Results (Table 3) revealed that only right OFA showed a significantly stronger response to ordered 5 Hz stimuli compared with static faces. However, bilateral OFA and FFA, right STS, and right IOG showed a significantly stronger response to ordered 12.5 Hz movies compared with static faces, with a trend at $P < 0.1$ found in the right FG. No regions showed a stronger response to ordered 25 Hz stimuli compared with static faces.

As can be seen in Figure 4, in almost all regions, the peak response was at 12.5 Hz and not for stimuli with the maximal

frame rate of 25 Hz. This is in agreement with previous findings, showing a limited temporal processing capacity of high-level object-selective areas (McKeeff et al. 2007). To determine the frame rate to which each ROI was most sensitive, we fitted the response of each subject in each ROI using a Gaussian function and identified the peak of the fitted function (these results are not dependent on the kind of the function used for the fitting: almost identical results were obtained when fitting a second-order polynomial function, suggesting that these results are reliable). This fitting was done twice for each ROI: once using the static frame and the 5, 12.5, and 25 Hz stimuli with correct frame order and the second time using the static frame and the 5, 12.5, and 25 Hz stimuli with scrambled frame order. The results are shown

Table 1

Details of clusters found in the whole-brain contrasts calculated on the main experimental data, ranked in decreasing order of *T*- and *Z*-values

Anatomy	H	Coordinates	Size	<i>T</i> -value	<i>Z</i> -value
Original 25 Hz movies > static faces					
STS	L	-57, -45, 6	64	5.26	5.08
STS	R	54, -33, 3	89	4.89	4.74
MTG (hMT+/V5)	R	54, -66, 0	102	4.61	4.49
MTG (hMT+/V5)	L	-45, -78, 6	59	4.53	4.42
Increase proportional to frame rate					
MTG (hMT+/V5)	R	51, -39, 6	672	5.72	5.5
STS	R	57, -33, 0		5.6	5.39
MTG (hMT+/V5)	L	-45, -78, 6	304	5.61	5.4
STS	L	-60, -45, 9		5.34	5.16
IFG	L	-48, 24, 0	31	4.49	4.38
Middle occipital gyrus	R	30, -96, 15	38	4	3.92
Movies with ordered frames > movies with scrambled frames					
STS	L	-54, -45, 6	71	4.6	4.48
MTS/STS anterior	R	51, -15, -15	64	4.48	4.36
STS	R	60, -42, 0	35	4.08	3.99

Note: Threshold is $P = 0.05$ whole-brain corrected (cluster level). H, hemisphere. Size indicates the number of voxels, and coordinates indicate position in *X*, *Y* and *Z* and are in MNI format. Missing values in the size column indicate an activation peak that is part of the cluster listed immediately above.

Table 2

Details of the ROIs found

Anatomy	Coordinates		<i>N</i>	Size	
	Mean	STD		Mean	STD
Regions responding more to 25 Hz movies > static phase-scrambled faces					
Left STS	-56, -48, 7	5, 7, 7	21	405	513
Right STS	54, -36, 3	6, 7, 5	20	486	459
Left FG	-39, -52, -24	4, 9, 5	12	432	513
Right FG	41, -53, -25	6, 8, 6	16	648	675
Left IOG	-39, -81, -10	5, 5, 5	20	675	729
Right IOG	44, -76, -11	3, 6, 5	20	756	675
Regions responding more to faces than to objects					
Left FFA	-37, -54, -19	4, 7, 6	13	918	675
Right FFA	41, -53, -22	3, 7, 6	25	756	567
Left OFA	-36, -79, -13	5, 7, 5	19	648	567
Right OFA	41, -79, -8	5, 7, 6	24	594	621

Note: Coordinates are in MNI standard, and mean and standard deviation (STD) values are for *X*, *Y*- and *Z*-axes, in that order. *N* indicates in how many participants each ROI was found. Size measures are in mm³.

in Figure 5 and highlight peak sensitivities from about 10 to 18 Hz. Interestingly, only the right IOG showed reduced peak sensitivities for frame-scrambled stimuli [right IOG: $t(14) = 3.77$, $P < 0.003$].

Discussion

Many previous studies have shown that several brain regions that had previously been thought to be important mostly for static face processing respond more to moving faces than to static faces (Fox et al. 2009; Schultz and Pilz 2009; Pitcher et al. 2011). However, up to now, it remained unclear whether the increased response to facial motion observed in classical face-sensitive regions was due to (i) an increased amount of static information in the stimulus, that is, the different frames of videos compared with single static frames, (ii) the deformation of face over time, that is, fluid facial motion per se, or (iii) increased attentional demands. In this fMRI study, we performed whole-brain and regions of interest analyses to test hypotheses (i) and (ii). Differences in attentional demands were excluded by having participants perform a

task unrelated to the stimuli of interest throughout the whole experiment.

We found that brain activations evoked by facial motion are not only due to the many frames that constitute a movie stimulus, but that the correct order of the frames is also important. Whole-brain analysis singled out STS as the region showing the greatest response to moving faces, with both the amount of static information and the fluidity of the motion influencing the response. No regions showed a greater response to stimuli with scrambled frame order compared with stimuli with ordered frames. The ROI analysis confirmed previous findings: OFA, FFA, and right STS face-sensitive regions respond more to moving faces than to static faces. The increased amount of activation for moving faces in face-sensitive areas seems to be mainly due to the greater amount of static information in the moving face stimuli, with only right STS showing sensitivity to facial motion fluidity.

These results are highly compatible with a neurophysiologically plausible neural model of biological motion processing (Giese and Poggio 2003): a form pathway appears to analyze stimuli as discrete event snapshots, whereas a motion pathway analyzes information based on optic-flow information. Both kinds of information are integrated in higher level areas such as STS and fusiform areas, resulting in different responses depending on the order of the presented images. Our results support this hypothesis and also suggest that form and motion information are weighed differently in STS and fusiform areas.

The Role of STS in Processing Facial Motion

Several of our analyses highlighted the role of STS in processing moving face stimuli. First, the whole-brain analysis revealed that BOLD response in STS (1) was higher during presentation of moving faces compared with static faces, (2) increased with the frame rate, (3) was higher when facial motion appeared fluid, and (4) was higher when the order of the frames constituting a moving face stimulus was correct (i. e. as recorded) rather than scrambled. Secondly, the ROI analysis confirmed that frame order influenced right STS activation reliably (a trend was found in left STS) and that moving face stimuli with 12.5 Hz evoked stronger responses than static face stimuli in right STS. The absence of significant effects in the left STS might be related to previous studies showing greater sensitivity to faces within the right hemisphere (Kanwisher et al. 1997; Pitcher et al. 2011) and a right hemispheric predominance in processing biological motion (Bonda et al. 1996; Grossman et al. 2000). Taken together, these results confirm previous studies, which indicated that the STS is the brain region most strongly associated with processing facial motion (Schultz and Pilz 2009; Pitcher et al. 2011). Our whole-brain results further suggest that STS is sensitive to both the amount of static information and the fluid facial motion, and our ROI analysis confirmed the sensitivity to motion fluidity. To our knowledge, only 2 brain imaging studies have previously reported decreased activation in STS in response to motion stimuli with scrambled frame order compared with ordered frames: one fMRI study in which subjects watched movie scenes with durations of several minutes (Hasson et al. 2008) and one magnetoencephalography study that used sequences of morph-based animations of facial expressions (Furl et al. 2010). Our results replicate their

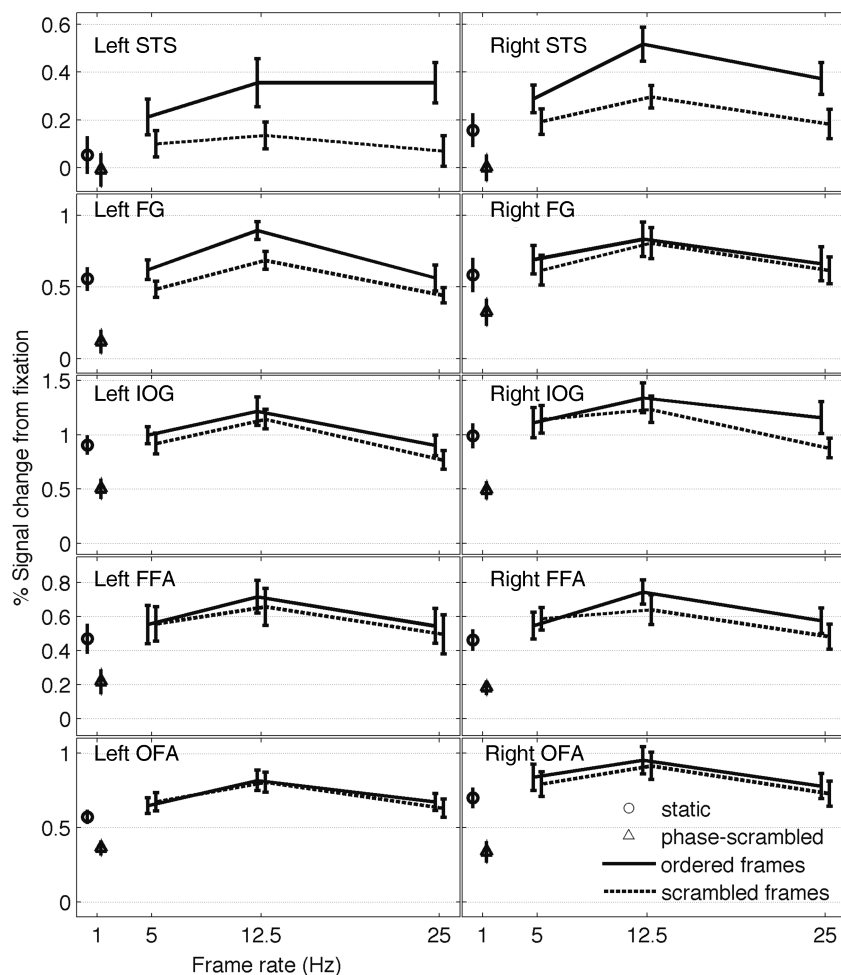


Figure 4. Percent signal change from fixation in individually defined ROIs. For abbreviations, see main text. Error bars represent standard errors of the mean across participants.

Table 3

Results of the tests performed on the ROI data

Anatomy	ANOVA			<i>t</i> -test movies > static face		
	Frame rate	Frame order	Interaction	5 Hz	12.5 Hz	25 Hz
Regions responding more to 25 Hz movies > static phase-scrambled faces						
Left STS	1.43	10.52*	1.39	1.95	3.05	3.15
Right STS	4.93	17.08**	0.88	1.67	3.74**	2.94
Left FG	8.2*	6.91	0.15	0.5	3.46	0.04
Right FG	6.42	0.74	0.05	0.95	3.66*	0.99
Left IOG	19.53***	5.47	0.16	1.02	2.64	-0.06
Right IOG	14.32***	6.03	2.47	1.52	4.35**	1.65
Regions responding more to faces than to objects						
Left FFA	7.33*	1.07	0.51	1.29	6.09****	1.68
Right FFA	13.61***	3.89	2.69	2	5.79***	2.78
Left OFA	10.96****	0.26	0.62	1.51	3.94**	2.4
Right OFA	18.1***	2.94	0.02	3.29*	4.28**	1.53

Note: ANOVA assessed the effects of frame rate and frame order (*F*-values are reported). *t*-tests compared the response to stimuli with multiple correctly ordered frames against static faces, separately for each frame rate (*t*-values are reported). ROIs responding to 25 Hz ordered > static scrambled faces were defined using this contrast calculated on the data of run 1 in the main experiment. Statistics for these ROIs are based on the data of run 2 only, which were not used to localize the ROIs. Face-sensitive ROIs were defined using the separate face localizer. Degrees of freedom for each *t*-test are the number of participants in which the ROI was found (column 4 in Table 2), minus one.

P* < 0.1, *P* < 0.05, ****P* < 0.001, and *****P* < 0.005. *P*-values are corrected for multiple tests using the Holm–Bonferroni method.

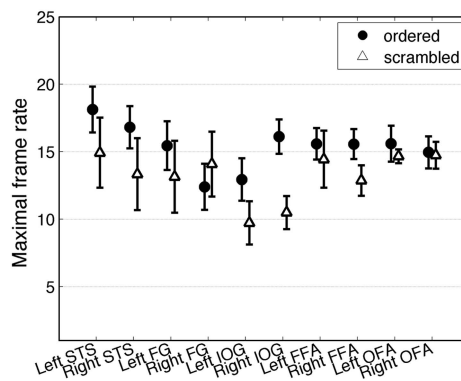


Figure 5. Stimulus frame rates evoking maximum activation in the different ROIs, assessed by the peak of a Gaussian function fitted to the data of each ROI and each participant, separately for the conditions with ordered frames and those with scrambled frame order. Error bars represent standard errors of the mean across participants.

findings that frame scrambling leads to decreased activation in STS using natural facial motion. In addition, our results allow a direct comparison between the effects of frame order and frame rate.

Our results are in complete accordance with the extensive literature associating STS with processing of signals relevant for social perception and communication (Allison et al. 2000), including facial motion (Puce et al. 1998; Campbell et al. 2001; LaBar et al. 2003; Bartels and Zeki 2004; Hasson et al. 2004; Hall et al. 2005; Pelphrey et al. 2007; Materna et al. 2008a, 2008b; Fox et al. 2009; Schultz and Pilz 2009; Lee et al. 2010; Said et al. 2010), point-light walkers (Bonda et al. 1996; Grezes et al. 1999; Grossman et al. 2000, 2004; Grossman and Blake 2001, 2002; Peelen et al. 2006), animations depicting social interactions between moving abstract shapes (Castelli et al. 2000; Schultz et al. 2004, 2005; Pavlova et al. 2010), and implied motion from static images (Puce et al. 1998, 2003; Castelli et al. 2000; Jellema and Perrett 2003; Puce and Perrett 2003; Schultz et al. 2004, 2005). The STS is also implicated in more cognitive aspects of social perception, including theory of mind (Fletcher et al. 1995; Frith and Frith 1999; Gallagher and Frith 2003; Samson et al. 2004; Saxe et al. 2004) or supramodal representation of emotional expressions (Peelen et al. 2010). Studies attempting to localize these functions with respect to each other within the STS region report overlapping activations in the right STS for many of these functions, and more specialized regions in left STS and bilateral temporo-parietal junction for the more complex processes (Bahnmann et al. 2010; Grosbras et al. 2011).

Ventral Temporal Face-Sensitive Regions and Facial Motion

Face movie stimuli with a frame rate of 12.5 Hz evoked a greater response compared with static faces in bilateral FFA and OFA as well as in right IOG, with a trend in right FG. As stated earlier, the aim of the present study was to find out whether response increases evoked by moving faces are due to (i) an increased amount of static information in the stimulus or (ii) the fluid facial motion per se. Our results show that most ventral temporal face-sensitive ROIs were sensitive to the frame rate of the stimuli but not to the frame order, suggesting that these regions are mainly sensitive to the higher amount of static information in moving face stimuli rather than to facial motion per se. These results are compatible with recent studies showing that not only superior but also ventral temporal face-sensitive regions respond strongly to moving face stimuli (Fox et al. 2009; Schultz and Pilz 2009; Pitcher et al. 2011). In agreement with our current results, 3 of these studies reported a stronger response to moving compared with static faces in the FG near the location of the FFA or in the FFA itself (Schultz and Pilz 2009; Trautmann et al. 2009; Lee et al. 2010), and 3 studies reported the same effect in the IOG or in the vicinity of the OFA (Fox et al. 2009; Schultz and Pilz 2009; Pitcher et al. 2011). In contrast, one study reported the absence of a difference in a direct comparison in both FFA and OFA (Pitcher et al. 2011). While it has been suggested that attentional mechanisms could be responsible for the increased activation for moving compared with static stimuli, especially in fusiform areas (Trautmann et al. 2009), it is unlikely that this factor has influenced our results, as the performance of our participants in the difficult detection task was similar across conditions.

We did not find any significant response increases when comparing stimuli with 25 Hz with static faces in any ROI. This discrepancy from our previous results (Schultz and Pilz 2009) might be related to the high number of tests performed in the current study and to the ensuing correction for multiple comparisons (we used the Holm–Bonferroni procedure, a relatively conservative method). Indeed, the response to the 25 Hz stimuli tended to be higher than that to static faces, and without correction for multiple comparisons, the difference reached significance in bilateral STS, right FFA, and left OFA.

Interestingly, although the results in FG and IOG show similarities to effects found in FFA and OFA, almost all effects tested in the former ROIs did not quite reach significance. The biggest difference was found between right FG and right FFA, suggesting that these ROIs might have different functional roles. This is striking given that their average coordinates are very similar (this similarity is also found between IOG and OFA, see Table 2). However, these ROIs do not overlap at all within participants (see Results). This can be explained by differences in the methods used to identify these ROIs: FG was defined using a less specific contrast compared with the one used to define FFA (see Materials and Methods and tables). While the functions of the right FFA have been widely investigated, those of the right FG as defined in the present study are much less well understood. Speculating about the function of FG on the basis of its location in ventral temporal cortex and of our current results, we propose that FG could be a region sensitive to a number of different stimuli without preference for faces (as there was little overlap with FFA), with a slight trend toward a sensitivity to the amount of static information and no sensitivity to deformations of the stimulus over time (as there was no effect of frame order or frame rate). Whether FG is particularly sensitive to another object category than faces cannot be ascertained from our data. In any case, its function contrasts with the sensitivity to frame rate we found in all the other ventral temporal regions we studied (OFA, FFA, IOG, and even a trend in left FG).

It is noteworthy at this point to remember that the effects reported in FG, IOG, and STS are based on only half the experimental data, to avoid statistical bias. In addition, right FG was identified in only 16 participants, whereas right FFA was found in 25 participants. With less participants and less data points, an increased effect of noise is inevitable and thus decreases in statistical values are to be expected. It is thus not completely unlikely that some of the functional differences between right FG and right FFA might disappear if more data were to be collected, but this would go beyond the scope of the present study.

Peak Response Sensitivities and the Percept of Facial Motion

We found that all the ROIs we tested had peak response sensitivities to frame rates around 10–18 Hz. This range of values is interesting as it can be related to our perceptual results: of all our stimuli with reduced frame rates, only the facial motion contained in the 12.5 Hz stimuli did not appear less fluid and meaningful than that at 25 Hz. Thus, a frame rate of 12.5 Hz seems high enough to induce a percept of natural fluid motion similar to the video recordings at 25 Hz. However, both a relatively high frame rate and a correct frame

order were necessary to induce the percept of fluid motion, as scrambling the order of the frames diminished the percept of fluid motion at all frame rates. Hence, it seems that at frame rates of 12.5 Hz and above, the frame-to-frame image changes in the stimuli with correct frame order were small enough and the presentation times short enough to evoke a percept of fluid motion rather than a sequence of successive images. Interestingly, 12.5 Hz was also the frame rate evoking the highest BOLD response. The percept of fluid motion might be the result of an integration of these small frame-to-frame image changes over time, and a candidate region for this process might be the STS. The fact that a frame rate of 12.5 Hz was sufficient to induce the percept of fluid motion agrees well with previous work on the wagon wheel illusion that showed a maximal percept of the illusion at alternation rates around 10 Hz (Purves et al. 1996; VanRullen and Koch 2003; VanRullen et al. 2005).

Our study suggests a link between the minimal frame rate leading to a percept of fluid facial motion and the frame rate evoking the peak BOLD response. One example of such a link between the temporal integration window in motion perception and a neural signal is provided by an electroencephalography (EEG) study, reporting that only 1 EEG spectral component, around 13 Hz, was affected by the continuous wagon wheel illusion (VanRullen 2006). Another study showed a decrease in activation for face stimuli with a presentation frequency higher than 10 Hz (McKeeff et al. 2007): McKeeff et al. presented diverse images of faces or houses at different frame rates and observed highest activation for stimuli presented at frame rates between 4 and 10 Hz (results varied slightly depending on the analysis method). These somewhat lower peak response rates found by McKeeff et al. (our peak responses were observed around 10–18 Hz) might be due to the fact that in their study, the images were unrelated to each other, that is, the identity and expression of the faces differed between successive frames of the presented stimulus sequence. Therefore, at a given frame rate, changes between successive images were greater than those in our study, in which all images depicted the same person showing the same expression within the same context, even in stimuli with scrambled frame order.

It has been previously suggested that there are temporal limitations on the amount of images the brain can process so that each image can be classified as a distinct event (Raymond et al. 1992; Duncan et al. 1994). Our results suggest that at lower frame rates, the brain processes each frame of a movie as a distinct event, yielding a percept of nonfluid motion. In contrast, when the low-level properties between successive frames are small enough and the frame rate is high enough, the successive images are successfully integrated into the percept of a single dynamic event. Such a percept can boost the encoding of information, for example, faces learned in motion are better recognized than static faces (Pilz et al. 2006). In a striking example, the percept of fluid motion from different (morphed) face images can lead these images to be perceived as one identity through temporal association of views (Wallis and Bülthoff 2001; Wallis et al. 2009).

Conclusion

In this study, we tested whether the response increases observed when comparing moving with static faces was due to

(1) an increased amount of static information in the stimulus or (2) the deformation of face over time, that is, fluid facial motion per se. We equated attentional demands as much as possible by having participants perform an unrelated task. We found that both factors are important, with differences between regions: STS response was mostly influenced by the fluidity of the motion (which depended mainly on frame order), whereas ventral face-sensitive regions were mostly influenced by static information (the amount of which was controlled by the frame rate).

In our experiments, we used stimuli with which we attempted to separate frame rate and motion fluidity as much as possible. While this approach was successful, it also led to some findings that might not be directly related to the habitual real-world experience of seeing facial motion, mainly the finding that stimuli with 12.5 Hz evoked the highest activation. Obviously, in daily life, we are normally not exposed to such kinds of low frame rate stimuli—except maybe under stroboscopic light in a night club or when blinking really quickly. However, our results nicely show that both dorsal and ventral areas such as STS and FFA/FG primarily process different kinds of information contained in facial motion and therefore contribute to understanding real-world brain function by understanding how facial motion is processed in the human brain.

Funding

This research was supported by the Max Planck Society and HHB was supported in part through the WCU (World Class University) program funded by the Ministry of Education, Science and Technology through the National Research Foundation of Korea (R31-10008). Funding to pay the Open Access publication charges for this article was provided by the Max Planck Society.

Notes

Conflict of Interest: None declared.

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 KJ. 1999. Nonlinear spatial normalization using basis functions. *Hum Brain Mapp.* 7:254–266.
- Ashburner J, Friston KJ. 1997. Spatial transformation of images. In: Frackowiak RSJ, Friston KJ, Frith CD, Dolan RJ, Mazziotta JC, editors. *Human brain function.* London: Academic Press. p. 43–59.
- Bahnemann M, Dziobek I, Prehn K, Wolf I, Heekeren HR. 2010. Sociotopy in the temporoparietal cortex: common versus distinct processes. *Soc Cogn Affect Neurosci.* 5:48–58.
- Bahrani B, Lavie N, Rees G. 2007. Attentional load modulates responses of human primary visual cortex to invisible stimuli. *Curr Biol.* 17:509–513.
- Bartels A, Zeki S. 2004. Functional brain mapping during free viewing of natural scenes. *Hum Brain Mapp.* 21:75–85.
- Bassili JN. 1976. Temporal and spatial contingencies in the perception of social events. *J Pers Soc Psychol.* 33:680–685.
- Blakemore S-J, Decety J. 2001. From the perception of action to the understanding of intention. *Nat Rev Neurosci.* 2:561–567.
- 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 D. 1997. The Psychophysics Toolbox. *Spat Vision*. 10:433–436.
- Campbell R. 1992. The neuropsychology of lipreading. *Phil Trans R Soc B*. 335:39–45.
- Campbell R, MacSweeney M, Surguladze S, Calvert GA, Mcguire PK, 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). *Cogn Brain Res*. 12:233–243.
- Castelli F, Happé F, Frith U, Frith C. 2000. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage*. 12:314–325.
- Corbetta M, Shulman G. 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci*. 3:201–215.
- Duncan J, Ward R, Shapiro K. 1994. Direct measurement of attentional dwell time in human vision. *Nature*. 369:313–315. Available from: <http://www.nature.com/doi/10.1038/369313a0>.
- Fletcher PC, Happe F, Frith U, Baker SC, Dolan RJ, Frackowiak RS, Frith CD. 1995. Other minds in the brain: a functional imaging study of “theory of mind” in story comprehension. *Cognition*. 44:283–296.
- 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.
- Franconeri SL, Simons DJ. 2003. Moving and looming stimuli capture attention. *Percept Psychophys*. 65:999–1010.
- Frisoni KJ, Ashburner J, Frith CD, Poline J-B, Heather J, Frackowiak RSJ. 1995a. Spatial registration and normalisation of images. *Hum Brain Mapp*. 2:165–189.
- Frisoni K, Holmes A, Price C, Buchel C, Worsley K. 1999a. Multisubject fMRI studies and conjunction analyses. *NeuroImage*. 10:385–396.
- Frisoni 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.
- Frisoni KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ. 1995b. Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapp*. 2:189–210.
- Frisoni KJ, Worsley KJ, Frackowiak R, Mazziotta J, Evans AC. 1994. Assessing the significance of focal activations using their spatial extent. *Hum Brain Mapp*. 1:210–220.
- Frisoni KJ, Zarahn E, Josephs O, Henson RN, Dale AM. 1999b. Stochastic designs in event-related fMRI. *NeuroImage*. 10:607–619.
- Frith CD, Frith U. 1999. Interacting minds—a biological basis. *Science*. 286:1692–1695.
- Furl N, van Rijsbergen NJ, Kiebel SJ, Frisoni KJ, Treves A, Dolan RJ. 2010. Modulation of perception and brain activity by predictable trajectories of facial expressions. *Cereb Cortex*. 20:694–703.
- Gallagher HL, Frith CD. 2003. Functional imaging of “theory of mind”. *Trends Cogn Sci*. 7:77–83.
- Gauthier I, Tarr MJ, Moylan J, Skudlarski P, Gore JC, Anderson AW. 2000. The fusiform “face area” is part of a network that processes faces at the individual level. *J Cogn Neurosci*. 12:495–504.
- Giese MA, Poggio T. 2003. Neural mechanisms for the recognition of biological movements. *Nat Rev Neurosci*. 4:179–192.
- Grezes J, Costes N, Decety J. 1999. The effects of learning and intention on the neural network involved in the perception of meaningless actions. *Brain*. 122:1875–1887.
- Grill-Spector K, Knouf N, Kanwisher N. 2004. The fusiform face area subserves face perception, not generic within-category identification. *Nat Neurosci*. 7:555–562.
- Grosbras M-H, Beaton S, Eickhoff SB. 2011. Brain regions involved in human movement perception: a quantitative voxel-based meta-analysis. *Hum Brain Mapp*. doi: 10.1002/hbm.21222.
- Grossman E, Blake R. 2002. Brain areas active during visual perception of biological motion. *Neuron*. 35:1167–1175.
- Grossman E, Blake R, Kim C. 2004. Learning to see biological motion: brain activity parallels behavior. *J Cogn Neurosci*. 16:1669–1679.
- Grossman E, Donnelly M, Price R, Pickens D. 2000. Brain areas involved in perception of biological motion. *J Cogn Neurosci*. 12:711–720.
- Grossman ED, Blake R. 2001. Brain activity evoked by inverted and imagined biological motion. *Vis Res*. 41:1475–1482.
- Halgren E, Dale A, Sereno M, Tootell R, Marinkovic K, Rosen BR. 1999. Location of human face-selective cortex with respect to retinotopic areas. *Hum Brain Mapp*. 7:29–37.
- Hall D, Füssell C, Summerfield A. 2005. Reading fluent speech from talking faces: typical brain networks and individual differences. *J Cogn Neurosci*. 17:939–953.
- Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R. 2004. Intersubject synchronization of cortical activity during natural vision. *Science*. 303:1634–1640.
- Hasson U, Yang E, Vallines I, Heeger D, Rubin N. 2008. A hierarchy of temporal receptive windows in human cortex. *J Neurosci*. 28:2539–2550.
- Haxby JV, Hoffman EA, Gobbini MI. 2000. The distributed human neural system for face perception. *Trends Cogn Sci*. 4:223–233.
- Hill H, Johnston A. 2001. Categorizing sex and identity from the biological motion of faces. *Curr Biol*. 11:880–885.
- Ishai A. 2008. Let’s face it: it’s a cortical network. *NeuroImage*. 40:415–419.
- Jellema T, Perrett D. 2003. Cells in monkey STS responsive to articulated body motions and consequent static posture: a case of implied motion? *Neuropsychologia*. 41:1728–1737.
- Josephs O, Henson R. 1999. Event-related functional magnetic resonance imaging: modelling, inference and optimization. *Phil Trans R Soc Lond B*. 354:1215–1228.
- Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*. 17:4302–4311.
- Kanwisher N, Yovel G. 2006. The fusiform face area: a cortical region specialized for the perception of faces. *Phil Trans R Soc B Biol Sci*. 361:2109–2128.
- Kleiner M, Brainard D, Pelli D. 2007. What’s new in Psychtoolbox-3? *Perception*. 36:2740. Available from: <http://www.perceptionweb.com/abstract.cgi?id=v070821>.
- Knappmeyer B, Thornton I, Bühlhoff H. 2003. The use of facial motion and facial form during the processing of identity. *Vis Res*. 43:1921–1936.
- Kruggel F, von Cramon DY. 1999. Modeling the hemodynamic response in single-trial functional MRI experiments. *Magn Reson Med*. 42:787–797.
- LaBar KS, Crupain MJ, Voyvodic JT, McCarthy G. 2003. Dynamic perception of facial affect 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, Chuang L. 2005. Why are moving faces easier to recognize? *Vis Cogn*. 12:429–442.
- Lander K, Chuang L, Wickham L. 2006. Recognizing face identity from natural and morphed smiles. *Q J Exp Psychol*. 59:801–808.
- Lee LC, Andrews TJ, Johnson SJ, Woods W, Gouws A, Green GGR, Young AW. 2010. Neural responses to rigidly moving faces displaying shifts in social attention investigated with fMRI and MEG. *Neuropsychologia*. 48:477–490.
- Liu TT. 2004. Efficiency, power, and entropy in event-related fMRI with multiple trial types. Part II: design of experiments. *NeuroImage*. 21:401–413.
- Materna S, Dicke P, Thier P. 2008a. Dissociable roles of the superior temporal sulcus and the intraparietal sulcus in joint attention: a functional magnetic resonance imaging study. *J Cogn Neurosci*. 20:108–119.
- Materna S, Dicke P, Thier P. 2008b. The posterior superior temporal sulcus is involved in social communication not specific for the eyes. *Neuropsychologia*. 46:2759–2765.
- McCarthy G, Puce A, Gore JC, Allison T. 1997. Face-specific processing in the human fusiform gyrus. *J Cogn Neurosci*. 9:605–610.

- McKeeff TJ, Remus DA, Tong F. 2007. Temporal limitations in object processing across the human ventral visual pathway. *J Neurophysiol.* 98:382–393.
- O'Toole A, Roark D, Abdi H. 2002. Recognizing moving faces: a psychological and neural synthesis. *Trends Cogn Sci.* 6:261–266.
- Pavlova M, Guerreschi M, Lutzenberger W, Krageloh-Mann I. 2010. Social interaction revealed by motion: dynamics of neuromagnetic gamma activity. *Cereb Cortex.* 20:2361–2367.
- Peelen M, Wiggett A, Downing P. 2006. Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. *Neuron.* 49:815–822.
- Peelen MV, Atkinson AP, Vuilleumier P. 2010. Supramodal representations of perceived emotions in the human brain. *J Neurosci.* 30:10127–10134.
- Pelli D. 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vision.* 10:437–442.
- Pelphrey K, Morris J, McCarthy G, LaBar K. 2007. Perception of dynamic changes in facial affect and identity in autism. *Soc Cogn Affect Neurosci.* 2:140–149.
- Perrett DI, Oram MW, Ashbridge E. 1998. Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. *Cognition.* 67:111–145.
- Pilz KS, Thornton IM, Bühlhoff HH. 2006. A search advantage for faces learned in motion. *Exp Brain Res.* 171:436–447.
- Pitcher D, Dilks DD, Saxe RR, Triantafyllou C, Kanwisher N. 2011. Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage.* 56:2356–2363.
- 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, Perrett D. 2003. Electrophysiology and brain imaging of biological motion. *Phil Trans R Soc Lond B Biol Sci.* 358:435–445.
- Puce A, Syngieniotis 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.
- Purves D, Paydarfar JA, Andrews TJ. 1996. The wagon wheel illusion in movies and reality. *Proc Natl Acad Sci USA.* 93:3693–3697.
- Raymond JE, Shapiro KL, Arnell KM. 1992. Temporary suppression of visual processing in an RSVP task: an attentional blink? *J Exp Psychol Hum.* 18:849–860.
- Rotshtein P, Henson R, Treves A, Driver J. 2005. Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nat Neurosci.* 8:107–113.
- Said CP, Moore CD, Engell AD, Todorov A, Haxby JV. 2010. Distributed representations of dynamic facial expressions in the superior temporal sulcus. *J Vis.* 10:1–12.
- Samson D, Apperly IA, Chiavarino C, Humphreys GW. 2004. Left temporoparietal junction is necessary for representing someone else's belief. *Nat Neurosci.* 7:499–500.
- Saxe R, Xiao D, Kovacs G, Perrett D, Kanwisher N. 2004. A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia.* 42:1435–1446.
- Schultz J, Friston KJ, O'Doherty J, Wolpert DM, Frith CD. 2005. Activation in posterior superior temporal sulcus parallels parameter inducing the percept of animacy. *Neuron.* 45:625–635.
- Schultz J, Imamizu H, Kawato M, Frith CD. 2004. Activation of the human superior temporal gyrus during observation of goal attribution by intentional objects. *J Cogn Neurosci.* 16:1695–1705.
- Schultz J, Pilz KS. 2009. Natural facial motion enhances cortical responses to faces. *Exp Brain Res.* 194:465–475.
- Sergent J, Ohta S, MacDonald B. 1992. Functional neuroanatomy of face and object processing: a positron emission tomography study. *Brain.* 115:15–36.
- Thornton IM, Kourtzi Z. 2002. A matching advantage for dynamic human faces. *Perception.* 31:113–132.
- Trautmann SA, Fehr T, Herrmann M. 2009. Emotions in motion: dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain Res.* 1284:100–115.
- VanRullen R. 2006. The continuous wagon wheel illusion is associated with changes in electroencephalogram power at ~13 Hz. *J Neurosci.* 26:502–507.
- VanRullen R, Koch C. 2003. Is perception discrete or continuous? *Trends Cogn Sci.* 7:207–213.
- VanRullen R, Reddy L, Koch C. 2005. Attention-driven discrete sampling of motion perception. *Proc Natl Acad Sci USA.* 102:5291–5296.
- Wager TD, Nichols TE. 2003. Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *NeuroImage.* 18:293–309.
- Wallis G, Backus B, Langer M, Huebner G, Bühlhoff H. 2009. Learning illumination- and orientation-invariant representations of objects through temporal association. *J Vis.* 7:1–8.
- Wallis G, Bühlhoff HH. 2001. Effects of temporal association on recognition memory. *Proc Natl Acad Sci USA.* 98:4800–4804.
- Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. 1996. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum Brain Mapp.* 4:58–73.