NeuroImage 102 (2014) 407–415

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Face processing regions are sensitive to distinct aspects of temporal sequence in facial dynamics

Maren Reinl, Andreas Bartels *

Vision and Cognition Lab, Centre for Integrative Neuroscience, University of Tübingen, and Max Planck Institute for Biological Cybernetics, Tübingen 72076, Germany

ARTICLE INFO

Article history:
Accepted 4 August 2014
Available online 12 August 2014

Keywords:
Faces
Motion
FFA
STS
Emotion
Predictive coding

ABSTRACT

Facial movement conveys important information for social interactions, yet its neural processing is poorly understood. Computational models propose that shape- and temporal sequence sensitive mechanisms interact in processing dynamic faces. While face processing regions are known to respond to facial movement, their sensitivity to particular temporal sequences has barely been studied. Here we used fMRI to examine the sensitivity of human face-processing regions to two aspects of directionality in facial movement trajectories. We presented genuine movie recordings of increasing and decreasing fear expressions, each of which were played in natural or reversed frame order. This two-by-two factorial design matched low-level visual properties, static content and motion energy within each factor, emotion-direction (increasing or decreasing emotion) and timeline (natural versus artificial). The results showed sensitivity for emotion-direction in FFA, which was timeline-dependent as it only occurred within the natural frame order, and sensitivity to timeline in the STS, which was emotion-direction-dependent as it only occurred for decreased fear. The occipital face area (OFA) was sensitive to the factor timeline. These findings reveal interacting temporal sequence sensitive mechanisms that are responsive to both ecological meaning and to prototypical unfolding of facial dynamics. These mechanisms are temporally directional, provide socially relevant information regarding emotional state or naturalness of behavior, and agree with predictions from modeling and predictive coding theory.

Introduction

Social interactions in real life are dynamic by nature, and numerous social signals, including those conveyed by faces, rely on timing and temporal sequences. In accord with this, evidence shows that temporal contingencies in facial dynamics play an important role in behavior. Facial dynamics can improve the recognition of subtle emotional expressions (Ambadar et al., 2005; Wehrle et al., 2000), the personal identity of others (O'Toole et al., 2002; Thornton and Kourtzi, 2002), gender (Hill and Johnston, 2001) and language (Campbell, 1992).

Despite their importance, dynamic face stimuli have not been studied nearly as extensively as their static counterparts. Brain regions responsive to static faces increase their activity in response to facial dynamics (Fox et al., 2009; Kilts et al., 2003; LaBar et al., 2003; Sato et al., 2004; Schultz and Pilz, 2009; Trautmann et al., 2009). Particularly the posterior superior temporal sulcus (pSTS) is more sensitive to facial dynamics than the fusiform face area (FFA) or the occipital face area (OFA) (Pitcher et al., 2011; Schultz et al., 2012). This partial regional differentiation parallels one proposed for the encoding of changeable versus non-changeable aspects of faces (Haxby et al., 2000; Ishai et al., 2005; Puce and Perrett, 2003; Said et al., 2011; Vuilleumier and Pourtois, 2007). Changeable aspects of a face have been shown to be encoded by the pSTS, such as emotional expression (Said et al., 2010), gaze-direction (Hoffman and Haxby, 2000; Puce et al., 1998), mouth movements (Campbell et al., 2001) and intention (Nummenmaa and Calder, 2009) while non-changeable aspects are thought to be mainly processed by the FFA, such as identity (Furl et al., 2011; Hoffman and Haxby, 2000; Nestor et al., 2011; Steeves et al., 2009) (but see Kriegeskorte et al., 2007), race (Natu et al., 2011), and gender (Kaul et al., 2011).

It is nevertheless unclear which aspects of facial dynamics drive the increasing responses in the core face processing regions. Schultz et al. (2012) showed that the amount of static information as well as the fluidity of the facial motion influences the activity of the core regions, and Furl et al. (2010) report sensitivity of pSTS and posterior fusiform gyrus to intact versus scrambled facial movement using MEG. Computational modeling and theory suggest that directionality is a key aspect of visual biological motion processing, requiring dedicated neural detectors (Giese and Poggio, 2003). For example, the direction of change from a neutral to a happy face conveys a distinct meaning from the reverse direction, differentially affecting amygdala responses (Sato et al., 2010).

* Corresponding author at: Vision and Cognition Lab, Centre for Integrative Neuroscience, University of Tübingen, Otfried-Müller-Strasse 25, 72076 Tübingen, Germany.
E-mail address: andreas.bartels@tuebingen.mpg.de (A. Bartels).

http://dx.doi.org/10.1016/j.neuroimage.2014.08.011
1053-8119/

Independent from the emotional direction, the sequence of facial movements during relaxation of an emotional expression is not necessarily the exact reverse of the increase of that expression. Therefore, neural detectors that have been exclusively exposed to natural facial dynamics throughout the lifetime of an observer may respond differentially when exposed to artificially reversed timelines that contain non-canonical temporal sequences.

It is unknown which of the core face processing regions are sensitive to these two independent aspects of directionality, emotional directionality and timeline directionality. From an ecological and physiological point of view, one may expect independent cortical detectors for the two: we are equally frequently exposed to increasing and decreasing emotional expressions, yet they differ in ecological meaning and valence. Facial static emotional content has mostly been shown to increase activity in FFA, pSTS and OFA, along with many other regions of the extended face processing network such as the amygdala, temporal, and prefrontal cortex (Pessoa et al., 2002; Winston et al., 2003) (see for review (Atkinson and Adolphs, 2011)). In contrast, temporal deviations from normal movement trajectories would be expected to affect responses of sequence-specific circuitries thought to be present in the pSTS — either by reducing responses due to suboptimal stimulation, or by enhancing responses as a result of violating predictions (Giese and Poggio, 2003; Rao and Ballard, 1999). We used genuine movie recordings to study these two aspects of facial dynamics in a 2-by-2 factorial design that balanced all visual aspects apart from directionality of motion trajectories. A distracting gender-discrimination task and a rapid event-related design with an unpredictable sequence of stimuli were used to emphasize results related to bottom-up, automatic stimulus processing. Our fMRI results show that dorsal and ventral core face processing regions are sensitive to timeline and emotional directionality.

Methods

Participants

31 healthy participants with normal or corrected-to-normal vision participated in this study. Data of 27 participants (15 male, mean age 27 ± 4 years, 1 left-handed) entered the final analyses, as a total of 4 participants had to be excluded due to spiking artifacts (2) or excessive head-movement (2). The study was conducted according to the declaration of Helsinki and was approved by the local ethics committee of the University of Tübingen. Participants provided written consent prior to participation.

Stimuli

Main experiment

The stimuli of the main experiment included static pictures and short movie clips of faces of eleven actors showing fearful expressions. Movies were recorded prior to the experiment or were obtained from the Video-Face-Database of the MPI Tübingen (Kaulard et al., 2012). All movies were captured in color with the actor placed in front of a black background. Actors showed fearful expressions starting from neutral face, going to peak expression and relaxing back to a neutral expression, and were asked to keep their head still to avoid rigid head movements.

These genuine video recordings were later cut (while maintaining the original frame order) to show either an increase or a decrease of emotional intensity ranging from low to high fear expression or vice versa using VirtualDub (virtualdub.org). The mean durations of the cut movie recordings showing increasing or decreasing fear did statistically not differ (588 ± 139 ms and 680 ± 235 ms respectively). The means of the luminance and of its spatial variance, i.e. root-mean-square (RMS) contrast, for all movies were 96.04 cd/m^2 and 109.03 cd/m^2, respectively. Duplicates of these movies of increasing and decreasing fear expressions were then reversed in frame order, giving rise to two additional conditions: decreasing and increasing fear in reversed frame order. In total, we obtained four dynamic conditions: increasing and decreasing fear in original frame order, and decreasing and increasing fear in reversed frame order, with 11 exemplars for each. Two static conditions were created using start and end frames of each movie (low and high fear expression, again with 11 exemplars of each). Circular grid-scrambles served as static baseline conditions (Gschwind et al., 2012; Sato et al., 2004). They were obtained by cutting images into tiles of a 10 × 10 grid, and pseudorandomly relocating each tile to a new position that was equidistant to the image center (hence 'circular').

Localizer experiment

For the localizer experiment, neutral and fearful frames of faces from the above videos were contrasted to pictures of houses (kindly provided by Bruno Rossion, http://www.nefy.ucl.ac.be/Face_Categorisation_Lab.htm) as well as to circular grid-scrambles of all pictures. Luminance and RMS contrast of house pictures were adjusted to match those of the faces.

Stimuli were back-projected on a screen of 24 × 18 visual degrees, viewed via a tilted mirror and placed centrally, such that stimuli subtended 6 × 9°. All stimuli were presented using Cogent Graphics 1.30 developed by John Romaya at the Wellcome Department of Imaging Neuroscience (http://vislab.ucl.ac.uk/cogent.php) running on MATLAB 2010a on a Windows PC.

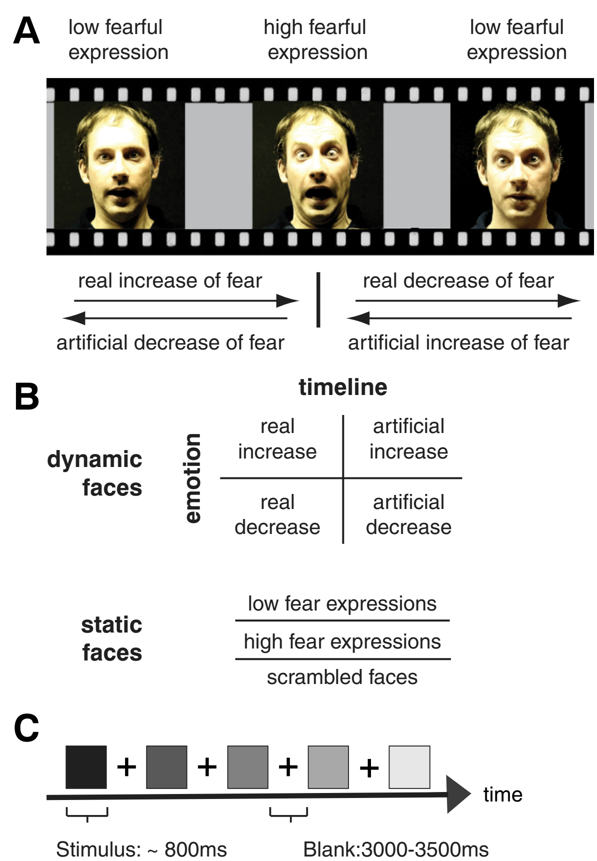

Fig. 1. Stimuli and experimental paradigm. (A) Illustration of how the four dynamic conditions of increasing and decreasing fearful expressions were obtained in both natural and artificial (reversed) timelines. (B) Overview of the 2-by-2 factorial design of the dynamic conditions with the factors emotion-direction (increasing or decreasing) and timeline (natural or artificial), and of the additional static conditions (movie start- and end-frames, scramble baseline). (C) Timing of the stimulus sequence of the event-related design.