Vision Research 143 (2018) 42–51

Contents lists available at ScienceDirect

Vision Research

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

The role of temporal inversion in the perception of realistic and morphed dynamic transitions between facial expressions

Olga A. Korolkova

Center for Experimental Psychology, Moscow State University of Psychology and Education, 2a Shelepikhinskaya Quay, 123290 Moscow, Russia

ARTICLE INFO

Keywords:
Face perception
Emotion
Dynamic facial expression
Automated face analysis
Timeline inversion
Video recording
Linear morphing

ABSTRACT

Recent studies suggest that video recordings of human facial expressions are perceived differently than linear morphing between the first and last frames of these records. Also, observers can differentiate dynamic expressions presented in normal versus time-reversed frame orders. To date, the simultaneous influence of dynamics (natural or linear) and timeline (normal or reversed) has not yet been tested on a wide range of dynamic emotional expressions and the transitions between them. We compared the perception of dynamic transitions between basic emotions in realistic (human-posed) and artificial (linearly morphed) stimuli which were presented in reversed or non-reversed order. The nonlinearity of realistic stimuli was demonstrated by automated facial structure analysis. The results of the behavioral study revealed that the recognition of emotions in time-reversed stimuli significantly differed from recognition of the normally presented ones, and this difference was substantially higher for videos of a dynamic human face than for linear morphs. Emotions displayed at the end of the transitions were recognized better than the first-frame emotions in all types of stimuli except in the time-reversed videos, which showed a similar recognition rate for both the starting and ending emotions. Our findings suggest that nonlinearity, which is present in a realistic facial display but absent in linear morphing, is an important cue for emotion perception, and that unnatural perceptual conditions (inversion in time) make the recognition of emotions more difficult. These results confirm the ability of the human visual system to use subtle dynamic cues on an interlocutor's face, and reveal its sensitivity to the timeline organization of the displayed emotions.

1. Introduction

In everyday communications we extensively use various non-verbal signals, and facial displays are probably the most important of them. The human face conveys both stable characteristics of its owner, such as gender, race and physical appearance, and transient ones: mood, intention and emotional state. Facial expressions constantly change, allowing us to infer socially and biologically relevant dynamic cues and to predict further actions of our counterpart. Studies of facial dynamics are therefore crucial for understanding the perceptual mechanisms underlying human interactions.

Recent behavioral studies have shown that dynamics can enhance the recognition of a face's emotional expressions (for reviews, see: Alves, 2013; Krumhuber, Kappas, & Manstead, 2013). The dynamic advantage is greater when the perception of a static expression is somehow disrupted: for example, in schematic (Bassili, 1978; Bruce & Valentine, 1988; Wehrle, Kaiser, Schmidt, & Scherer, 2000), blurred (Barabanschikov, Korolkova, & Lobodinskaya, 2015), and low-detailed (Cunningham & Wallraven, 2009) faces; with reduced configural, color and texture information (Kätsyri, Saalasti, Tiippana, von Wendt, & Sams, 2008; Kätsyri & Sams, 2008); or in faces with low intensity of the displayed emotions (Ambadar, Schooler, & Cohn, 2005; Bould & Morris, 2008). When a facial expression is already recognizable by its static image, the dynamic advantage is substantially reduced. For example, expressions displayed by human actors are easily identified both in static and dynamic modes (Fiorentini & Viviani, 2011), which is not the case with computer-generated 3D avatars without detailed facial textures. Evidence thus suggest that motion plays a key role in emotional face perception, facilitating it when perceptual conditions are suboptimal.

The most popular way to create dynamic stimuli for use in research is linear morphing, or interpolation, between static emotional and neutral faces. This procedure makes it possible to obtain highly controlled intermediate images with a predetermined percentage of each of the original faces which are blended together to produce a smooth dynamic change between them. An earlier study did not reveal any differences in the emotion recognition between dynamic computer-generated faces with all features moving synchronously, as in simple

E-mail address: olga.kurakova@psyexp.ru.

https://doi.org/10.1016/j.visres.2017.10.007
Received 7 February 2017; Received in revised form 4 October 2017; Accepted 4 October 2017

linear morphing, and the same faces with asynchronous movements (Wehrle et al., 2000). However, more recent studies suggest that linear morphs may lack some perceptual cues that help us to extract the meaning of a facial expression. The dynamics of a real face are not necessarily linear, since different facial muscles can start and stop moving at different times, with different velocity and intensity (Krumhuber & Scherer, 2011; Scherer & Ellgring, 2007), and this asynchrony may influence the perceived plausibility and credibility of the conveyed emotion. For example, the order of facial actions is important for judging the sequentially unfolding fear by the eye region (Krumhuber & Scherer, 2016). Observers also seem to be particularly sensitive to the naturalness of facial dynamics which cannot be efficiently modeled with simple linear morphing procedures, but requires advanced animation techniques (Dobs et al., 2014). Spatial non-linearity, after controlling for velocity differences, influences the perceived naturalness as well (Cosker, Krumhuber, & Hilton, 2010, 2015). Fine-grain discrimination between similar emotional displays, such as surprise and fear, relies on a small number of dynamic patterns, each including several action units that may be asynchronous in different face areas (Delis et al., 2016). Although these studies model the dynamics of real actors' faces, observers usually view and rate only artificial dynamic avatars (3D models), and this could also influence the accuracy of an emotion's perception and its naturalness. Realistic expressions on human faces, videotaped and parametrically described, can be more relevant for use as visual stimuli (Fiorentini, Schmidt, & Viviani, 2012).

The role of nonlinear dynamics of facial expressions in the perception of emotions has been supported by another important line of research, in which the natural timeline of expression is scrambled or reversed. A study using realistic movies of actors' emotional faces with increasing or decreasing intensity of fearful expressions revealed that timeline inversion of the stimuli changed the impression of the observers and led to lower perceived intensity, realism and convincingness of the presented emotion (Reinl & Bartels, 2015). Moreover, the time-reversed dynamic expression of an increase of fear, compared to the unaltered movie of decreased fear, resulted in different patterns of brain activation in left and right posterior superior temporal sulci, a bilateral region involved in the processing of biological motion, socially relevant stimuli, facial expressions and gaze direction (Reinl & Bartels, 2014). Activation in this area is higher for realistic unfolding of expressions than for time-scrambled frame sequences, and it correlates with the frame rate at which the dynamic expressions are presented (Schultz, Brockhaus, Bulthoff, & Pilz, 2013). Differences have also been found in the activation of the left fusiform face area for realistic increasing fear compared to realistic decreasing fear expressions (Reinl & Bartels, 2014), and in the left amygdala for linearly morphed increasing versus decreasing happiness and fear (Sato, Kochiyama, & Yoshikawa, 2010). The same neutral face at the end of a dynamic expression is perceived differently depending on whether it started from happiness or sadness (Jellema, Pecchinenda, Palumbo, & Tan, 2011; Marian & Shimamura, 2013; Yoshikawa & Sato, 2008). These data are in accordance with other behavioral and neuroimaging studies showing the importance of motion fluidity on emotion perception (e.g., Ambadar et al., 2005; Schultz & Pilz, 2009). Therefore, findings to date support the idea that an observer is highly sensitive to alterations of the expression timeline. The direction in which an expression changes indeed seems to convey important information for emotion recognition.

A growing body of research thus suggest that the direction of facial expression change (increase or decrease), its fluidity (Schultz & Pilz, 2009), velocity (Kamachi et al., 2001) and onset time (natural or altered), as well as timeline (normal or reversed) and dynamics (realistic nonlinear or artificial linear) can influence the recognition of emotion. Until now, however, the perception of timeline inversion has been studied using only the dynamic unfolding of emotions from a neutral face or their fading from intense emotion to neutrality, which represent just a small fraction of all possible emotional displays. To our knowledge, no study has explored the combined influence of altering the realistic frame order and non-linearity of facial dynamics on the perception of a wide range of dynamic transitions from one basic facial emotional expression to another. Such a study can extend previous results and provide important new information about face perception in realistic situations. Transitions between expressions are even more ecologically valid than the unfolding of emotions, since in the course of natural communication the expression on an interlocutor's face may change not only from a neutral state to an emotion and back, but also between different emotions, and these situations are psychologically different. For example, a shift from happiness to anger would probably occur in situations and lead to observer's perceptions and reactions which are opposite to those caused by a shift from anger to happiness.

The results of earlier research allow us to expect that the transitions between emotions presented in normal and time-reversed manners would be perceived differently. Moreover, there can exist differences depending on whether the emotion to be recognized is shown at the start or at the end of the dynamic stimuli. For example, in a study by Reinl and Bartels, presentations starting from a neutral face and ending with intense fear were perceived as significantly more fearful than stimuli starting from intense fear and ending with a neutral expression (Reinl & Bartels, 2015), which implies that fear displayed at its maximum at the end of the sequence has greater impact on emotion perception than maximal fear at its start. These situations have different implications: relaxing from fear means that actual conditions are not threatening anymore, which is not the case with increasing fear. However, participants in the mentioned study did not estimate the amount of neutrality in the stimuli, so in fact they rated only the intensity of the ending emotion when fear increased and the intensity of the starting emotion when fear decreased. Also, when a face was sequentially presented with two different expressions (anger or sadness), the observers' inferences about the actor's intentions and emotional states were more influenced by the last expression (Hareli, David, & Hess, 2016). Based on these results, one can expect that in dynamic transitions between two intense emotional states, observers would pay more attention to the last-frame emotion and recognize it better than the first-frame emotion, because the former is more relevant to the actor's actual emotional state.

Emotion recognition during transitions between basic expressions can also depend on the amount of non-linearity in the forward- compared to backward-presented facial dynamics. For example, in linear morphing, normal and time-reversed stimuli are equally smooth, and therefore perceptually equivalent (i.e., one would be unable to tell whether a morphing sequence is presented normally or reversed in time). In this case, emotions at the end of the morphed stimuli would be recognized equally well, and better than the first-frame emotions. In realistic movements, however, natural nonlinearities are ecologically valid and plausible only when presented with a normal frame order. Their perception would follow the same pattern as in morphed transitions: the emotion in the last frame would be recognized better than the starting emotion. Timeline reversal probably makes realistic facial dynamics less natural (Dobs et al., 2014), and therefore the emotions displayed at the end of transitions would be recognized with more difficulty. Starting emotions in this case may be perceived even more easily than with non-reversed stimuli, as they are observed prior to the unnatural movements and are not yet affected by them.

In the current study, we explore the simultaneous influences of the following factors on the recognition of emotions in dynamic transitions: realistic (human-posed) or artificial (linearly morphed) types of stimuli; and normal or reversed frame order. For the purposes of the study, we use high-speed video recordings of an actor who displayed emotional transitions on his face, and linear morphs which are based on these records. The dynamic stimuli were quantified to obtain measures of (non)linearity and then presented in a behavioral study to test the following hypotheses: 1) Recognition of the ending expression in morphed and realistic dynamic transitions is better than recognition of the starting expression; 2) Emotion recognition does not differ in