Analysis of Dynamic Characteristics of Spontaneous Facial Expressions

Masashi Komori (komori@oecu.jp)   Yoshitaro Onishi (mi13a002@oecu.jp)

Division of Information and Computer Sciences, Osaka Electro-Communication University,

18-8 Hatsucho, Neyagawa, Osaka, 572-8530, JAPAN

Abstract

The relationship between emotions elicited by film clips and spontaneous dynamic facial expressions was investigated. Participants (n = 10) watched 13 emotional film clips, and their facial responses were recorded using a motion capture system. We extracted 3-sec length intervals in which facial events occurred from motion sequences. The participants were asked to self-assess their felt emotional arousal and positive and negative affect for each interval. To find the spatiotemporal components of dynamic facial expressions, we employed the multiway decomposition method, PARAFAC, on a time sequence of facial landmark coordinates standardized via methodologies of geometric morphometrics. The second component was related to facial movement that appears slowly and then maintains a stable state over a long term. Finally, the third component was linked to movement that appears and rapidly returns to the initial state. Local regression analysis was performed to obtain the distribution of the component scores on a two-dimensional plane: pleasure–displeasure and arousal–sleepiness. The third component was negatively correlated with arousal level.

Keywords: Dynamic Facial Expressions; Spontaneous Expressions; Emotion; Motion Capture

Introduction

Many studies have investigated morphological features of facial expressions. Most researches on facial expressions have relied on the use of photographed images or "static" information. However, affective facial expressions are dynamic in nature. Several researchers have claimed that dynamically changing facial configuration plays an important role in the perception of facial patterns and attribution of emotional category labels (Krumhuber, Kappas, & Manstead, 2013). Particularly when either morphological information is limited or intensity of expressions is lower, the presence of facial motion signals helps perceivers to identify emotion (Cunningham & Wallraven, 2009; Bould & Morris, 2008; Krumhuber & Manstead, 2009).

Although the presence of a motion signal is important for emotion recognition, the benefit of dynamic display does not seem to be solely attributed to an increase in the amount of static information. Natural facial events would not occur at a constant speed as in morphing animations; however, they occur in a nonlinear manner. Such nonlinear facial motion leads to more accurate emotion recognition compared to linear motion animation (Wallraven, Breidt, Cunningham, & Bülthoff, 2008). Therefore, for elucidating dynamic facial expressions, complex spatiotemporal information embedded in facial motion is also considered to be at least as important as "static" facial information.

In many studies of facial expressions, portraits of actors' stereotypical emotion expressions or facial actions of specific predefined patterns (e.g., FACS (Ekman & Friesen, 1978)) have been employed. However, it is difficult to capture temporal facial changes using posed expressions. Thus, in the present study, we investigated dynamic facial expressions based on spontaneous facial expressions elicited by emotional film clips. Although some conventional facial action coding techniques based on human observations are designed to code temporal changes in facial expressions, it is difficult to quantitatively describe facial dynamic changes with such coding systems. To obtain dynamic facial configurations, we employed a facial motion capture system.

Facial motion capture systems have been widely used in the film industry. Recently, such landmark-based approaches have been applied to research on spontaneous dynamic facial expressions (Valstar, Gunes, & Pantic, 2007; Zhang, Yin, Cohn, Canavan, Reale, Horowitz, & Girard, 2014). The present study aims at extracting the components of dynamic facial expressions using a combination of the methodologies of geometric morphometrics and the multiway decomposition method PARAFAC (Kroonenberg, 1983), being a type of modified principal component analysis based on three dimensional landmark coordinate data. Furthermore, we investigated the relationship between the extracted facial-expression components and the self-reported emotion of the owners of the faces.

Method

Participants

Japanese undergraduate and graduate students (n = 10: 8 men and 2 women; age: 18 to 24 years, mean age = 21.4, SD = 1.96) participated on a voluntary basis. The participants gave written consent to participate in the study, and Osaka Electro-Communication University ethics committee approved the study.

Procedure

The participants watched thirteen clips intended to elicit differentiated emotional states, such as positive/negative mood, anger, sadness, fear, or disgust. The clips were taken from a database of emotion-eliciting films (Schaefer, Nils, Sanchez, & Philippot, 2010) (Table 1). Each of the clips was shown in a Japanese-dubbed version and was approximately 2 min in length.

Table 1: Sources of film clips used

Title	Release Year	Emotion
The Blair Witch Project	1999	fear
Schindler's List (1)	1993	arousal, negative
There's Something About Mary	1998	amusement
Life Is Beautiful	1997	tenderness, positive
Seven	1995	disgust
Dead Poets Society	1989	positive
The Professional	1994	sadness, tenderness
Trainspotting	1996	disgust
Schindler's List (2)	1993	anger
Misery	1990	arousal
The Dinner Game	1998	amusement, positive
Sleepers	1996	anger
Forrest Gump	1994	tenderness

Participants were instructed to watch all the clips attentively, without diverting their attention from the monitor. When participants were ready to begin the experiment, lighting in the room was dimmed. Each trial started with a countdown leader (3s). It was followed by a presentation of explanatory text of the outline of the narrative of the film (30s), and a presentation of the clip. This procedure was repeated for each film excerpt. The order of items within each set was randomized. Each session lasted approximately 40 min. The film clips were displayed on a 12-inch monitor placed about 1m in front of the participant. The audio of the films was played through speakers placed in front of the participant.

The sequences of 30 facial landmark coordinates were recorded using a facial motion capture system (OptiTrackFLEX: V100R2) at 100Hz (Figure 1). The landmark locations for each participant were decided based on a front facial image that was photographed before the experiment. We also set up a video camera to capture the entire scene and speakers for monitoring and as a reference for synchronization between facial motion sequence and stimuli.

To extract intervals in which facial events occurred from the motion data, we applied principal component analysis (PCA) based on automatic motion segmentation technique to the facial motion sequences (Barbič, Safonova, Pan, Faloutsos, Hodgins, & Pollard, 2004). This technique is based on the observation that simple motions exhibit lower dimensionality than more complex motions. We set the segmentation parameters to $k = 3s$, $\tau = 0.85s$, and $l = 75s$. As a result, all the facial motion sequences of 10 participants were divided into 1325 sections. It is expected that some of the beginnings of the sections (hereafter referred to as segmentation points) correspond to the moments just after facial expressions began. Three observers determined visually whether each segmentation point neighbors the starting point of facial expressions according to the criteria; (1) the segmentation point must be included in the period of film presentation, (2) motion segmentation must be caused by facial movement changes but not by head movements, and (3) a segmentation point that was segmented by solely eye blinking is not the starting point of facial expressions. Consequently, 98 segmentation points were selected as analysis targets. For each selected segmentation point, we extracted a 3-sec length interval of facial motion sequence from 1s before to 2s after the segmentation point.
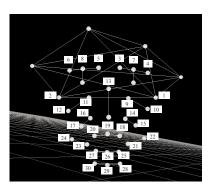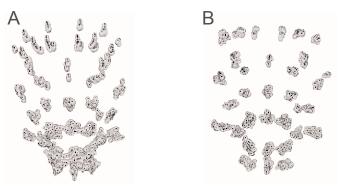

Figure 1: Thirty facial landmark locations

Figure 2: Landmarks of faces (a) before standardization and (b) standardized via the methodologies of geometric morphometrics

Facial Motion Analysis

Each face in each frame differed in location, size, and orientation. To standardize them, we performed a Generalized Procrustes analysis (GPA) on the facial landmarks of all frames of all faces (Bookstein, 1997; Dryden & Mardia, 1998). A GPA is an analytical method used for multivariate statistical analysis of landmark locations expressed in Cartesian coordinates. This method preserves information on the relative spatial relationships of landmarks throughout the standardization. Using this method, we standardized the three dimensional sequential landmark coordinates of facial motion (Figure 2). Moreover, landmark coordinate values were standardized for each coordinate so that the mean equals 0 and the standard deviation equals 1.

This motion data of the participants can be regarded as a three-mode data $\mathbf{X} = \{x_{ijk} | i = 1,2,\dots,I, j = 1,2,\dots,J, k = 1,2,\dots,K\}$ (i corresponds to the time sequence, j corresponds to the landmark, and k corresponds to each extracted interval). To find the components of facial expressions from the motion data, we performed PARAFAC analysis on the three-mode data. PARAFAC is a generalization of PCA to higher order arrays, given by

$$x_{ijk} = \sum_{p=1}^{p} a_{ip}\, b_{jp} c_{kp} + e_{ijk}$$

In the above, a_{ip}, b_{jp}, c_{kp} correspond to the mode of time sequence, and landmark and interval, respectively.

We calculated the principal component scores and loadings up to the third factor with PARAFAC. Figure 3 illustrates the changes of each factor loading concerning time sequence. The first component was found to be a "static" factor, indicating that it was not linked to any facial motion but to static morphological features that are a geometric arrangement of facial parts. However, the second and the third component were related to facial movement. In this study, we examined the second and third components that were linked to facial movement.

The second component score started increasing at around 60 msec and peaked at around 150 msec, after which the score becomes stable. Thus, the second component is considered to be related to a facial movement that appears slowly and maintains a stable state. The third component was linked to a movement that appears at around 50 msec, peaks at 100 msec, and then rapidly returns to the initial state.

To visualize the facial changes along the second and third component, facial landmark movements were reconstructed based on the loadings of both time sequence and landmarks for each component (Figure 4). The results of the reconstruction indicated that a higher second component score of extracted interval was related to eyebrow raising and mouth opening, whereas a lower score was related to eyebrow lowering and mouth closing (Figure 4-A). A higher third component score was related to horizontal movement, including mouth opening in a horizontal direction, and a

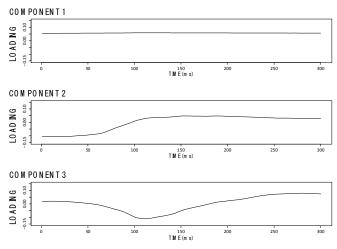lower score was related to a movement of contracting lips into a rounded shape and knitting eyebrows (Figure 4-B).
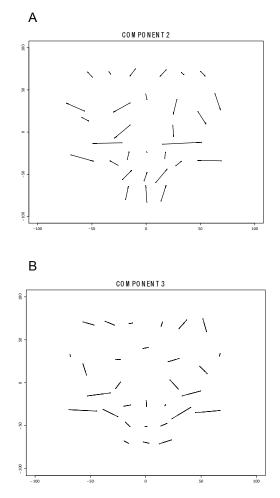

Figure 3: PARAFAC loadings for time-sequence

Figure 4: Facial movement along the second and third component

Assessment of Participants' Affect

Subjective emotional states during observation of emotional film clips were examined using the same film clips as used in the facial motion capture session. The participants were also the same as those in the facial motion capture session. For each participant, we presented each film clip and paused the clip at the moment the facial expression had occurred, which marked the segmentation point. Participants were asked to rate their mood using the Affect Grid technique (Russell, Weiss, & Mendelsohn, 1989) for each moment that their facial expressions had occurred. The Affect Grid is a mood scale that requires participants to place a mark to report their current mood on a 9 x 9 grid with the horizontal dimension representing affective valence (from unpleasantness to pleasantness) and the vertical dimension representing a degree of perceived activation (ranging from sleepiness to high arousal). Both the valence and arousal scores were standardized for each participant.

The relationship between facial motion components and subjective emotional state was investigated. It was considered inappropriate to assume an *a priori* model regarding the relationship between facial expressions and emotion. Therefore, we adopted LOESS, a robust leveling technique based on local polynomial regression (Cleveland, Grosse, & Shyu, 1992), that can graphically demonstrate the relationship between emotion and facial motion components.

The relationship between Affect Grid scores and smoothed component scores of both the second component and the third component is shown in Figure 5 (*span*, kernel function range was set as 3.0). Figure 5-A shows that both negative and low arousal affect lead to higher scores for the second component, suggesting that the component of facial movement that appears slowly is related to expressions both of emotional valence and arousal. For this component, both positive and high arousal expressions are similarly correlated to eyebrow raising and mouth opening. On the other hand, it can be seen from Figure 5-B that arousal level was basically linked to the third component that reflects short-term facial deformations, except for the cases of strong unpleasant affect. As for this component, a higher arousal level corresponds to the movement of contracting lips into a rounded shape and knitting eyebrows.
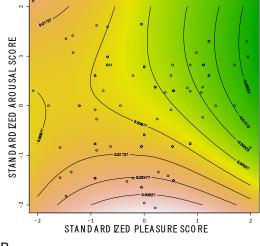
Discussion

This study suggests that dynamic facial expressions consist of plural components differing in spatiotemporal characteristics (i.e., long-term facial deformations and short-term facial deformations). One component was found to be connected to facial deformations that appear slowly and maintains a stable state over a long term, and the other was connected to rapid appearance and disappearance of facial deformations. Each component corresponds to movements in different directions and of different facial parts. The findings of the present study indicate that dynamically changing facial expressions can be described by synthesizing a few components of facial movement that differ in spatiotemporal characteristics.

Moreover, we show the connection between the components of expressions and emotional valence and arousal. The results suggested that long-term facial deformation was related both to valence and arousal whereas short-term deformation was related solely to arousal.

The results of the study also suggest that the combination of motion segmentation technique, methodology of geometric morphometrics, and modified principal component analysis is a valid method for finding components of dynamic facial expressions.
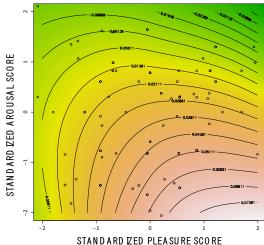
A

B

Figure 5: Contour of component scores on the plane of pleasure–displeasure (horizontal) and arousal–sleepiness (vertical)

Because of the small number of participants in this study, it is difficult to consider the facial components found were stable. Future studies should examine the accuracy of PARAFAC-based expression model in detail with a larger number of participants.

Acknowledgments

References

Barbič, J., Safonova, A., Pan, J. Y., Faloutsos, C., Hodgins, J. K., & Pollard, N. S. (2004, May). Segmenting motion capture data into distinct behaviors. In *Proceedings of the 2004 Graphics Interface Conference* (pp. 185-194). Canadian Human-Computer Communications Society.

Bookstein, F. L. (1997). *Morphometric tools for landmark data: geometry and biology*. Cambridge University Press.

Bould, E., & Morris, N. (2008). Role of motion signals in recognizing subtle facial expressions of emotion. *British Journal of Psychology*, *99*(2), 167-189.

Cleveland, W. S., Grosse, E., & Shyu, W. M. (1992). Local regression models. *Statistical models in S*, 309-376.

Cunningham, D. W., & Wallraven, C. (2009, September). The interaction between motion and form in expression recognition. In *Proceedings of the 6th symposium on applied perception in graphics and visualization* (pp. 41-44). ACM.

Dryden, I. L., & Mardia, K. V. (1998). *Statistical shape analysis* (Vol. 4). Chichester: Wiley.

Friesen, E., & Ekman, P. (1978). Facial action coding system: a technique for the measurement of facial movement. *Palo Alto*.

Kroonenberg, P. M. (1983). *Three-mode principal component analysis: Theory and applications* (Vol. 2). DSWO press.

Krumhuber, E. G., Kappas, A., & Manstead, A. S. (2013). Effects of dynamic aspects of facial expressions: A review. *Emotion Review*, *5*(1), 41-46.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: a single-item scale of pleasure and arousal. *Journal of personality and social psychology*, *57*(3), 493.

Valstar, M. F., Gunes, H., & Pantic, M. (2007, November). How to distinguish posed from spontaneous smiles using geometric features. In *Proceedings of the 9th international conference on Multimodal interfaces* (pp. 38-45). ACM.

Wallraven, C., Breidt, M., Cunningham, D. W., & Bülthoff, H. H. (2008). Evaluating the perceptual realism of animated facial expressions. *ACM Transactions on Applied Perception (TAP)*, *4*(4), 4.

Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., & Girard, J. M. (2014). BP4D-Spontaneous: a high-resolution spontaneous 3D dynamic facial expression database. *Image and Vision Computing*, *32*(10), 692-706.