2013 Humaine Association Conference on Affective Computing and Intelligent Interaction

Computational Analysis of Emotion Dynamics

Ayesha Hakim
Massey University, New Zealand
Email: a.hakim@massey.ac.nz

Stephen Marsland
Massey University, New Zealand
Email: s.r.marsland@massey.ac.nz

Hans W. Guesgen
Massey University, New Zealand
Email: h.w.guesgen@massey.ac.nz

***Abstract*—Emotions are dynamic. They vary continuously with regard to intensity, duration, persistence with time, and other attributes. In addition, their appearance on the face of subjects varies, and the transition in facial expressions is based on both the change in emotion and physiological constraints. In this paper, we examine the trajectories between emotions in activation-evaluation space and show that emotion trajectories are smooth and follow 'common' paths between different emotions.**

***Keywords*-emotion dynamics; emotion trajectories; activation-evaluation space.**

I. Introduction

While psychologists have been trying to define emotions for over 100 years, a non-subjective definition still eludes us. More recently, there has been significant efforts to computationally recognise, represent, and analyse emotions based on their appearance such as in the face, or tone of voice [1]. In this paper we focus on the change in emotions over time; research in psychology [2] and neuroscience [3] has shown that emotions change (more specifically, *fade*) over time due to either external stimuli or their natural progression. We focus on the transitions between six 'basic' emotions states (happiness, excitement, anger, frustration, sadness, and neutral) as they appear in activation-evaluation space.

In this paper, we use facial expressions as the representative of underlying emotional states. Just like the underlying emotions, facial expressions of emotions change over time, due to external stimuli and the way the face works. The facial dermal tissues comprise collagen (72%) and elastin (4%) fibres which help resist deformation of tissues. Therefore, the facial tissues effected by the active facial muscles need to relax before stretching to another form (i.e., expressing another emotion) [4]. The movement of facial points between different emotional states is analysed by tracking the paths of change in facial expressions through time. We do not consider the effect of external stimuli on the facial expressions of emotions.

The paths of emotional expressions are observed on the valence-activation (or activation-evaluation) space, which is the most widely used bipolar circumplex model for representing emotions in psychological studies [5]. This model represents emotions based on activation (how motivated a person is by that emotion) and valence (how positive or negative an emotion is). It forms a circular representation of emotion space in which each emotion is located at a specific angle based on its similarity to other emotions. The radial distance from the centre of the circle represents the intensity of emotion. The greater the distance, the stronger the emotion and vice-versa [6]. Sometimes, the intensity of emotion is too slight to notice, or too subtle to effect what we do. Ekman [7] therefore suggested that we may better say that there are times when there is no emotion, i.e., the 'neutral' state. Neutral lies at the centre of the circle and is assumed to be the presence of any emotion with very low intensity.

II. Dataset

In practical applications of emotion recognition the ideal would obviously be to deal with images of the face alone. However, for simplicity we have chosen to start with a simpler problem. We used the Interactive Emotional Dyadic Motion Capture (IEMOCAP) dataset, in which a pair of actors were recorded using a high-speed camera capturing 120 frames per second. One of the actors had a set of reflective markers on their face and the 3D positions of these markers was tracked with very high accuracy [8]. We based our analysis on the locations of these marker points. The videos of the actors were watched by three human experts who labelled the dataset, providing a ground truth labelling of the data. However, the experts did not always agree. Each conversation consists of almost 50 utterances with an average duration of 4.5 seconds each, It is these utterances that were annotated by the three human evaluators into categorical labels (neutral, happy, angry, sad, surprise, disgust, fear, frustration, and excitement) as well as psychological data about emotion intensity (valence, activation, and dominance). Although nine emotional labels were used by the humans, we chose to consider only six of them, as for the missing emotions (disgust, surprise, and fear) there was insufficient data. For further details on any part of the data capture and labelling, see [8].

We consider the two actors separately in everything that follows. We started by creating a training set of 4,000 frames of each of the six emotions (five basic emotions and neutral) to make a total set of 24,000 frames for each of the actors. These 24,000 frames were chosen from the set where all three human experts agreed on the annotation. We then created a test set based on seven continuous conversations, each lasting around 3 minutes, comprising almost $152,000$ frames in total.

III. Methodology

We begin by reducing the size of the data for each frame from 61 to 28 marker points, each comprised of a 3D position. We then generated three different shape models, on the full face (28 markers), the upper face (17 markers), and the lower face (11 markers) separately [9]. 4 principal components

978-0-7695-5048-0/13 $26.00 © 2013 IEEE
DOI 10.1109/ACII.2013.37

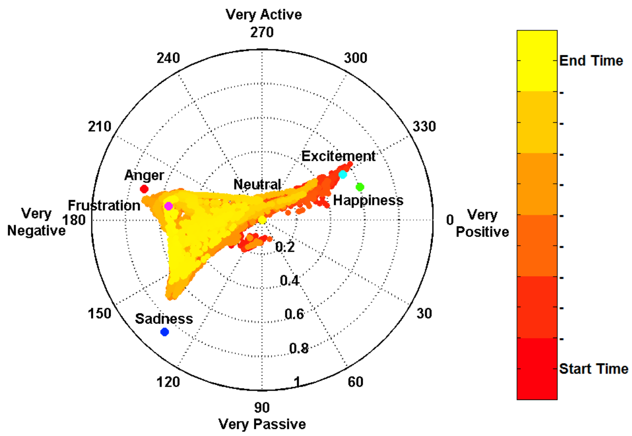


Fig. 1: The mapping of continuous emotions during a conversation (Ses01F_impro01) onto the activation-evaluation space. The movement of emotions through time is represented by changing colour spectrum from dark/red (start) to light/yellow (end).

were used for each of these models to represent the possible variation they showed, and clusters representing the six 'basic' emotions were generated from training data. Locations for these basic emotions were identified that broadly match those proposed by Whissell [10]. The Mahalanobis distance was used to classify each new test frame as a mixture of von Mises distributions using the methodology described in [11].

Fig. 1 shows the mapping of continuous emotions through time corresponding to an unsegmented conversation in the activation-evaluation space. The colour variation represents time, ranging from red (dark in grayscale) to yellow (light in grayscale). The figure shows some sequence information in the transition from one emotion state to another. This visual pattern and the effect of mechanical properties of human skin on facial expressions motivate us to analyse these emotion trajectories. We propose the following hypotheses about the emotion paths representation in the activation-evaluation space:

1) The paths form 'smooth' trajectories in the space.
2) If the end-point emotions are not positively correlated, then the path goes through the neutral state.
3) If the end-point emotions are positively correlated, then the path does not go through the neutral state.

IV. Evaluation

Within the following section the processes used to test the proposed hypotheses are presented, together with the results of those tests.

Hypothesis 1: The paths between emotions form 'smooth' trajectories in the space: In order to test the first hypothesis, we first need to define 'smoothness' of an emotion trajectory. After extensive review of the psychological literature, we could not find any standard definition. However, if we consider an emotion trajectory as a time series (sequence of values at successive time points following a non-random order), then its smoothness may be defined as a measure of its persistence with time. A random time series, e.g., Brownian motion is not smooth, as it is not persistent with time. On the basis of this definition, we may say that if the points in the emotion space move in a predictable manner then the resulting trajectory is smooth/persistent with time.

We measure the smoothness of emotion points trajectories in the activation-evaluation space using two approaches: *first*, by measuring the time derivative of angular displacement (angular velocity) and *second*, by estimating the Hurst exponent (H) [12]. The time derivative of angular displacement determines the change in the angle with time; the smaller the change, the more the smoothness of trajectory and vice-versa. It is calculated as,

$$\dot{\theta}_t = \theta_t - \theta_{t-1}$$

where $t = 2, 3, \cdots, n$ and $n =$ total number of frames in the video. Fig. 2(a) plots the time derivatives of angular displacement of the emotion points during each conversation in the testing set. Smaller values imply smooth emotion trajectories.

The Hurst exponent is a statistical measure of persistence and predictability of a time series, calculated by rescaled range $\left(\frac{R_t}{S_t}\right)$ analysis. The greater the value of H $(0.5 < H < 1)$, the smoother the time series, $H = 0.5$ means a random time series. Fig. 2(b) plots the the Hurst exponent as a measure of smoothness of emotion points trajectories in the activation-evaluation space for each conversation in the testing set. We estimated H for the time series representing the size of 'change' between pairs of consecutive points in the space as a function of valence and activation. Suppose X_t denotes the time series where $t = 2, 3, \cdots, n$ and $n =$ total number of frames in the video. The size of 'change' for each frame is calculated as the Euclidean distance between the time derivative of valence $(\dot{V}_t)$ and the time derivative of activation $(\dot{A}_t)$ in the activation-evaluation space:

$$\dot{V}_t = \parallel V_t - V_{t-1} \parallel$$
$$\dot{A}_t = \parallel A_t - A_{t-1} \parallel$$

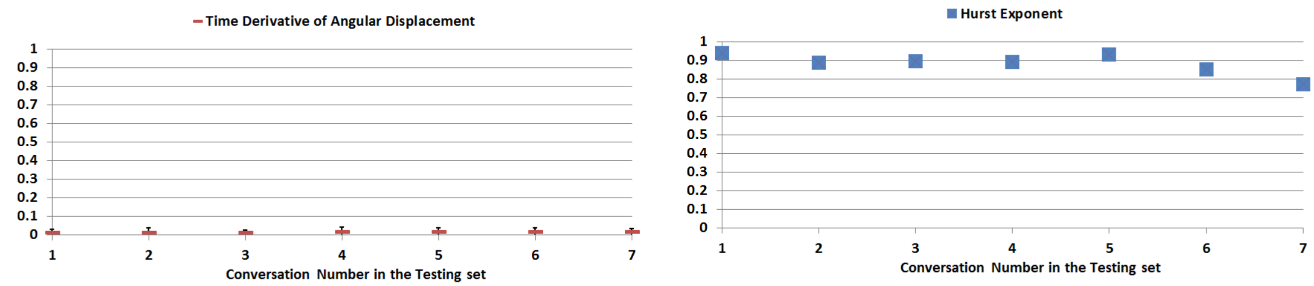
where the t index represents the t^{th} element of the time series.

$$\text{size of change} = \sqrt{\dot{V}_t^{\,2} + \dot{A}_t^{\,2}} \quad (1)$$

The rescaled range R of time series (X_t) is calculated by:

1) Calculate the mean-centred time series $Y_t = X_t - \mu$, where μ is the mean of the time series.
2) Calculate the cumulative sum of Y_t,

$$Z_t = \sum_{i=1}^{t} Y_i$$



(a) (b)

Fig. 2: Test of smoothness of emotion paths corresponding to each conversation in the testing set

3) Calculate the rescaled range R_t of time series,

$$R_t = \max(Z_1, Z_2, \cdots, Z_t) - \min(Z_1, Z_2, \cdots, Z_t)$$

S_t is the standard deviation of the time series. The ratio $\frac{R_t}{S_t}$ scales as a power law with time so that:

$$H = \frac{log((\frac{R}{S})_t) - c}{log(t)}$$

where c is a constant and the slope of the regression line (R/S versus t in log-log axes) approximates the Hurst exponent.

Fig. 3 plots the size of 'change' between two consecutive emotion points in the activation-evaluation space for each conversation in the testing set separately. The histogram shows that the size of 'change' (which is the motion within 120^{th} of a second) is mostly very small. However, in a few cases the size of change is bigger. In order to find the reason behind these intensity jumps (those beyond the first standard deviation), we monitored those paths of trajectories and compare them with the original videos in the dataset. We noticed that the bigger size of change in the trajectories are false positives due to closing the eyes. We had tried to avoid this by removing the eyelid markers, but still the closing of eyes is captured by the muscles around the eyes, especially those near the eyebrows. Fig. 1 also shows that false positives arise due to closing the eyes. As lowering eyelids/eyebrows is one of the expressions of sadness, the outliers in the emotion space lies in the direction of sadness. It should be noted that the size of 'change' refers to the size of displacement between the two consecutive points in the activation-evaluations space, not the displacement of markers on the face.

Hypotheses 2 and 3: Transitions between uncorrelated or negatively correlated emotions need to pass through the neutral state while transitions between positively correlated emotions do not: According to differential emotion theory, each discrete emotion is related to certain other discrete emotions with a distinct pattern [13]. In the activation-evaluation space, emotions lie along particular angles on the basis of their similarity measures. The neighbouring emotions (i.e., those separated by only a few degrees) are assumed to be *positively correlated*, those 90° apart are considered as *uncorrelated*, while those 180° apart are *negatively correlated* [14].

While looking at continuous trajectories in the activation-evaluation space (e.g., Fig. 1), we noticed some patterns of emotion transitions. We fitted regression lines to the trajectories between six emotions (neutral, happiness, excitement, anger, frustration, and sadness). There are 15 possible symmetric paths between the six emotions, but the regression analysis shows that actually only seven paths are represented for all possible transitions between these emotions. For instance, the *smiling* expression (getting happy) shows a linear relationship of valence and activation between neutral and happiness. The onset of smiling occurs at neutral, reaches apex at some intensity of happiness and returns to neutral for offset. In the case of *laughing*, which is another expression of happiness, the same pattern repeats several times depends on its duration and intensity.

By observing the trajectories between uncorrelated and negatively correlated emotional states, we found that the

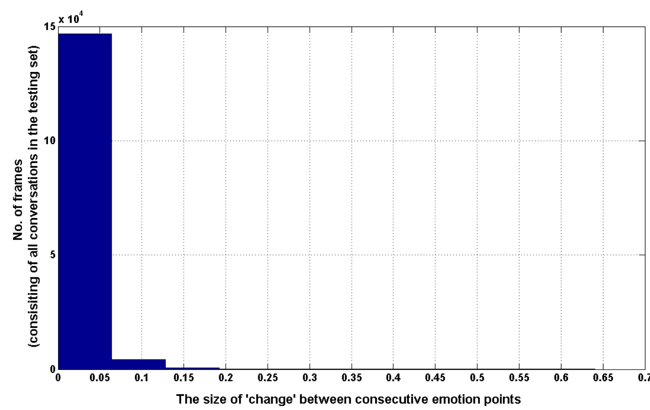
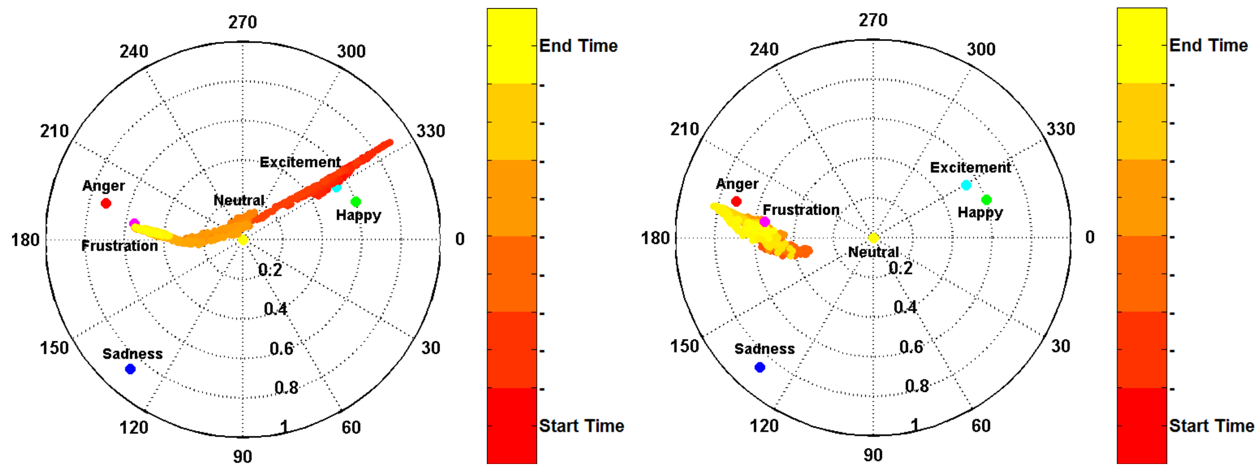


Fig. 3: The size of the 'change' between consecutive emotion points in the activation-evaluation space, for all conversations in the testing set. The maximum possible motion is 2, since the circle is radius 1.



(a) Continuous transition between excitement and frustration, which are negatively correlated emotions

(b) Continuous transition between anger and frustration, which are positively correlated emotions

Fig. 4: Transitions between negatively correlated emotions tend to pass through the neutral state, while transitions between positively correlated emotions do not.

transition between these states tends to pass through the neutral state. The intensity of the current emotion must decrease to neutral (shown as linear motion along a radial line), before the intensity of the next emotion increases. Fig. 4(a) shows a trajectory followed by the transition from excitement to frustration. However, the positively correlated emotions (such as anger and frustration, as well as happiness and excitement) may move from one state to another with slight change in intensity and angle simultaneously, as shown in Fig. 4(b). These findings are also supported by the mechanical properties of facial dermal tissues [4]. Under low stress (transition between positively correlated emotions), dermal tissue applies low resistance to stretch as the collagen fibres uncoil in the direction of the strain. However, under high stress conditions (transition between negatively or uncorrelated emotions), the elastin fibres behave like elastic springs to return the collagen fibres to their original no-stress condition. According to these properties, to express a very different emotion the facial muscles have to pass through a 'no-stress' condition.

Table I shows a comparison between the coefficient of determination (R^2) of linear, quadratic, and cubic polynomial regression models fitted to each of the seven symmetric paths (i.e., between neutral and happiness, neutral and excitement, neutral and anger, neutral and frustration, neutral and sadness, anger and frustration, and happiness and excitement). For the paths from neutral to any of the five emotions, there is no significant difference between the R^2 values of linear, quadratic and cubic regression, which implies that a linear model may be used to fit these paths. For the path between neighbouring emotions such as anger to frustration as well as happiness to excitement, the quadratic regression is significantly better than the linear regression. Moreover, there is no significant difference between quadratic and cubic regressions, which implies that quadratic model may be used to fit the trajectories between the neighbouring emotions onto the activation-evaluation space.

On the activation-evaluation space, the travel along the emotion flows/trajectories is a matter of intensity change and the angle change. As already discussed, the emotion trajectories follow 'common' paths; which in turn suggests that there exists some relationship between the intensity change and angle change through time. In order to analyse this relationship, we plot the polar coordinates (r: changing intensity, θ: changing emotion) of continuous points in the space during emotion transitions. Fig. 5 consists of four subplots, the first and second subplots show r and θ respectively. In these plots, the three horizontal lines represent the mean and mean $\pm$ 1 standard

Emotion Transitions	Linear (R^2)	Quadratic (R^2)	Cubic (R^2)
Neutral-Anger	0.9131	0.9297	0.9139
Neutral-Frustration	0.9642	0.9854	0.9924
Neutral-Happiness	0.9223	0.9323	0.9418
Neutral-Excitement	0.9374	0.9383	0.9388
Neutral-Sadness	0.902	0.9144	0.9164
Anger-Frustration	0.0922	0.6494	0.6497
Happiness-Excitement	0.0457	0.5914	0.6083

TABLE I: Coefficient of determination (R^2) of linear, quadratic, and cubic polynomial regression models fit to the seven symmetric paths of emotion transitions into the activation-evaluation space.

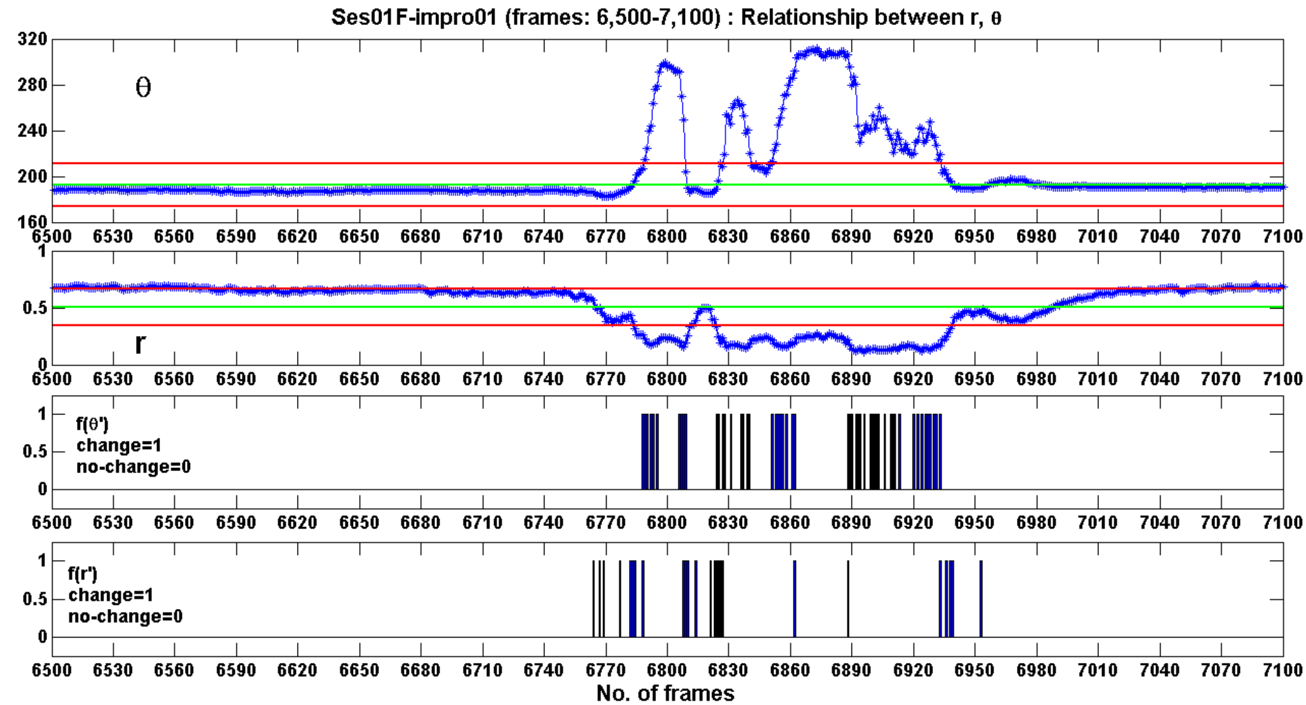


Fig. 5: Ses01F_impro01, continuous frames during emotion transition from angry to happy. The temporal relationship between intensity change and angle change can be seen.

deviation. The third and fourth subplots show change versus no-change using a binary plot by applying the following rules to the time derivatives $(\dot{r}_t, \dot{\theta}_t)$ of (r, θ) respectively (where the t index represents the t^{th} element of the time series):

$$f(\dot{\theta}_t) = \begin{cases} 0, & \text{if } \mu_{\dot{\theta}_t} - \sigma_{\dot{\theta}_t} < \dot{\theta}_t < \mu_{\dot{\theta}_t} + \sigma_{\dot{\theta}_t} \\ 1, & \text{otherwise} \end{cases}$$

$$f(\dot{r}_t) = \begin{cases} 0, & \text{if } \mu_{\dot{r}_t} - 2\sigma_{\dot{r}_t} < \dot{r}_t < \mu_{\dot{r}_t} + 2\sigma_{\dot{r}_t} \\ 1, & \text{otherwise} \end{cases}$$

Fig. 5 shows that there is a relationship between angle change and intensity change such that whenever there is a large 'change' in angle (according to the given rules), the intensity decreases.

V. Validation

We used 4-fold cross validation to evaluate the consistency of our models. We randomly chose 10,000 frames of each of the six emotions to form a total set of 60,000 frames. This data includes all frames from the original training set that was used to build the shape models, as well as another 36,000 frames that represented part of a conversation that included emotion changes. We set up four different datasets, each of which contained 45,000 frames for training and 15,000 frames for testing. Each of these datasets was used to build new shape models, and then map the emotion trajectories in activation-evaluation space using the method described in section III. The purpose of this validation is to test the consistency of each model and the reliability of the obtained results.

Fig. 6 shows the outputs of the emotion trajectories for the four test sets. It can be seen that the first, third, and fourth show very similar shapes to each other, but that the second one does not have any instances of sadness. However, otherwise the match is still very good.

VI. Conclusion

In this paper, we have presented an analysis of emotion trajectories in activation-evaluation space based on shape models of facial points. On the basis of trajectory-level analysis, we evaluated some hypotheses related to the smoothness of emotion trajectories, and the 'common' paths between emotions based on their correlation.

By measuring the size of 'change' between consecutive frames, we found that the emotions move in a continuous flow, which implies that there are no sudden jumps within the trajectories. Further, we measured the smoothness of continuous emotion trajectories on the basis of the time derivative of angular displacement and estimated the Hurst exponent, which suggests that the emotion trajectories are smooth and persistent with time. This trending property of emotions implies that it

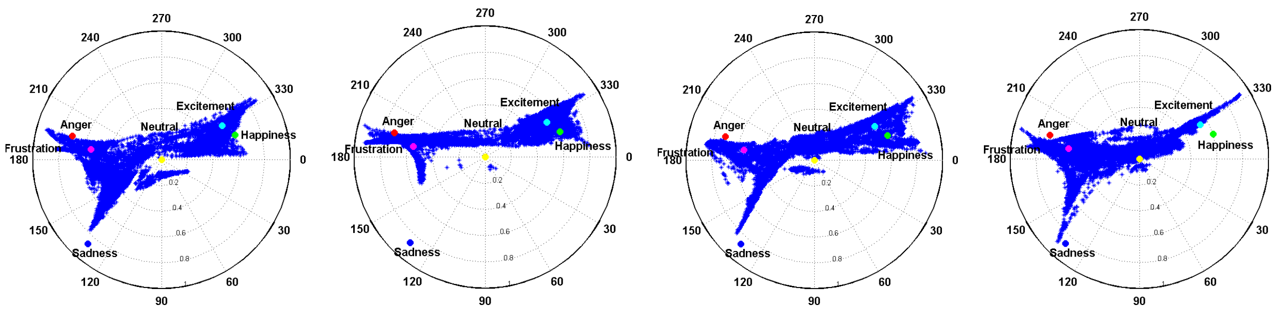


Fig. 6: Outputs of the emotion trajectories on the test sets in activation-evaluation space.

may be possible to predict the future emotions based on past and present emotional states in the absence of external stimuli.

By visualising the emotion trajectories, it appeared that there may be 15 symmetric paths between the six emotions in the activation-evaluation space. To test it, we fitted regression lines to the trajectories and found that there are actually 7 symmetric paths to travel between these six emotions. We showed that the trajectories between uncorrelated and negatively correlated emotions can be fitted with straight lines, while the trajectories between positively correlated emotions are fitted well by a quadratic regression model.

By analysing the relationship between the change in angle and change in intensity, we may conclude that the transition between negatively correlated or uncorrelated emotions causes a decrease in intensity, while the transition between positively correlated emotions may occur with a slight change in intensity and angle simultaneously.

The presented analysis might be used and extended in several directions, such as examining the 'abnormal' paths of emotions, which might give some cues about underlying deception, or some illness. The mapping of continuous trajectories to the activation-evaluation space might be a useful tool to build emotional conversation agents displaying realistic emotions and going through smooth emotion transitions.

ACKNOWLEDGEMENTS

This work was supported by funds from the Higher Education Commission (HEC) Pakistan. The authors would like to thank Massey University Smart Environments (MUSE) Group for providing useful feedback and support throughout the research. We also wish to thank Claude McCarthy Scholarship for funding the conference participation.

REFERENCES

[1] G. Matthews, M. Zeidner, and R. D. Roberts, Emotional intelligence: Science and myth, Cambridge, MA: MIT Press, 2002.
[2] T. Ritchie, J. Skowronski, J. Hartnett, B. Wells, and W. Richard Walker, "The fading affect bias in the context of emotion activation level, mood, and personal theories of emotion change," Memory, vol. 17, pp. 428-444, 2009.
[3] W. Cunningham and P. Zelazo, "Attitudes and evaluations: A social cognitive neuroscience perspective," Trends in cognitive sciences, vol. 11, no. 3, pp. 97-104, 2007.
[4] D. Terzopoulos and Keith Waters, "Physically-based facial modeling, analysis, and animation," Visualization and Computer Animation, 1990.
[5] R. Plutchik and H. R. Conte, Circumplex Models of Personality and Emotions, APA, Washington, 1997.
[6] R. Plutchik, Emotions and Life: Perspectives from psychology, biology, and evolution, American Psychology Association, Washington, 2002.
[7] P. Ekman, Emotions revealed: Recognizing faces and feelings to improve communication and emotional life, Holt Paperbacks, 2007.
[8] C. Busso, M. Bulut, C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan, "IEMOCAP: Interactive Emotional Dyadic Motion Capture Database," Language Resources And Evaluation, vol. 4, no. 42, pp. 335-359, 2008.
[9] A. Hakim, S. Marsland, and H. Guesgen, "A Robust Joint Face Model for Human Emotion Recognition," Proc. Conference on Image and Vision Computing New Zealand (IVCNZ), 2012.
[10] C. Whissell, "The dictionary of Affect in Language," in Emotion: Theory, research and experience, vol. 4, R. Plutchik and H. Kellerman Eds., Academic Press, 1989.
[11] A. Hakim, S. R. Marsland, and H. W. Guesgen, "Statistical Modelling of Complex Emotions using Mixture of von Mises Distributions," Proc. Humaine Association Conference on Affective Computing and Intelligent Interaction, in press.
[12] B. Qian and K. Rasheed, "Hurst Exponent and Financial Market Predictability," Proc. International Conference on Financial Engineering and Applications, pp. 203-209, 2004.
[13] S. H. Blumberg and C. E. Izard, Patterns of emotion experiences as predictors of facial expressions of emotion, Merrill-Palmer Quarterly, 37, pp. 183197, 1991.
[14] J. Larsen, P. McGraw, and J. Cacioppo, "Can People Feel Happy and Sad at the Same Time?," Journal of Personality and Social Psychology, 81, pp. 684696, 2011.