

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

# COMPUTATIONAL MODELLING TO TRACK HUMAN EMOTION TRAJECTORIES THROUGH TIME



A THESIS PRESENTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE  
DEGREE OF  
*Doctor of Philosophy*  
IN  
*Computer Science*  
AT SCHOOL OF ENGINEERING AND ADVANCED TECHNOLOGY,  
MASSEY UNIVERSITY, PALMERSTON NORTH,  
NEW ZEALAND.

**AYESHA HAKIM**

2013



© 2013 by Ayesha Hakim

All rights reserved.



# Acknowledgements

I owe the most sincere gratitude to my supervisors, Prof. Hans Guesgen and Prof. Stephen Marsland, for the continuous support during my Ph.D. study and research. Their patience, kind guidance, friendship, and immense knowledge helped me not only in my research and thesis writing, but also to deal with the daily-life challenges. Of course, it is not easy to do a PhD with a young family! I couldn't have asked for more supportive advisors and talented teachers.

Thanks to all the members of Massey University Smart Environment (MUSE) group for their encouragement, insightful comments, and valuable feedback on my research and presentations. Thanks to Michele Wagner for all the administrative support throughout my degree.

I would like to thank my office-mate and friend, Le Thu, for keeping my spirits high and giving useful advice to cope with personal and study-life problems. Many thanks to An, Linda, and Amissa for providing me with an ear to listen after hard comments from the supervisors. We did a lot of back-biting!

Heaps of acknowledgement go to Higher Education Commission, Pakistan for funding my research. Thanks to Universities New Zealand for granting me the Claude McCarthy Fellowship to participate in the conference. Thanks to Dilantha for kind guidance related to account issues, and Natalia for her friendly attitude and kind support for university issues. Many thanks to the authors of IEMOCAP dataset for

providing me access to the continuous unsegmented data. A special thanks to Prof. Carlos Busso for helping me understand and use this dataset throughout the research.

Last but not the least, I would like to thank my parents Abdul Hakim Khanzada and Nasira Hakim and sisters Sahar, Mahrukh, Rabia, who encouraged me to go for Doctoral studies and gave me full support mentally and emotionally throughout my PhD, and my life. Many thanks to my husband Tariq and son Ali, who gave me happy moments after a tiring day at Massey. I love and respect all of you from the depth of my heart.

# Abstract

There has been a lot of research into the field of affective computing over the past three decades. In the context of this thesis, affective computing is the computing that relates to emotion recognition, representation, and analysis. Much of the past work has focused on the basic emotions. However, most human emotions are not pure examples of one basic emotion, but a mixture of them, known as complex emotions. Emotions are dynamic, they change continuously over time. This thesis focuses on computational modelling to recognise, represent, and analyse continuous spontaneous emotions through time.

Emotions are internal, and hence impossible to see directly. However, there are some external presentations of emotions enabling computational tools to be used to identify them. This thesis focuses on the use of facial points as a measure of underlying emotions. The main focus is the development of computational models to track the patterns of facial changes in order to analyse the paths followed by emotions over time.

While there has been lots of work on shape models to classify facial expressions into discrete basic emotion categories, they are generally based on the analysis of the full face. However, the research shows that some expressions are better recognized by muscle activity in the upper half of the face, while others use muscles primarily from the lower half of the face. This thesis introduces a joint face model based on



shape models of full, upper, and lower parts of the face separately that significantly improves the accuracy.

The set of shape models gives a degree of match to each basic emotion. Using this information, this thesis addresses the problem of complex emotion recognition by developing a mixture model that combines each basic emotion in an appropriate amount. The proposed model represents emotions in the activation-evaluation space, which is the most widely-used representation of emotions in psychological studies. It represents emotions on the basis of their polarity and similarity to each other. This thesis uses a mixture of von Mises distributions for emotion recognition, which is an approximation to the normal distribution for circular data and is the most common model for describing directional data. The results show that the proposed mixture model fits the data well.

Emotions vary continuously with regard to intensity, duration, persistence with time, and other attributes. In addition, their appearance on the face varies, and the transition in facial expressions is based on both the change in emotion and physiological constraints. This thesis examines the trajectories between emotions in activation-evaluation space and shows that these trajectories are smooth and follow ‘common’ paths between different emotions. In the past, very few efforts have been made on the analysis of continuous emotion dynamics. The findings presented in this thesis can be used and extended in several directions to improve the emotion recognition as well as emotion synthesis.

For my papa and mama



# Contents

<b>Acknowledgements</b>	<b>v</b>
<b>Abstract</b>	<b>vii</b>
<b>1 Introduction</b>	<b>13</b>
1.1 Motivation . . . . .	13
1.2 The Problem . . . . .	15
1.3 Scope of the Study . . . . .	17
1.4 Aims, objectives, and hypotheses . . . . .	18
1.5 Experimental Methodology . . . . .	20
1.6 Main Research Contributions of Thesis . . . . .	21
1.7 Overview of Thesis . . . . .	23
<b>2 Emotions from the Perspective of Psychology</b>	<b>27</b>
2.1 Introduction . . . . .	27
2.2 Understanding Emotions . . . . .	28
2.3 Basic and Complex/Mixed Emotions . . . . .	29
2.4 Emotions and Facial Expressions . . . . .	31
2.5 The Space of Emotions . . . . .	32
2.5.1 Circumplex models of emotions . . . . .	33

2.5.2	Activation-evaluation space . . . . .	35
2.6	Categorical and Attribute-based Descriptions . . . . .	38
2.7	Summary . . . . .	39
<b>3</b>	<b>Literature Review</b>	<b>41</b>
3.1	Introduction . . . . .	41
3.2	Direct Classification Approaches . . . . .	42
3.2.1	Template-based techniques . . . . .	42
3.2.2	FACS-based techniques . . . . .	45
3.2.3	Rule-based techniques . . . . .	47
3.3	Mapping to Emotion Space Approaches . . . . .	48
3.3.1	Quantised approaches . . . . .	48
3.3.2	Continuous approaches . . . . .	51
3.4	Analysis of Emotion Dynamics . . . . .	54
3.5	Summary . . . . .	55
<b>4</b>	<b>Datasets for Emotion Recognition and Analysis</b>	<b>57</b>
4.1	Introduction . . . . .	57
4.2	Criteria for Selecting a Dataset . . . . .	58
4.3	Review of Datasets . . . . .	59
4.4	Data Annotation Tools . . . . .	68
4.5	The Selected Dataset . . . . .	71
4.6	Description of IEMOCAP . . . . .	73
4.7	Data Preprocessing . . . . .	76
4.8	Summary . . . . .	80
<b>5</b>	<b>Basic Emotion Recognition</b>	<b>81</b>
5.1	Introduction . . . . .	81

5.2	Shape Models . . . . .	81
5.2.1	Using the upper and lower face separately . . . . .	85
5.2.2	Classification . . . . .	88
5.3	The Effects of Each Principal Component . . . . .	89
5.3.1	Effect of each full, upper, and lower face PC . . . . .	90
5.4	Performance Comparison . . . . .	94
5.4.1	A rule-based emotion classifier . . . . .	94
5.4.2	Support Vector Machines . . . . .	98
5.5	Results . . . . .	98
5.5.1	Robustness to mislabelled data . . . . .	105
5.6	Discussion . . . . .	105
<b>6</b>	<b>Statistical Modelling of Complex Emotions using Mixtures of von Mises Distributions</b>	<b>109</b>
6.1	Introduction . . . . .	109
6.2	Analysis of Circular Data . . . . .	110
6.2.1	Circular distribution . . . . .	111
6.2.2	Statistical approaches to modelling circular data . . . . .	112
6.2.3	Selection of an appropriate circular distribution . . . . .	114
6.3	Description of Our Method . . . . .	115
6.3.1	Mapping emotions to the activation-evaluation space . . . . .	116
6.3.2	von Mises Mixture Model . . . . .	117
6.3.3	Estimating the Parameters of the Mixture Model . . . . .	119
6.3.3.1	Estimating the concentration parameter . . . . .	119
6.3.3.2	Estimating the weights . . . . .	120
6.4	Experimental Results . . . . .	120
6.5	Summary . . . . .	128

<b>7</b>	<b>Computational Analysis of Emotion Dynamics</b>	<b>131</b>
7.1	Introduction . . . . .	131
7.2	Hypotheses . . . . .	132
7.3	Evaluation . . . . .	133
7.4	Validation . . . . .	144
7.5	Conclusion . . . . .	144
<b>8</b>	<b>Conclusions</b>	<b>149</b>
8.1	Research Overview and Significant Findings of the Thesis . . . . .	149
8.2	Future Research . . . . .	157
8.2.1	Using full facial images instead of marker locations . . . . .	157
8.2.2	Emotion recognition using multiple modalities . . . . .	158
8.2.3	Adding contextual information . . . . .	159
8.2.4	Observing the time-offset of emotion activation and recovery . . . . .	160
8.2.5	Life-long learning . . . . .	161
8.3	Conclusions . . . . .	163
<b>A</b>	<b>List of Publications</b>	<b>165</b>
<b>B</b>	<b>Emotion Words from Whissell and Plutchik</b>	<b>167</b>
<b>C</b>	<b>The effects of varying the second principal component on the female mean upper face</b>	<b>171</b>
<b>D</b>	<b>Glossary</b>	<b>175</b>
	<b>References</b>	<b>177</b>

# List of Tables

2.1	List of basic emotions by various theorists. Presented by Ortony and Turner [163]. . . . .	30
2.2	Emotion classes in the activation-evaluation space [221] . . . . .	37
4.1	A review of existing visual (V) and audio-visual (AV) datasets for human emotion recognition and analysis with respect to the presented criteria. . . . .	67
5.1	The effect of each PC on the mean face. For explanation, see the text.	90
5.2	Rules for classifying emotions based on the direction and magnitude of the selected full face principal components and the corresponding vocabulary of FAPs associated with the expression of each basic emotion. The vocabulary of FAPs is taken from [181]. The description of each of the mentioned FAP is listed in Table 5.3. The relationship between PCs and the corresponding FAPs is discussed in the text. . . . .	95
5.3	Description of each MPEG-4 FAPS mentioned in Table 5.2. . . . .	96
5.4	Mean accuracy and standard deviation of the joint face model, full, upper, lower face models, SVM Classifier, and the rule-based Classifier on both 4 and 6 emotion classes. . . . .	101
6.1	Angular values from Whissell’s study and those estimated by the models.	122



7.1	Coefficient of determination ( $R^2$ ) of linear, quadratic, and cubic polynomial regression models fitted to the fifteen symmetric paths of emotion transitions into the activation-evaluation space. The emotion categories are denoted as <i>Neu</i> for Neutral, <i>Ang</i> for Anger, <i>Fru</i> for Frustration, <i>Hap</i> for Happiness, <i>Exc</i> for Excitement, and <i>Sad</i> for Sadness	141
B.1	Emotion Words from Whissell and Plutchik. List taken from [44]. The first two numerical values represent valence and activation of each emotion word that was found in the study by Whissell [221], and the fourth column represents the corresponding angular location in the activation-evaluation space that was found in the study by Plutchik [174]. The values of valence range from 1 (negative extreme) to 7 (positive extreme), and the values of activation range from 1 (very passive) to 7 (very active).	170
C.1	The numerical values demonstrating the direction of 3D movement of 17 marker points effected by varying the second PC between $\pm 3\sigma$ on the female mean upper face. $\mu$ denotes the mean face and $\sigma$ denotes the standard deviation. Considering the mean face as a reference, the values identify the upward movement of forehead and eyebrows marker points by varying PC2 from $-1\sigma$ to $-3\sigma$ and the downward movement of forehead and eyebrows marker points by varying PC2 from $+1\sigma$ to $+3\sigma$ .	173

# List of Figures

1.1	A simple illustration of four temporal phases (neutral, onset, apex, and offset back to neutral) of a facial expression of emotion. . . . .	16
2.1	A circumplex structure of emotions obtained on the basis of similarity measures. Picture taken from [174]. . . . .	34
2.2	A bipolar circumplex of emotions, commonly known as activation-evaluation space. Picture taken from [234]. . . . .	37
4.1	Example of FeelTrace display during a tracking session. Cursor colour changes from red/orange at the left hand end of the arc, to yellow beside the active/passive axis, to bright green on the negative/positive axis, to blue-green at the right hand end of the arc [43]. . . . .	69
4.2	Iconic representation of valence (top row), activation (middle row), and dominance (bottom row) by Self Assessment Manikins [127]. . . . .	70
4.3	Marker Layout used in recording for IEMOCAP dataset [25]. . . . .	74
4.4	The ANVIL annotation tool used for emotion evaluation of the utterances based on categorical and continuous attributes [25]. . . . .	75
4.5	The distribution of data for each emotion category, <i>Neu</i> : Neutral, <i>Dis</i> : Disgust, <i>Hap</i> : Happiness, <i>Sur</i> : Surpsie, <i>Sad</i> : Sadness, <i>Ang</i> : Anger, <i>Fea</i> : Fear, <i>Fru</i> : Frustration, <i>Exc</i> : Excitement, <i>oth</i> : Other [25]. . . . .	77

4.6	The 28 marker points in 3D used for emotion recognition and analysis.	78
4.7	An example of data layout in a continuous conversation between two actors (male and female). The term ‘overlap’ refers to the situation where both actors are talking at the same time. . . . .	80
5.1	The scree graph for the full face data plotting percentage variance covered by the first ten principal components. . . . .	85
5.2	The effect of varying the first PC (PC1) for the full face of female actor. $\mu$ denotes the mean face and $\sigma$ denotes the standard deviation. . . . .	86
5.3	The 17 marker points in 3D on the mean upper face of the female actor.	87
5.4	The 11 marker points in 3D on the mean lower face of the female actor.	87
5.5	The scree graphs plotting the percentage variance covered by first ten principal components for (a) the upper face data, and (b) the lower face data. . . . .	87
5.6	The proposed joint face model. . . . .	89
5.7	The effects of varying the second principal component (PC2) on the female full face model. $\mu$ denotes the mean face and $\sigma$ denotes the standard deviation. . . . .	91
5.8	The effects of varying the second principal component (PC2) on the female upper face. $\mu$ denotes the mean face and $\sigma$ denotes the standard deviation. The numerical values demonstrating the direction of variation of marker points are listed in Table C.1 in Appendix C. . . .	92
5.9	The effects of varying the second principal component (PC2) on the female lower face. $\mu$ denotes the mean face and $\sigma$ denotes the standard deviation. . . . .	93

5.10	Positions of images in the space of the first three principal components of some of the transformed emotion clusters of full face data in 4D space. All the figures are in 3D, but to get a clear view the viewpoint has been rotated by setting the azimuth and elevation to some suitable value. The figure needs coloured print to differentiate between different clusters. . . . .	99
5.11	The comparison of the mean accuracy of the four shape models, the rule-based classifier, and the SVM-based classifier on the female testset. Lines mark one standard deviation. . . . .	102
5.12	The comparison of the mean accuracy of the four shape models, the rule-based classifier, and the SVM-based classifier on the male testset. Lines mark one standard deviation. . . . .	103
5.13	The robustness of the joint face model and the SVM-based classifier to mislabelled training data. . . . .	104
5.14	Some examples of emotion blends. . . . .	106
6.1	Illustration of the resultant vector direction and length (in grey) obtained by vector addition of (a) $60^\circ, 180^\circ, 300^\circ$ , (b) $120^\circ, 180^\circ, 240^\circ$ , and (c) $150^\circ, 180^\circ, 210^\circ$ . All the angles start from the horizontal zero-direction. The light grey circle is the unit circle. In (a), we cannot see the resultant vector of length zero, it resides at the centre of the circle. This picture is taken from [19]. . . . .	111
6.2	The probability density plots for three different sets of datapoints modelled using wrapped normal distributions with selected parameters $\mu$ and $\sigma^2$ . The red circles show the density estimate based on the observed data. Picture taken from [87]. . . . .	113

6.3	The probability density plot for 100 datapoints randomly drawn from a wrapped Cauchy distribution. The red circle shows the density estimate based on the observed data. Picture taken from [87]. . . . .	114
6.4	The effect on the density of von Mises distribution for (a) fixed mean direction $\mu$ and varying concentration parameter $K$ and (b) fixed concentration parameter $K$ and varying mean direction $\mu$ . . . . .	118
6.5	The position of each basic emotion (based on the training set) in the activation-evaluation space. . . . .	121
6.6	(a) von Mises Probability Distributions in the mixture model with unit weight, (b) von Mises Probability Distributions characterising one frame of an utterance labelled as [angry, angry, frustrated] by three human experts. . . . .	123
6.7	The test of fit for (a) the mean ground truth <i>directions</i> and those estimated by the mixture model (b) the mean ground truth <i>intensities</i> and those estimated by the mixture model, for each of the seven conversations in the test set. Lines mark one standard deviation. . . .	124
6.8	Illustration of the Kuiper's statistic. Blue line is model estimated CDF of (a) anger, (b) frustration, (c) happiness, (d) excitement, and (e) sadness, red line is an empirical CDF of the first conversation in the testing set, and the green line is the Kuiper's statistic. . . . .	126
6.9	The mapping of continuous emotions during an utterance on the activation-evaluation space, along with the mean ground truth direction and intensity and those estimated by the model. The movement of emotions through time is represented by changing colour spectrum from dark/red (start) to light/yellow (end). . . . .	127

7.1	The mapping of continuous emotions during a conversation (Ses01F_impro01) into the activation-evaluation space. The movement of emotions through time is represented by changing colour spectrum from dark/red (start) to light/yellow (end). . . . .	132
7.2	Test of smoothness of emotion paths corresponding to each conversation in the testing set. For explanation, see the text. . . . .	134
7.3	A linear regression model fitted to $R/S$ analysis for all emotion trajectories in the testing set. . . . .	136
7.4	The size of the ‘change’ between two 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. . . . .	137
7.5	The false positives appear during transition from neutral to excited. The movement of emotions through time is represented by changing colour spectrum from dark/red (start) to light/yellow (end). . . . .	138
7.6	Transitions between negatively correlated emotions tend to pass through the neutral state, while transitions between positively correlated emotions do not. . . . .	139
7.7	Ses01F_impro01, continuous frames during emotion transition from anger to happiness. The temporal relationship between intensity change and angle change can be seen. For explanation, see the text. . . . .	143
7.8	Outputs of the emotion trajectories for the four different testing sets in the activation-evaluation space. . . . .	145

7.9	A synchronised frame-by-frame comparison of videos from the dataset and the corresponding mapped emotions into the activation-evaluation space. The snapshot shows the <i>laughing</i> expression during a happy conversation mapped as <i>excitement</i> of high intensity on the space. The female actor was being recorded. . . . .	146
7.10	The sarcastic expression lead to an abnormal transition towards happiness during an angry conversation. The female actor was being recorded.	147
8.1	Ses01F_impro01, continuous frames (6500-7100) during emotion transitions. Time-offset between intensity change and angle change. . . .	161
8.2	The temporal structure of emotional life, figure taken from [44]. . . .	162