Adaptive artificial brain for humanoid robot using Pattern Recognition and Machine learning.

Master Prince[†] and Suliman Alsuhibany[††]

[†]Department of Computer Science, Qasim University, College of Computer, Qasim University, Al Qasim, KSA
[††] Department of Computer Science, Qasim University, College of Computer, Qasim University, Al Qasim, KSA

Summary

Human face is a very special and it has been used to express emotions by the human being in order to express their feelings to others. But when talking about Human Robot Interaction (HRI), it becomes difficult to identify the exact emotion. So far, there are seven recognized emotions: Natural, Happy, Sad, Anger, Surprise, Fear, and Disgust. However, apart from these emotions, human being can express combination of these emotions. Therefore, recognition of the expression and what emotional expression should be given to the person as a response becomes more difficult. In this paper, voice features of the human has also been considered apart from emotions recognized through facial expression, in order to recognize the actual emotion with accurate intensity, and also to increase the effectiveness of the HRI. On the other side, the corresponding real time artificial facial expression is generated and LED pattern around the eyes of the robot has been used to make it more effective. Moreover, the transition between two consecutive emotions has been made smoother for producing humanlike emotion. To evaluate the effectiveness of the proposed method a questionnaire survey is conducted. Thus results show that artificial robotic head based on proposed method appropriately responds to users.

Key-words:
LED pattern, voice feature, emotional intensity, emotional expressions, artificial facial expression.

1. Introduction

Robotics becomes very emerging area in today's world, as it is playing very important role in various fields like medical science, military applications, home appliances, education etc. In recent years, one of the major attractions of the researchers is to develop intelligent robot that can interact with people like companion unlike machine. To interact with a humanoid robot, a study about HRI is very important. A humanoid robot must be able to understand the person's emotion state at a particular instance. The major advantages of this application are in crime control, psychiatric clinic and HRI. Researches still facing challenges in making robots behave like human. In this proposal, we introduce a framework to make a humanoid robot works more effective in order to understand the emotion of the person. The approach applies pattern recognition as well as machine learning.

As a human being, we can recognize emotion of ones by seeing his face, listing his voice and his walking style or gait. Through facial expression human's expression can be recognized effectively [1]. And research shows the color (LED pattern) around the eyes can also play important role in order to reveal emotional state [2]. The facial expressions keep most information about human's emotional state. So, if robots can automatically recognize the facial expressions, those artificial systems are easily able to understand or estimate human's emotion or mood. This recognition technique can be also used as a component of human-robot interaction (HRI). A sociable robot, i.e. Leonardo is presented in [3] that was able to express emotions close to human. Another reception robot, SAYA, has been developed to realize natural voice and natural interactive behaviors with six typical facial expressions by Hashimoto et al. [4], [5]. Walking style, or gait, can also reveal the walker's emotional state [6] [7]. As stated, voice is the common way we express emotions. Changes in properties like pitch, tempo and loudness of speech create emotional differences [8]. As the most effective way to recognize human's emotion is to notice human's facial expressions K.T Song et al., proposed a method based on image processing [9]. The willingness to interpret human gestures and facial expressions is making interaction with robots more human-like. So M. Wimmer et al., proposed in [10] a model–based approach for automatically recognizing facial expressions, and its application to human-robot interaction. In Japan, Waseda University [11] has developed a series of robots named WE-R (four in total) since 1996. One of them is SAYA developed by professor Hiroyasu [12]. In University of Kaiserslautern [13], German, a project was focusing on humanoid robot heads has been launched which has designed a very complex robot head that not only can mimic human facial expressions but also sensor the environment around it.

Previous efforts provide enough tools for designing emotional robots. However, it is recognized that all these representations lake fusion of these parameters i.e., image of face, voice and gait into an emotion recognition model. This fetched our attention to look for a model for emotion

recognition and imitate it to the humanoid robot that can produce human like emotion unlike machine.
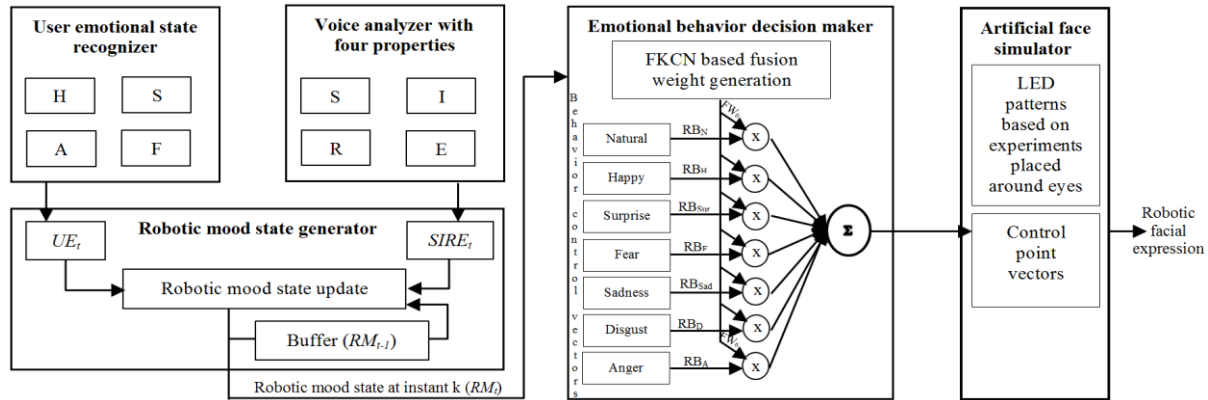

Fig. 1 Block diagram of ABERGS

On the other hand, the proposed method is intended to produce smoother transition between consecutive emotions and irrespective of all parameters new emotion should also depend on previous robot's emotion or mood of the robot. Once actual facial expression is recognized it can be imitated onto the artificial robot face simulator. One additional feature is added to the robot to make it more effective, by using LED pattern around the eyes [1].

The rest of the paper is structured as follows. Section 2 shows Adaptive Artificial Brain for Emotion Recognition and Generation. Section 3 explains Robotic Mood State Generator. The Emotional Behavior Decision Maker is presented in Section 4. Section 5 discusses Artificial Face Simulator

2. Adaptive Artificial Brain for Emotion Recognition and Generation

Fig. 1 shows block diagram of proposed artificial brain emotion recognition and generation system (ABERGS). The robotic facial expression as this is ultimate goal of the purposed system is meant to react to user's emotional state and at the same time reflect user's mood as well. And behavior of the robotic face should be effective enough at the same time emotion transition should be smoother like human. For this reason, we combine different (five) modules to construct the ABERS. A camera and microphone are placed in front of the robot. The received image is send to the processing unit to obtain the emotional state of the user at   an instant t (UE_t) is recognized [14] and represented in the form of vectors of four emotional intensities: happy($UE_{H,t}$), fear($UE_{F,t}$), angry ($UE_{A,t}$), and sad ($UE_{S,t}$) [15]-[18]. The intensity of each recognized emotional state at a particular instant is ranges between 0 to 1. In Fig. 1 N, S A, and F is representing

intensities value of natural, sad, angry, and fear respectively at an instant. Same as voice received via microphone is processed through Praat and four parameters are recorded speed, intensity, regularity, and extent ($SIRE_t$) [19], Fig. 1 represents these parameters as S, I, R and E respectively. All of these parameters for a particular instant t, passes to the robotic mood state generator and accordingly robotic mood state is updated, is represented as (RM_t).

As human face is very complex it is not easy to limit it under fixed number of classified emotions. In order to generate human like emotion a fuzzy Kohonen clustering network (FKCN) is applied to generate fusion weights (FW_e, $e= 0 \sim 6$) against each recognized user's emotional intestines (refer to Section IV for detailed description) [20]. So we are able to generate combination of seven basic emotions which

leads to the human like behavior. Emotional behavior decision maker generates summation of the product of each behavior control vectors and its corresponding fusion weight.

Artificial face simulator is actually a facial image with eight feature points extracted [21]. The movement of these feature points generates different emotions depending on fused emotional behavior i.e., control point vectors are described by combination of weighted basic facial expressions linearly. Using a model provided by Ekman [22], moving control points are determined and facial expressions are described accordingly. And LED pattern based on experiments are placed around eyes with fixed period of rise and fall time [1]. David O. Johnson et al. provide a table with LED patterns associated with each emotion. Finally, robotic facial expression is shown by the simulator.

3. Robotic Mood State Generator

Through the real life experience of emotional behavior it has been observed that human being's happy and sad intensities behave inversely, and in the same way anger and fear. Thus, it can be plotted on 2D plane. Based on psychologist study a relationship between mood and emotion is plotted on 2-dimentional space, the axes of which could be interpreted as pleasure–displeasure and arousal–sleepiness has been proposed [23].

$$(UE_t) = \begin{bmatrix} happy\ intensity\ at\ instant\ t \\ sad\ intensity\ at\ instant\ t \\ angry\ intensity\ at\ instant\ t \\ fear\ intensity\ at\ instant\ t \end{bmatrix} \quad (1)$$

$$= \begin{bmatrix} UE_{H,t} \\ UE_{F,t} \\ UE_{A,t} \\ UE_{S,t} \end{bmatrix} \quad \in 0 \sim 1 \quad (2)$$
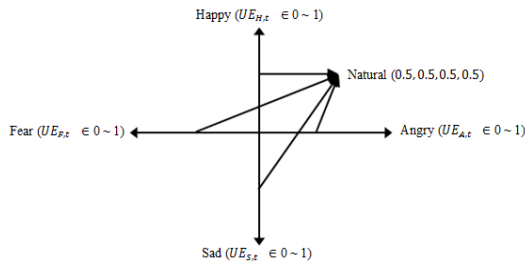

Fig. 2 A sample 2D plot of a user emotional state at an instant (UEt) recognized through image.

Fig. 2 shows, if the intensities of all four considered emotional behaviors (happy, sad, angry, and fear) are in the middle (0.5) of the range (0~1), represents natural emotion. For another instant, if the value of (UEH,t, UES,t, UEA,t, UEF,t) is (1, 0,0,0), represent excitement of the user.

In the same way the voice sample is analyzed based of four parameters speed, intensity, regularity, and extent and represented as SIREt. Table 1 and Table 2 describe these parameters and Feature distributions of the Berlin Emotional Database dataset respectively.

Table 1: Properties of voice and associated emotional features [19]

Parameter	Description	Voice Feature	Measurement
Speed	Slow vs. fast	Speech rate [21],	syllables/sec
Intensity	gradual vs. abrupt	voice onset rapidity [25],	dB/sample^2
Regularity	smooth vs. rough	jitter [25]	dB/sample
Extent	small vs. large	pitch range [21],	Hz

Table 2: Feature distributions of the Berlin Emotional Database dataset

Feature	μ	σ
Speech rate (syll/sec)	5.87	1.26
Voice onset rapidity (dB/sample2)	11.36	4.56
Jitter (dB/sample)	871.91	269.39
Pitch range (Hz)	111.57	44.04

Once we analyze different emotional voice especially happy, angry, sad, and fear, it is noticed that for each of these emotional voice one of the voice feature is more affected. That is happy voice reflect speed rate (speed) on an average by more than one standard deviation, angry voice reflect voice onset (intensity) on an average by more than one standard deviation, sad voice reflect jitter (regularity) on an average by more than one standard deviation and fear voice reflect pitch range (extent) on an average by more than one standard deviation.
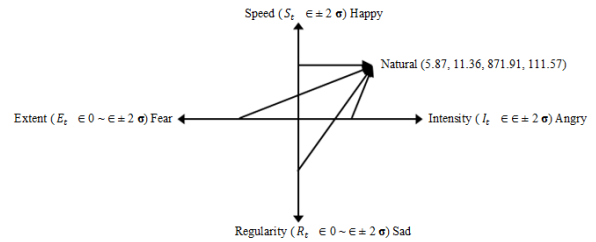

Fig. 3 A sample 2D plot of a user emotional state at an instant (SIREt) recognized through voice.

Fig. 3 shows, if all the feature values are at the mean value so it represents natural voice. And if speed rate varies with one or two standard deviation, represents happy emotion.
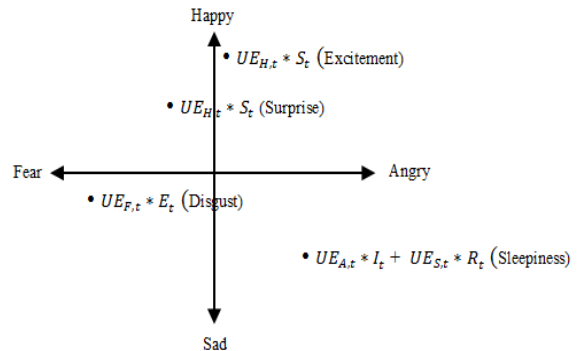

Fig. 4 2-D scaling for facial expressions based on image and voice fusion.

Fig. 4 show, mapping of these two scaling on the same plane, ten facial expressions are observed and portion of them found on the space. Like for excitement fusion weight is calculated as $(UE_{H,t} * S_t + UE_{A,t} * I_t + UE_{S,t} * R_t + UE_{F,t} * E_t)$. But in the Fig. 4 only happy (image) and

speed (voice) is shown just because of only these two parameters have highest influence on excitement.

$$(SIRE_t) = \begin{bmatrix} value\ of\ speed\ at\ instant & t \\ value\ of\ intensity\ at\ instant & t \\ value\ of\ regularity\ at\ instant & t \\ value\ of\ extent\ at\ instant & t \end{bmatrix} \quad (3)$$

$$= \begin{bmatrix} S_t, \in \pm 2\,\sigma \\ I_t, \in \pm 2\,\sigma \\ R_t, \in \pm 2\,\sigma \\ E_t,, \in \pm 2\,\sigma \end{bmatrix} \quad (4)$$

So, the robotic mood state at instant t can be represent as below, before multiply value of voice features must be normalized between 0 ~ 1. For an example speed range was linearly scaled between a minimum f0 of 3.35 syll/sec (-2 σ) and a maximum f0 of 8.39 syll/sec (+2 σ). So if speed is 3.35 or less it is considered as 0 and if 8.39 or more considered as 1.

$$RM_t = RM_{t-1} + (UE_{H,t} * S_t + UE_{A,t} * I_t + UE_{S,t} * R_t + UE_{F,t} * E_t) \quad (5)$$

Because robotic mood will also depend on current mood state (RM_{t-1}) as well. And the value of RM_t passed to the emotional behavior decision maker for further processing.

4. Emotional Behavior Decision Maker

Here in this section we generate control point vectors, FKCN is employed to determine the fusion weight of the emotion corresponding to each basic recognized emotions.
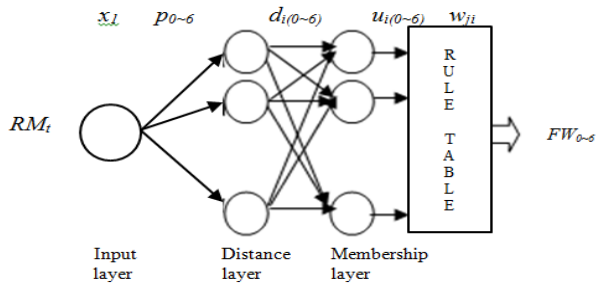

Fig. 5 Feature Fusion Weight Generation

In Fig. 5 the input layer receives current robotic mood state (RMt) form mood state generator and provide to the next layer; distance layer calculates the difference of current robotic mood from each standard patterns [24].

$$d_{ij} = \| x_i - p_j \|^2 \quad (6)$$

where xi represents the input pattern and pj represents the jth standard pattern, in our case total number of prototype patterns are seven. The another layer of the network

determine the degree of witch the input pattern and standard pattern are similar through mapping the distance dij to membership values uij. This degree of similarity is expressed by a membership value u_{ij} (0 to 1), As membership values renges between 0 – 1, sum of all these membership value must be equal to 1.

Now current fusion weight against each prototype pattern (FWi, i =0 ~ 6) is obtained as follow;

$$FW_i = \sum_{j=0}^{c-1} w_{ji}\, u_{ij} \quad (7)$$

where wji represents the standard-pattern weight of the ith output behavior determined through the rule table as shown in Table 3. This rule table is formed after analysis of ten different expressions and located on the 2D plane of Fig. 4.

Once fusion weight is determined summation of the product of fusion weights and prototype pattern weights is calculated as an artificial facial expression and passed to the artificial face simulator.

5. Artificial Face Simulator

Once the edge has been detected (canny edge detection method is applied to do so), features have been extracted through fusion of grid and global features [21]. The areas like fore head, lip region, skin color, eye tail, moustache region, nose wing and eyelid of the face are used as grid feature. On the other hand the distences like; interocular, between nose tip to the line joining two eyes, between lips to the nose tip, between lips to the line joining two eyes, width of lips, and eccentricity of the face, ratio of dimension included under global features. So eight feature points are selected as shown in fig. 6.
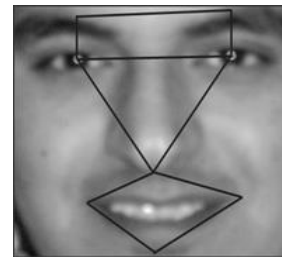

Fig. 6 Eight feature points extracted.

As human being uses facial muscles to express their emotions, artificial face simulator uses the control point vectors to move these feature points accordingly. And LED patterns are place around eyes in order to enhance the effectiveness [2].

6. Experiment

In order to evaluate the performance, the experiments are carried out by observation of fifty numbers of people. They interact with the face simulator and the robotic reflections are observed. The satisfaction level is noted between 0~1. The percentage satisfaction level is shown in the Table 4.

Table 4: satisfaction level with and without voice

Emotions	Satisfaction Level (without voice)	Satisfaction Level (With voice fusion)
Natural	96.5%	98%
Happiness	97%	98%
Surprise	96.5%	97.5%
Fear	97.5%	99%
Sadness	96%	98%
Disgust	97.5%	98.5%
Anger	95.0%	98%

7. Conclusion and Futre Work

A method of ABERGS has been developed. The proposed method used both image and voice to recognize and generate robotic mood state. Furthermore, robotic mood state transition using FKCN together with rule table provides satisfactory combination capacity for a robot to create enthusiastic collaborations. A survey has been conducted and its results show that the humanoid robotic face entertains humans like a companion rather than machine. In future work human gait can also be introduced in order to recognize human behavior more appropriately. For future work, we will first try to extract more feature points on the face to mimic artificial emotion and human gait can also be considered in order to recognize emotion of the human being. Second, we will focus more on processing speed by using GPU with efficient algorithm.

References

[1] Do Hyoung Kim, Sung U k Jungt, Kwang Ho An, Hui Sung Lee and Myung Jin Chung, Development of a Facial Expression Imitation System, Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems October 9 - 15, 2006, Beijing, China.

[2] David O. Johnson· Raymond H. Cuijpers· David van der Pol, "Imitating Human Emotions with Artificial Facial Expressions," Int J Soc Robot (2013) 5:503–513 DOI 10.1007/s12369-013-0211-1.

[3] C. Breazeal, D. Buchsbaum, J. Gray, D. Gatenby, and B. Blumberg, "Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots," J. Artif. Life, vol. 11, no. 1/2, pp. 1–32, Jan. 2005.

[4] T. Hashimoto, S. Hiramatsu, T. Tsuji, and H. Kobayashi,, "Realization and evaluation of realistic nod with receptionist

robot SAYA," in Proc. IEEE 16th Int. Symp. RO-MAN Interactive Commun., Jeju Island, Korea, 2007, pp. 326–331.

[5] T. Hashimoto, S. Hiramatsu, T. Tsuji,and H. Kobayash I, "Development of the face robot SAYA for rich facial expressions," in Proc. Int. Joint Conf. SICE-ICASE, Busan, Korea, 2006, pp. 5423–5428.

[6] Roether, C.L., Omlor, L., Christensen, A., Giese, M.A.: Critical features for the perception of emotion from gait. J. Vision 9(6), 15, 1–32 (2009)

[7] Montepare, J.M., Goldstein, S.B.: The identification of emotions from gait information. J. Nonverbal Behav. 11(1), 33–42 (1987)

[8] Scherer, K.H.: Vocal affect expression: A review and a model for future research. Psychol. Bull. 99, 143–165 (1986).

[9] K. T. Song, M. J. Han, and J. W. Hong, "Online learning design of an image-based facial expression recognition system," Intell. Service Robot., vol. 3, no. 3, pp. 151–162, Jul. 2010.

[10] Matthias Wimmer, Bruce A. MacDonald, Dinuka Jayamuni, and Arpit Yadav, Facial Expression Recognition for Human-Robot Interaction – A Prototype, G. Sommer and R. Klette (Eds.): Rob Vis 2008, LNCS 4931, pp. 139–152, 2008.Springer-Verlag Berlin Heidelberg 2008.

[11] Hiroyasu Miwa, TomohikoUmetsu, Atsuo Takanishi, et.al. Human like robot head that has olfactory sensation and facial color expression. Proceeding of the 2001 IEEE, ICRA,2001: 459 - 464.

[12] Kobayashi H, Ichikawa Y, Senda M, et al.Realization of realistic and rich facial expressions by face robot. Proceedings of the 2003 IEEE International Conference on Intelligent Robots and Systems Las Vegas, Nevada, USA, Oct 27－23, 2003: 1123 - 1128.

[13] Karsten Berns, Jochen Hirth, "Control of facial expressions of the humanoid robot head ROMAN" Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems October 9-15, 2006, 3119-3115.

[14] K. T. Song, M. J. Han, and J. W. Hong, "Online learning design of an image-based facial expression recognition system," Intell. Service Robot., vol. 3, no. 3, pp. 151–162, Jul. 2010

[15] M. A. Amin and H. Yan, "Expression intensity measurement from facial images by self-organizing maps," in Proc. Int. Conf. Mach. Learn. Cybern., Kunming, China, 2008, pp. 3490–3496.

[16] M. Beszedes and P. Culverhouse, "Comparison of human and automatic facial emotions and emotion intensity levels recognition," in Proc. Int. Symp. Image Signal Process. Anal., Istanbul, Turkey, 2007, pp. 429–434.

[17] M. Oda and K. Isono, "Effects of time function and expression speed on the intensity and realism of facial expressions," in Proc. IEEE Int. Conf. Syst., Man Cybern., Singapore, 2008, pp. 1103–1109.

[18] K. K. Lee and Y. Xu, "Real-time estimation of facial expression intensity," in Proc. IEEE Int. Conf. Robot. Autom., Taipei, Taiwan, 2003, pp. 2567–2572.

[19] Angelica Lim, Tetsuya Ogata, and Hiroshi G. Okuno, "Converting emotional voice to motion for robot telepresence", 2011 11th IEEE-RAS International Conference on Humanoid Robots Bled, Slovenia, October 26-28, 2011.

[20] K. T. Song and J. Y. Lin, "Behavior fusion of robot navigation using a fuzzy neural network," in Proc. IEEE Int. Conf. Syst., Man Cybern., Taipei, Taiwan, 2006, pp. 4910–4915.

[21] M. Prince "Enhancing Face Normalization based on Novel Normal facial Diagram", ISBN: CAINE 2013: 978–1–880843–93–226th International Conference on Computer Applications in Industry and Engineering, September 25-27, 2013, Omni Hotel, Los Angeles, California, USA,

[22] Grimace Project. [Online]. Available: http://grimace-project.net/

[23] J. A. Russell and M. Bullock, "Multidimensional scaling of emotional facial expressions: Similarity from preschoolers to adults," J. Personality Social Psychol., vol. 48, no. 5, pp. 1290–1298, May 1985.

[24] M. Prince "Clustering- Based Spam Image Filtering Considering Fuzziness of the Spam Image" U.S ISSN: 2156-5570 (Online) International Journal of Advanced Computer Science and Applications (IJACSA), Volume 7 Issue 12 December 2016 pp-269-270.

University. Awarded Ph.D. by Pune University, India. Head of Imaging and Robotics Research Groups. My area of research interest is Digital Image Processing, Pattern recognition. Joined Qasim University as Assistant Professor in 2009. Teaching B.S and M.S program.

Master Prince received the B.S. degree from Patna University and M.S. degrees in computer Science from Inidra Gandhi Open

Suliman Alsuhibany, PhD, is an assistant professor in the Computer Science department and the head of the department at Qassim University, Saudi Arabia. He received his PhD in information security from Newcaslte University, UK, and MSc in computer security and resilience from Newcastle University, UK
.

Table 3 RULE TABLE FOR INTERACTIVE EMOTIONAL BEHAVIOR DECISION

If part (prototype patterns)				THEN part weighting						
$UE_{H,t} * S_t$	$UE_{A,t} * I_t$	$UE_{S,t} * R_t$	$UE_{F,t} * E_t$	Natural	Happiness	Surprise	Fear	Sadness	Disgust	Anger
0.51	0.52	0.58	0.49	1						
•		•			•				•	
•		•			•				•	
0.12	0.81	0.23	0.08							1