A Dynamic Model for Identification of Emotional Expressions

Rafael A.M. Gonçalves, Diego R. Cueva, Marcos R. Pereira-Barretto, and Fabio G. Cozman

***Abstract*—This paper discusses the dynamics of emotion recognition on faces, layering basic capabilities of an emotion sensor. It also introduces a model for the recognition of the overall conveyed emotion during a human-machine interaction, based on the emotional trajectory over an emotional surface.**

***Index Terms*—Emotion dynamics, emotion recognition, emotional surface, Kalman filtering.**

I. Introduction

PERSON-to-person communication is highly non-verbal: face, body and prosody demonstrate much of what is not being said but loudly spoken. Therefore, it is expected that human-machine communication may benefit from non-verbal expressions. This may have already been started as the so-called "user centric experience", by having applications and games with voice and gesture recognition, for instance. But recognizing emotions is not easy not even for humans: it has been shown humans correctly recognize the conveyed emotion in the voice 60% of the cases and 70% to 98% on the face [1], [2]. This paper focuses on emotion recognition on faces.

In the 70´s, Ekman and co-workers established FACS (Facial Action Coding Sys-tem), a seminal work for emotion recognition on faces [3], by decomposing the face into AUs (Action Units) and assembling them together to characterize an emotion. The universality of AUs was strongly debated for the last two decades but inter-cultural studies and experiences with pre-literate populations lead to its acceptance [2]. A state-of-the-art review of emotion detection on faces can be found in [4]. Among the most recent works, we cite eMotion, developed at Universiteit van Amsterdam [5] and FaceDetect, by Fraunhofer Institute [6].

Both eMotion and FaceDetect detect an emotion from each frame on a video (or a small sequence of frames). Therefore, they show excelent results in posed, semi-static situations. But during a conversation, the face is distorted to speak in many ways, leading these softwares to incorrectly detecting the conveyed emotion. Even more, a movement of the mouth during a conversation, similar to a smile, does not mean the speaker is happy; it may be an instantaneous emotion: the speaker saw something not related to the conversation which made him smile. Consider, as an example, the frames from a video, shown in Figure 1 and the outputs from eMotion, on Figure 2.

From the frames on Figure 1, a human would conclude nothing. From eMotion outputs on Figure 2, a human would conclude nothing, also. Or perhaps for Sadness, which seems to display a higher mean value. But by seeing the video, even without sound, a human would easily conclude for Anger.

Therefore, during a conversation, there is a "slow dynamic" related to the overall emotion conveyed, lasting longer than a single video frame. During a conversation, many "emotional modes" (as vibrational modes in Mechanics) may be displayed, invoked by events (internal or external) to the speaker but probably out of the reach for the listener. These modes are interleaved within the conversation, somewhat as it happens with appositive phrases [7]. This work discusses a general model for the detection of emotional modes and presents a model to detect slow dynamic emotions. Some reference material is presented on Section 2, while Section 3 presents the general model and Sections 4 to the end present the proposed model for the detection of emotional modes.

II. Reference Material

Behaviorist theories dominated the Psychology scene from the 30´s to 60´s. According to them, emotions are only a dimension of human behavior, corresponding to a certain degree of energy or activity. The determinist characteristic and one-dimensional associations of event-emotion are on the basis of these theories: for each event, an associated emotion. Appraisal Theories took their place during the 80´s, although started during the 60´s. Simply put, they postulate emotions are elicited from appraisals [8]. Appraisals differ from person to person but the appraisal processes are the same for all persons. Therefore, they offer a model which justifies a common behavior but, at the same time, allows for individual differences.

This work is based on the concept of emotions which, on Scherer (2001) words, are "... an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism".

The appraisal process starts with an event. We argue the perceived emotion should be considered as an event, as much as a strong noise such as an explosion. Therefore, it will be evaluated for its relevance, according to (i) novelty, (ii) intrinsic pleasantness; (iii) Goal/need relevance.

Manuscript received June 10, 2011. Manuscript accepted for publication September 14, 2011.

This work is supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), FAPESP, under research project 2008/03995-5 and the University of São Paulo.

Rafael A.M. Gonçalves, Diego R. Cueva, Prof. Dr. Marcos R. Pereira-Barretto and Prof. Dr. Fabio G. Cozman are with the Decision Making Lab of the Mechatronics Department, University of São Paulo, São Paulo, Brasil. Electronic correspondence regarding this article should be sent to Prof. Dr. Marcos R. Pereira-Barretto to mpbarre@usp.br.

Fig 1. From left to right, eMotion classified the frames as happiness (100%), sadness (70%), fear (83%) and anger (76%).

A specialized sensor such as the one proposed here is need to detect this kind of event. Going further, the perception of an emotion is altered by Attention and conditioned by Memory and Motivation; an emotion sensor should be adaptable as the eyes and ears.

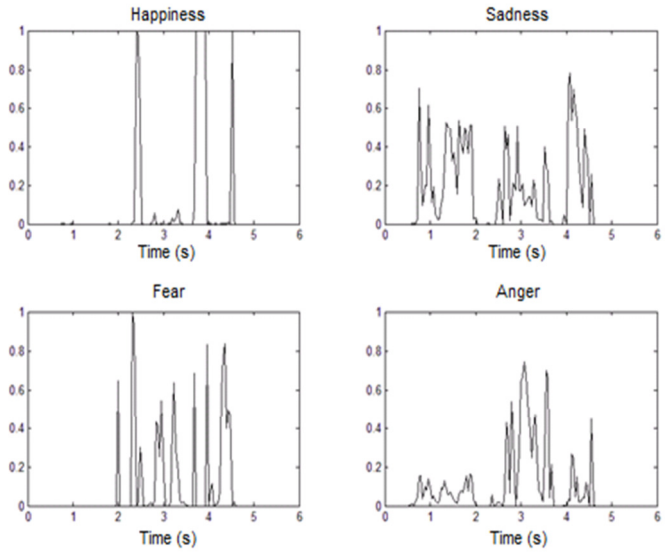

Fig 2. eMotion output for the video of Fig 1.

III. Reference Model

Emotion detection from video frames has been subject to research by many authors, as described on [4], [5] and [6]. Despite the advances, it is still an open subject; re-search on compensation of lightning, speech and body movements are some examples. These works are "raw sensors", on Figure 3. On the top of these sensors, we argue for the need of "emotional mode" detectors, for fast and slow dynamics. Con-sider, for instance, a conversation with a friend: the overall conveyed emotion could be Happiness, the slow dynamics. But suddenly the speaker reminds of someone he hates: Anger may be displayed. The event could be external: the speaker may see someone doing something wrong and also to display Anger. In both cases, Anger is displayed as the fast dynamics, enduring for more than just a few video frames. For the listener, the appraisal process could lead to just continue the conversation, ignoring Anger. Or change the subject to investigate what caused this change in speaker´s face.

IV. Proposed Model for Detection of Emotional Modes

The proposed model to determine the perceived emotion from instantaneous facial expressions is based on the displacement of a particle over a surface, subject to ve-locity changes proportional to the current probability of each emotion, at every mo-ment. This surface will be called here Dynamic Emotional Surface (DES). Over the surface, attractors corresponding to each detectable emotion are placed.

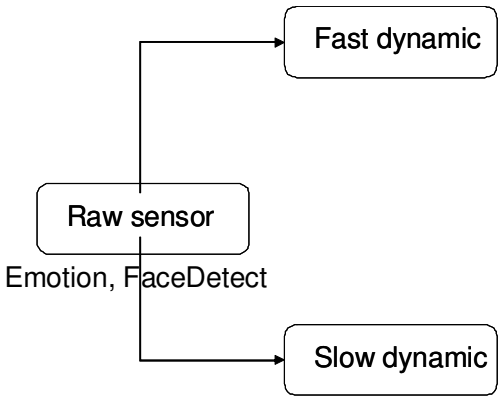

Fig 3. Reference model.

The particle, which represents the instantaneous emotion, moves freely over the DES, subject to attraction to each detectable emotion. It also moves toward the neutral state, placed at the origin of the coordinate system, the point of minimum energy. Therefore, the particle velocity can be determined by Eq. 1.

$$\vec{V}_p = \vec{V}_e + \sum_{a=1}^{N} \vec{V}_a \quad (1)$$

where:

$\vec{V}_p$: particle velocity

$\vec{V}_e$: initial velocity

$\vec{V}_a$: velocity in direction of each attractor or detectable emotion.

The idea, an emotional surface, comes from works such as [10], [11], [12], [13], [14], shown in Figure 4. This surface represents the appraised emotion while DES represents the

perceived emotion; they keep some relationship because the speaker should display a "reasonable" behavior.

DES keeps also some relationship with the Arousal-Valence plane [15], but differs for the same reasons as from Zeeman´s surface.

As an example, suppose the following paraboloid is a DES with the attractors listed in Table I:

$$z = f(x,y) = ax^2 + by^2 \quad (2)$$

$$\gamma(x,y) = (x, y, ax^2 + by^2) \quad (3)$$

$$a = b = 0{,}6 \quad (4)$$

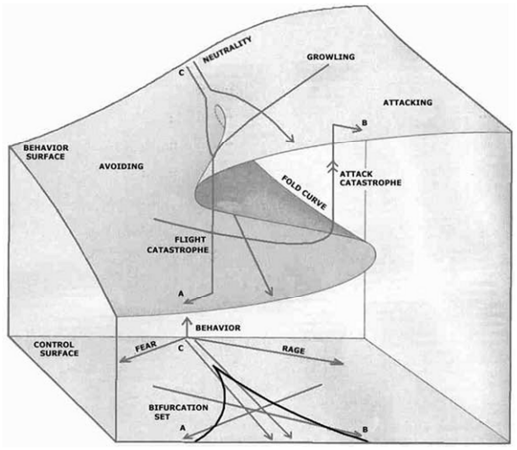

Fig 4. Zeeman´s Emotional Surface [10]

TABLE I
ATTRACTOR PROJECTIONS

Emotion	Attractor Projection
Happiness	$[60, 60, 0]$
Anger	$[-60, 60, 0]$
Sadness	$[-60, -60, 0]$
Fear	$[60, -60, 0]$

This example DES helps in highlighting the differences to the A-V Plane: Fear has been placed on 4th quadrant, while on the A-V Plane is positioned on the 3rd, close to the origin. But this positioning follows Zeeman´s surface, where Fear and Anger are orthogonal.

The velocity in direction of each attractor, $\vec{V}_a$, is proportional to the probability of each emotion as detected by existing software such as eMotion. Considering as $\vec{P}$ the current particle position and $\vec{A}$ the position of the attractor (emotion) being calculated, V_a can be calculated as:

$$\overrightarrow{AP} = \vec{A} - \vec{P} = [a_{px}, a_{py}, a_{pz}] \quad (5)$$

$$S(x) = \gamma(x, rx) \quad (6)$$

$$r = \left|\frac{a_{py}}{a_{px}}\right|, a_{px} \neq 0 \quad (7)$$

$$scale = \left|\frac{dS(x)}{dx}\right| = \sqrt{1 + r^2 + [2(a + br^2) * P_x]^2} \quad (8)$$

$$V_{x,a} = \frac{dS}{dt} * \frac{signal(a_{px})}{scale} * \frac{dS}{dx} \cdot \vec{\imath} \quad (9)$$

$$V_{y,a} = \frac{dS}{dt} * \frac{signal(a_{py})}{scale} * \frac{dS}{dx} \cdot \vec{\jmath} \quad (10)$$

Sensor input is always noisy; that is the case for eMotion also: in this case, noise comes from the frame-by-frame emotion detection. A pre-filtering can be applied to its output prior to submit to the model. Both Kalman filtering and moving-average filtering were tested, as shown in what follows.

V. EXPERIMENTS

Experiments were conducted to test the proposed model for the detection of slow dynamic.

Videos from eNTERFACE'05 Audio-Visual Emotion Database corpus were selected from those available displaying the emotions under study: 7 showing Fear, 9 showing Anger, 5 showing Happiness, 9 showing Sadness and 4 for Neutral. For each video, the facial mesh was adjusted on eMotion and its output collected.

A Kalman filter with process function in the form of Eq. 9 was adjusted, using 4 eMotion outputs for each emotion, given the parameters shown in Table II.

$$\frac{Y(s)}{R(s)} = \frac{K_k}{\tau s + 1} \quad (11)$$

TABLE II
PARAMETERS OF KALMAN FILTER

	Q	R	K_k	τ
Happiness	0.1	0.080	5	1.5
Anger	0.1	0.100	5	1.5
Sadness	0.1	0.035	5	1.5
Fear	0.1	0.010	5	1.5

TABLE III
COMPARISON BETWEEN UNFILTERED SIGNALS, MOVING AVERAGE AND PROPOSED KALMAN FILTERING WITH DES

	Original		Moving Average		Kalman	
Emotion	μ	σ	μ	σ	μ	σ
Happiness	0.175	0.634	0.175	0.237	0.114	0.127
Sadness	0.377	0.532	0.377	0.254	0.207	0.108
Fear	0.211	0.544	0.211	0.206	0.234	0.203
Anger	0.236	0.434	0.236	0.257	0.445	0.434

Applying this filter and moving-average filtering to the video whose frames are displayed on Figure 1 gave the results shown on Figure 5.

For both implementations, mean value and standard deviation were calculated as shown in Table III.

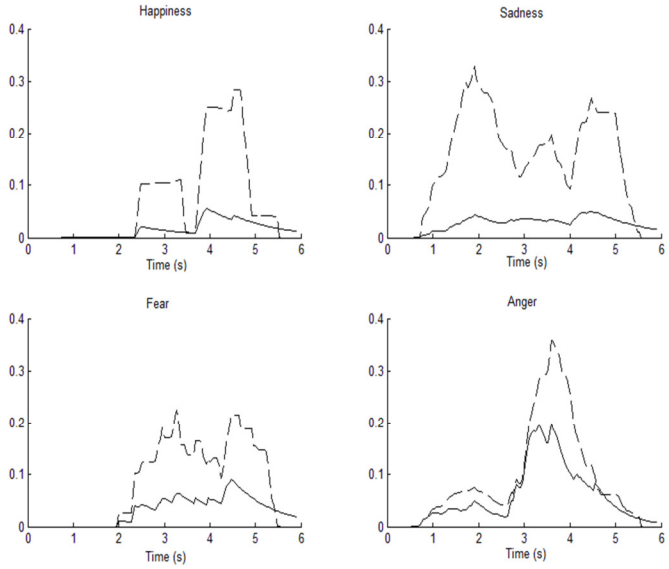

Fig 5. Moving Average (dashed) and Proposed Kalman Filtering with DES algorithm (solid) outputs for example video (see Fig 1).

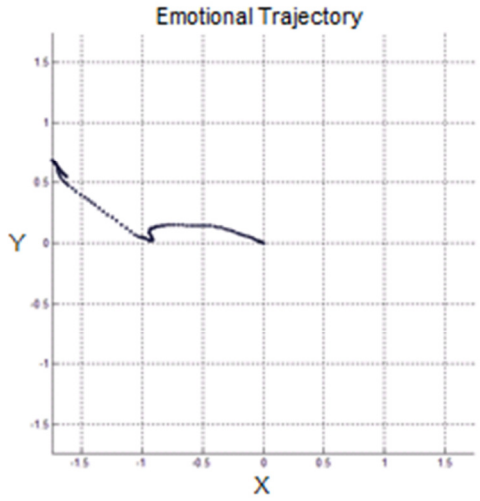

Fig 6. Projection of the Emotional Trajectory of the Sample Video.

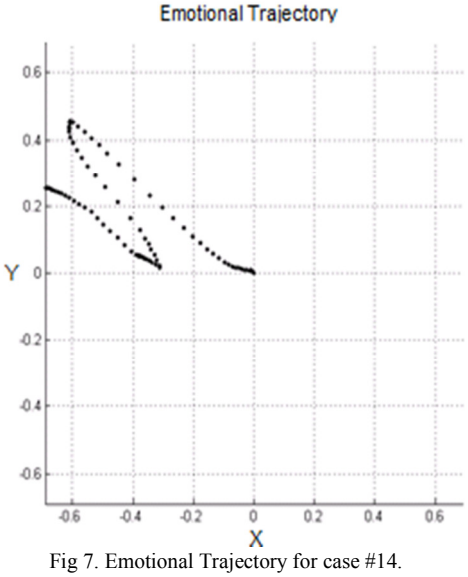

Fig 7. Emotional Trajectory for case #14.

The 14 remaining videos, i.e., those not used for adjusting Kalman filter, were then submitted to the system, yielding to the results shown in Table IV.

As it can be seen, the overall emotion conveyed by the video, Anger, has been correctly detected with Kalman filtering, although with a large std deviation. The projection on X-Y plane of the trajectory over the DES is shown in Figure 6.

As it can be seen, Anger is present almost all the time, starting mild but going stronger as the video continues. This corresponds to the human observation of the video.

TABLE IV
COMPARISON BETWEEN HUMAN EVALUATION AND THE PROPOSED KALMAN FILTERING WITH DES ALGORITHM

#	File	Classifications	
		Human	System
1	S1sa1	Sadness	Sadness
2	S38an1	Anger	Anger
3	S38fe3	Fear	Fear
4	S42sa1	Sadness	Sadness
5	S43ha1	Happiness	Happiness
6	S43an2	Anger	Anger
7	S43an3	Anger	Anger
8	S43an4	Anger	Anger
9	S43fe2	Fear	Fear
10	S42fe1	Fear	Fear
11	S43sa1	Sadness	Sadness
12	S43sa3	Sadness	Sadness
13	S43sa4	Sadness	Sadness
14	S43sa5	Sadness	Anger

As shown before for the sample video (see fig. 6), we may plot the emotional trajectory estimated for S43sa05 (#14):

Note the special case of video #14, showing Anger/Sadness swings, shown in Figure 7, which corresponds to author´s analysis.

VI. CONCLUSION

A reference model for recognition of emotions on faces has been introduced, besides a computational model for computing fast and slow conveyed emotions. The model has been tested for the detection of slow emotions, demonstrating good results.

As future works, the authors plan to test the model for fast emotions. The main obstacle foreseen is the lack of a corpus for this kind of test. The authors also plan to investigate the interaction between the proposed model and CPM processes.

REFERENCES

[1] R.W. Piccard, *Affective Computing*, MIT Press, 1997.
[2] P. Ekman and W.V. Friesen, *Unmasking the face*, Malor Books, 2003.

[3] P. Ekman and W.V. Friesen, *Facial Action Coding System: a technique for the measurement of facial movement*, Consulting Psychologists Press, 1978.

[4] M. Pantic and L.J.M. Rothkrantz, "Automatic analysis of facial expressions: state of art," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol. 22 no. 12, 2000.

[5] A. Azcarate, F. Hageloh, K. Sande, and R. Valenti, *Automatic facial emotion recognition*, Universiteit van Amsterdam, 2005.

[6] Fraunhofer Facedetect.
http://www.iis.fraunhofer.de/en/bf/bv/ks/gpe/demo
Accessed 07/04/2001.

[7] R. Cooper, *Quantification and Syntactic Theory*, D. Reidel Publishing Company; 1983.

[8] I.J. Roseman and C.A. Smith, "Appraisal Theory - Overview, Assumptions, Varieties, Controversies," *Appraisal Processes in Emotion - Theory, Methods, Research,* edited by K. Scherer, A. Schorr, and T. Johnstone, Oxford University Press, 2001.

[9] R.R. Scherer, "Appraisal considered as a process of multilevel sequential checking," *Appraisal Processes in Emotion - Theory, Methods, Research,* edited by K. Scherer, A. Schorr, and T. Johnstone, Oxford University Press, 2001.

[10] E.C. Zeeman, "Catastrophe theory," *Scientific American,* vol.4 no.254 pages 65−83, 1976.

[11] I.N. Stewart and P.L. Peregoy, "Catastrophe theory modeling in Psychology," *Psychological Bulletin,* vol. 94, no. 2, pp. 336-362, 1983.

[12] K. Scherer, "Emotions as episodes of subsystem synchronization driven by nonlinear appraisal processes," *Dynamic Systems Approaches to Emotional Development,* ed. by M.D. Lewis and I. Granic, Cambridge Press, 2000.

[13] H.L.J. van der Maas and P.C.M. Molenaar, "Stagewise cognitive development: an application of Catastrophe Theory," *Psychological Review,* vol.99, no.2, pp. 395−417, 1992.

[14] D. Sander, D. Grandjean, and K.R. Scherer, "A systems approach to appraisal mechanisms in emotion," *Neural Networks,* vol.18, pp. 317−352, 2005.

[15] H. Gunes and M. Piccardi, "Observer Annotation of Affective Display and Evaluation of Expressivity: Face vs. Face-and-Body," in *HCSNet Workshop on the Use of Vision in HCI (VisHCI 2006)*, Canberra, Australia, 2006.