Presented at the *1st Workshop of the EC HUMAINE Network of Excellence on Research on Emotions and Human-Machine Interaction*, (Geneva, Switzerland, June 2004).

MAUI: a Multimodal Affective User Interface Sensing User's Emotions based on Appraisal Theory - Questions about Facial Expressions…

Christine Lisetti, Guillaume Bastard
Institut Eurécom, Multimedia Communications Department, Sophia-Antipolis, France,
{lisetti, bastard}@eurecom.fr

I. Introducing MAUI

We are developing a Multimodal Affective User Interface (MAUI) framework shown in Figure 1 and described in [5], aimed at recognizing its users emotions by sensing their various user-centered modalities (or modes), and at giving the users context-aware feedback via an intelligent affective agent by using different agent-centered modes. The agent is built on an *adaptive system architecture* which takes in input three systems (**V, K, A**): **V**isual, **K**inesthetic and **A**uditory helped by subjective experience collected via **L**inguistic tools (**L**). The feedback is then rendered by a multimodal anthropomorphic interface agent that *adapts* to the user emotional state [3,4].

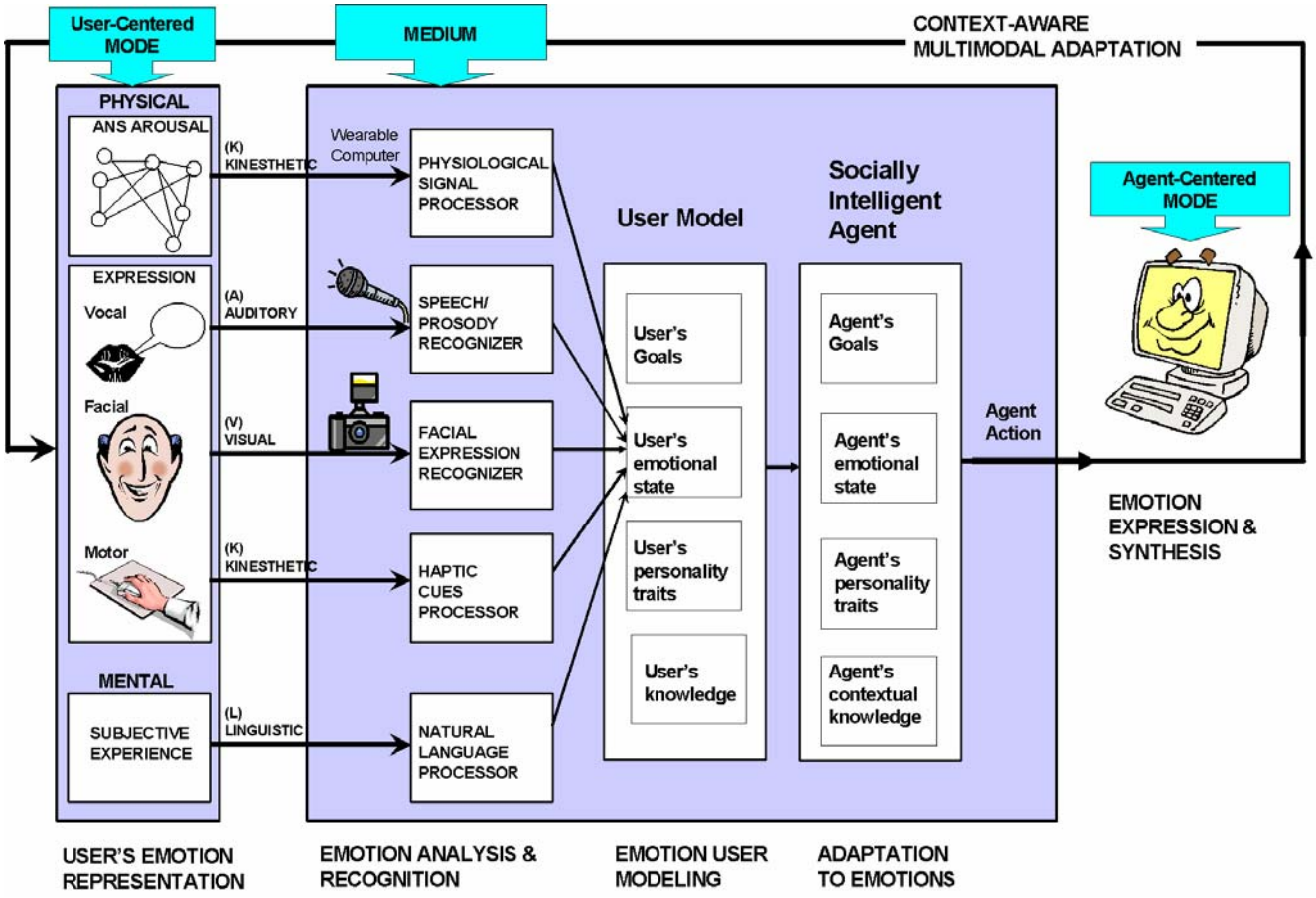

Figure 1: The MAUI framework

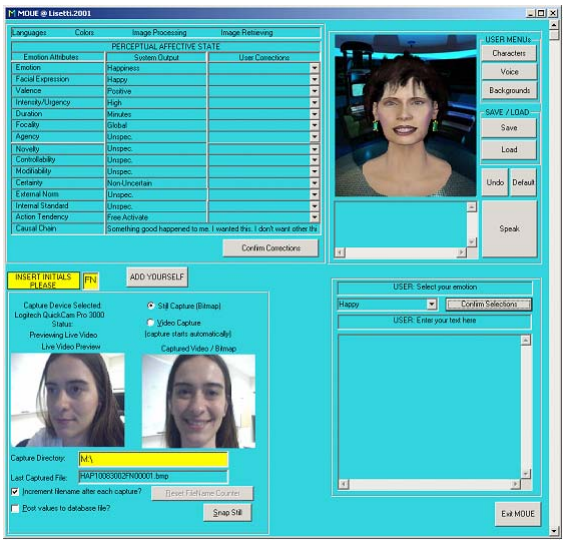

Figure 2: The current graphical interface of MAUI, our working tool

The current interface of the MAUI, shown in Figure 2 and fully described in [4], is made up of four parts:

1. A multimodal perception-sensing system:

- **Wearable computers** are used to capture **physiological signals (body temperature, GSR, heart rate)** and are processed by machine learning algorithms to categorize them in terms of signals associated with emotions.
- **Cameras** are used to capture **ongoing videos** and **captured still images** of the user during human-computer interaction and are displayed in the bottom left part of the interface. The captured images are taken regularly and used for automatic facial recognition of the user. We are also currently using these visual inputs to develop facial expression recognition algorithms based on Scherer's appraisal theory [6;7], for which have numerous questions, some of which are listed in Section II below.
- The user is also able to **input text** in the right bottom part and describe her/his current state to the system, or input an emotion label.

2. Adaptive Multimodal Feedback is displayed in the upper part of the interface.

- On the left, a **list of parameters** (detailed in Table 1 inspired from Scherer's [6;7]) describing the assessment of the user state by the system
- Emotion components are also used by a collection of multi-ethnic **avatars** (on the upper right) chosen by the user, able to adapt his/her emotions and facial expressions according to the user emotional state (e.g. mirrors the user's emotions, or adapt appropriately).
- We are also interested in refining our avatar's expressions and displays with some concepts described in [2].

The detailed list of the feedback parameters closely related to Scherer's SECs [6] is shown in Table 1:

Emotion	used to label the user's emotion assessed by the system
Facial expression	stores the facial expression label associated with the emotion.
Valence	Positive or negative, describes the pleasantness of an affective state. Only "surprise" can have either positive or negative valence.
Intensity	used to represent the intensity of an affective state which is relevant to the importance, relevance and urgency of the message that the state carries.
Duration	used to distinguish between personality, mood and emotions; emotions are measured in minutes, mood in days, and personality can be permanent
Focality	used to indicate whether the affective phenomena is global as in moods in which the

	cause has become detached from the felt action readiness, or as in emotions which are mostly about something: an event or an object. Globality can also differentiate emotions: depression from sadness, bliss from joy, anxiety from fear.
Agency	used to indicate who was responsible for the emotion, the agent itself, or someone else.
Novelty	used to refer to whether a novel and unexpected stimulus occurred causing mismatch with the subject's expectations regarding the stimulus triggered.
Intentionality	used to refer to whether the triggering event is perceived as caused by some live intending agent.
Controllability	used to refer to how much the subject believes s/he/it can control the current situation
Modifiability	used to refer to duration and time perspective, or to the judgment that a course of events is capable of changing.
Certainty	used to refer to anticipation of effects to come, and how (subjectively) certain the subject is about the consequences of the situations
External (social) norm	used to refer to whether the event conforms to social norms, cultural conventions, or expectations of significant other.
Internal (self) standard	used to refer to whether the event is consistent with internalized personal standards as part of the self concept or ideal self.
Action tendency	identifies the most appropriate (suite of) actions to be taken from that emotional state.
Causal chain	identifies the causation of a stimulus event.

Table 1: MAUI Appraisal Components

II. Specific Open Questions to Emotion Theorists

After working on the categorization of physiological signals, we are currently planning to use Scherer's appraisal theory [6;7;9] to guide our **facial expression recognition algorithms** which will study and incorporate state-of-the-art facial information processing results [1;8]. The theory will also guide the design of our **user model and adaptive avatar** [3]. We have just began our investigation and following is a list of some of the questions we would like to have answers to in order to enhance MAUI's abilities to acknowledge the communication power of facial expressions.

What does facial expression express?

It is said that emotion episodes come from appraisal-determined cumulative patterns of component synchronizations ([7] p.150), why it is contrary to dimensionalists? Why not represent the emotion in a N-dimensions space since there is a cumulative process?
Even from the table of Hypothetical outcomes of SECs, emotional states are somewhat represented in an 11D space. What are exactly "push" and "pull" effects in expression? ([7] p.148)

What is the precise meaning of the empty fields of Power Sadness [7] p. 151 (table of outcomes of SECs)? (even if it has been fixed in recent tables [6], we wondered why it was empty then).

Studying dynamic models of facial expression of emotion using synthetic animated faces

A synthesized face using FACE is said to be in 3-D ([9] p.107) and contours are represented by splines. Are those splines at fixed places (like wireframes) or are they computed from the separation of the current position of the face and the background? In which manner is it 3-D?

Are the faces only monocolor or grey-scaled? Grey-scaled would bring more refinement of the face and maybe more believable (and recognizable) expressions. ([9] p.110)

In study 2 ([9] p.110), there are 3 conditions; in the first, there is a cumulation of AUs; precisely, do they start at different times or together (but some with quick onset, some with slow onset)? In the second, facial expression appears in a "simultaneous way"; does that mean each AU composing the expression start at the

same time, and end together reaching their maximum intensity, changing in a linear way? Are the final expression of the two conditions exactly the same?

Does adding the same AUs in different order produce the same result or are they order-independent?

Appraisal considered as a process of multi-level sequential checking

Do SE Checks (former subchecks) within the groups checks also have a fixed processing order or not [6]?

Will sequentially cumulated checks of the same group lead to the same result if they are applied in different order? (for example, applying Control then Power, and Power then Control lead to the same result?)

Are there more details about the network architecture presented in figure 5.3 presented in [6] p.104?

III. References

[1] **Andrés del Valle, A. C. and Dugelay, J-L. (2002).** Expressions Facial Expression Analysis Robust to 3D Head Pose Motion, ICME, Switzerland.

[2] **de Rosis, F., Pelachaud, C., Poggi,I., Carofiglio, V. and De Carolis, B. (2003).** From Greta's mind to her face: Modeling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies Special Issue on Applications of Affective Computing in Human-Computer Interaction*, Vol. 59 (1-2):81-118.

[3] **Lisetti, C. L., Nasoz, F. Lerouge, C., Ozyer, O., and Alvarez, K. (2003).** Developing Multimodal Intelligent Affective Interfaces for Tele-Home Health Care. *International Journal of Human-Computer Studies Special Issue on Applications of Affective Computing in Human-Computer Interaction*, Vol. 59 (1-2):245-255.

[4] **Lisetti, C., Nasoz, F. (2004).** Using Non-Invasive Wearable Computers to Recognize Human Emotions from Physiological Signals. EURASIP Journal on Applied Signal Processing – Special Issue on Multimodal Interfaces, (in press).

[5] **Lisetti, C., Nasoz, F. (2002).** MAUI: A Multimodal Affective User Interface. In *Proceedings of the ACM Multimedia International Conference 2002*, (Juan les Pins, France, December 2002).

[6] **Scherer, K. R. (2001).** Appraisal considered as a process of multi-level sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.). *Appraisal processes in emotion: Theory, Methods, Research* (pp. 92-120). New York and Oxford: Oxford University Press.

[7] **Scherer, K. R. (1992).** What does facial expression express? In K. Strongman (Ed.), *International Review of Studies on Emotion.*Vol. 2 (pp. 139-165). Chichester: Wiley.

[8] **Valente, S.. and Dugelay, J.L. (2001).** A visual analysis/synthesis feedback loop for accurate face tracking. Image Processing Image Communication.

[9] **Wehrle, T., Kaiser, S., Schmidt, S. & Scherer, K. R (2000).** Studying dynamic models of facial expression of emotion using synthetic animated faces. *Journal of Personality and Social Psychology, 78(1)*, 105-119.