Issue 04

Emotional IT Fellows

July 30, 2009 Volume 2, Issue 4

The CALLAS Project: **C**onveying **A**ffectiveness in **L**eading-edge **L**iving **A**daptive **S**ystems

A CALLAS Newsletter

In this issue:

Visit us at:
www.callas-newmedia.eu

Contact us:
info@callas-newmedia.eu

Emotional models in CALLAS

By STEVE GILROY

The CALLAS project combines a variety of emotion-based input and output components.

In order to be able to easily assemble these components into a variety of arts and entertainment projects, a common descriptive model of emotional information or transformation between alternative representations is needed.

There also needs to be compatibility with our approaches to affective multimodal fusion and higher level modelling of emotion-based processes.

Existing low-level models of emotional state fall into two general types: either positing categories of universal emotions (whether based on theories of physiology or cognitive psychology) or dimensions of a continuous emotion space in which the current state lies.

The exact number of distinguishable emotions, as well as the representation or even existence of "combined" emotions is an open research topic, with no overall consensus. Some of these models are implicitly grounded into specific theories of emotional processing.

As CALLAS components utilise a wide range of input data, it is hard to choose a coherent set of discrete emotion categories that is compatible with all components. In addition, temporal aspects of more abstract artistic installations would benefit from continuous data derived from emotional information. A final consideration is the integration of the emotional reactions of multiple members of a diverse audience.

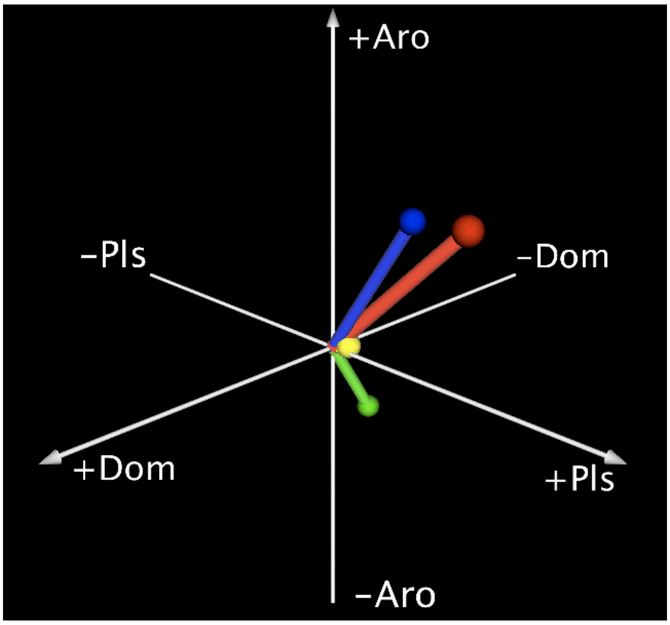
The emotional model adopted in CALLAS is an attempt to address some of these issues.

It is a continuous representation of emotional input that supports aggregation from multiple sensors and participants over the duration of an artistic/entertainment experience. It does not preclude the development of other emotional models within components or for a specific experience, but does support transformation between some other representations of emotion.

The model is a psychological model of emotion developed by Albert Mehrabian, which describes behavioural tendencies as an aggregation of emotional reactions to situations. This model distilled such reactions into three factors, which constitute three dimensions in an emotional space: Pleasure, Arousal and Dominance.

(Continued on page 2)

(Continued from page 1)



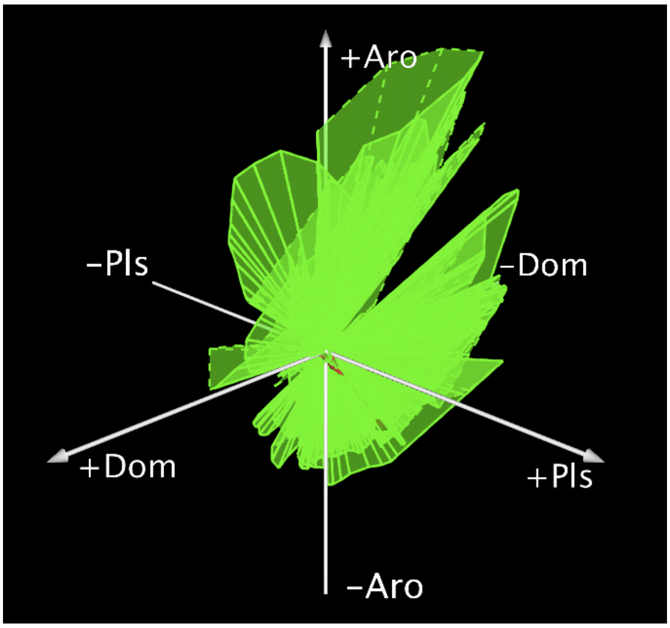
Component emotional vectors in PAD dimensional space

This model is therefore known as the PAD model.

The three dimensions of the PAD model are thought to correspond to orthogonal aspects of emotional response: pleasure-displeasure, corresponding to cognitive evaluative judgements; arousal-nonarousal to levels of alertness and physical activity; and dominance-submissiveness to the feeling of control and influence over others and surroundings. This was devised as a universal model, in that the states represented in the emotional space can also stand for aggregated tendencies of individuals, and overall emotional representation of a combined series of events. It is also universally independent of higher-level models of how emotions are processed or generated.

In CALLAS, we represent the output of a component as a 3D vector in the PAD emotional space, and combine these representations into an overall characterisation of the emotional response of an audience en masse to an installation, through an affective multimodal fusion process.

The continuous nature of this dimensional model is appealing, as explained earlier, as it allows us to model intermediate states of affect that may not have an a priori label or categorisation. This means that a component's output can be a combination of input from multiple persons, or an overall interpretation of a mass response that does not rely on precise identification of a discrete emotion. Temporal aspects come into play as we represent the dynamics of a changing emotional response.



Overall emotional vector path through PAD dimensional space during an interactive session

Different sampling rates, timelines and temporal congruence of input modalities make fusion of on-going data into a series of discrete emotions difficult, but, with a continuous PAD representation, each input component has a constantly changing vector representation which can capture the temporal characteristics of component modalities. This is augmented by additional temporal smoothing techniques in our fusion process. The path of the combined PAD representation over the duration of an experience reveals the changes in mood as transitions through different volumes of the emotional space.

Mehrabian, A. 1995. Framework for a comprehensive description and measurement of emotional states. Genetic, Social, and General Psychology Monographs, 121, 339-361.

Greta features and use in CALLAS showcases

By RADOSLAW NIEWIADOMSKI, SYLWIA HYNIEWSKA and CATHERINE PELACHAUD

Greta is a 3D embodied conversational agent that implements the MPEG-4 animation standard. It is able to communicate using verbal as well as nonverbal channels like gaze, head and torso movements, facial expressions, and gestures. Greta is a general purpose agent that allows one to build interactive applications. Our agent architecture is modular, distributed and compatible with the SAIBA framework. Each module exchanges information and data through a central message system. We use the concept of whiteboard that allows internal modules and external software to be integrated easily.

Two standard languages can be used to control the behavior of the agent: FML-APML is a high level language that is used to describe its communicative intentions, while BML describes verbal and nonverbal behaviors of the agent. The system is designed to be used in interactive applications working in real-time. Several interactive applications of our system were developed within the SEMAINE and CALLAS projects.

The Architecture

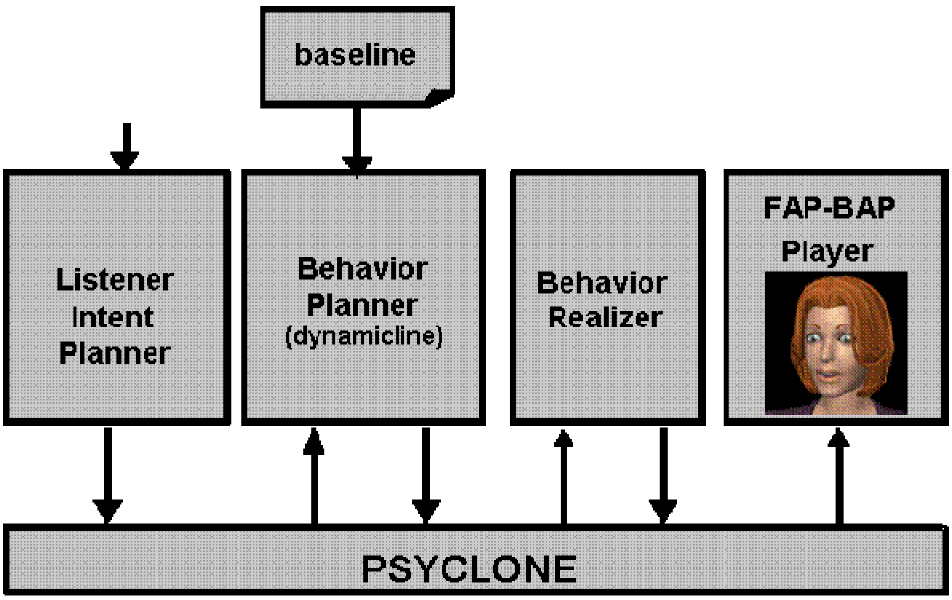
Greta's architecture is an almost full implementation of the SAIBA framework.

It is composed of three main modules (see Figure below), offering solutions for the *Behavior Planner* and the *Behavior Realizer* and a partial implementation of the *Intent Planner*. In the SAIBA standard the *Intent Planner* is dedicated to the intention generation of a virtual Speaker agent.

To be able to control a Listener agent, we have introduced the *Listener Intent Planner*, which generates automatically the communicative intentions of the listener.

In the current state of our system, when the agent is the Speaker (and not the Listener) its intentions are pre-defined manually in an FML-APML input. In the future these files should be generated by a *Speaker Intent Planner*.

Each communicative intention generated by the *Intent Planner* (whether the agent is a Speaker or a Listener) is transmitted to the *Behavior Planner*, in the FML-APML language. The *Behavior Planner* proposes a list of possible corresponding nonverbal behaviors, written in the BML language. These behavior signals are sent to the Behavior Realizer that generates MPEG-4 Facial Animation Parameters (FAP) and Body Animation Parameters (BAP) frames. Finally, the animation is played in the *FAP-BAP Player*.



GRETA's architecture

(Continued on page 4)

(Continued from page 3)

All modules are synchronized by the Central Clock and communicate with each other through the Psyclone white-board (Thórisson et al., 2005).

Multimodal sequential expressions of emotions

Initially, the emotional expressions of ECAs were based mainly on the work of Paul Ekman and his colleagues (Ekman1975, Ekman2003) who postulate the existence of universal emotional expressions that can be described at the apex. They described six such expressions displayed on the face: anger, disgust, joy, fear, sadness, surprise. The recent results in psychology (e.g. (Keltner1995, Shio-ta2003)) show that several emotions are expressed rather by a set or a sequence of different nonverbal behaviors which are arranged in a certain interval of time rather than by a static facial expression. The expressions of emotional states are dynamic and they can be displayed over different modalities like the face, gaze and head movement (Keltner1995), gestures (Keltner1995), or even posture (Pollick2001, Wallbott1998). Interestingly, these signals do not have to occur simultaneously (Keltner1995).

We called these expressions *multimodal sequential expressions of emotions*. They may be composed of nonverbal behaviors displayed over different modalities, of a sequence of behaviors or of expressions within one modality that change dynamically. A model of multimodal sequential facial expressions of emotion for an Embodied Conversational Agent was developed within the Greta system. The model is based on video annotations and on descriptions found in the literature. A language has been derived to describe expressions of emotions as a sequence of facial and body movement signals.

The data from the observational studies need a formal representation. In our approach we use a symbolic high level notation which gives us some flexibility in the generation of possible behaviors. New emotions may be added manually at any moment and their definition is easy to interpret by a human user.

Our XML-based language defines multimodal sequential expressions in two steps: behavior set and constraint set. Single signals like a Duchenne smile, a shake or a bow are described in the repositories of the character's non-verbal behaviors.

Each of them may belong to one or more behavior sets. Each emotional state has its own behavior set, which contains signals that might by used by the character to display that emotion. However, according to the observational studies, the signals' occurrence in an emotional display is not accidental.

The relations that occur between the signals of one behavior set are more precisely described in the constraint sets. In our algorithm, the appearance of each $signal_{si}$. in the animation is defined by two values: its start time, $start_{si}$ and its stop time $stop_{si}$. During the computation, the constraints influence the choice of values $start_{si}$ and $stop_{si}$. for each signal to be displayed.

A study was run to evaluate the recognition of the multimodal sequential expressions synthesized with the Greta agent. 53 participants took part in the study. They evaluated eight animations displaying an affective state (anger, anxiety, cheerfulness, embarrassment, panic fear, pride, relief, and tension).

Participants were asked to choose which label from a list of eight possible emotions described the expression best. In this evaluation we considered eight different emotional states: anger, anxiety, cheerfulness, embarrassment, panic fear, pride, relief, and tension. To create believable multimodal expression for the agent for each emotion, data is needed on the sequence of appearance of different behaviors. We took two approaches to reach this aim. On one hand we gathered data from the literature and on the other hand we manually annotated visual data. The recognition rate of these expressions is higher than the chance level making us believe that our model is able to generate recognizable expressions of emotions, even for the emotional expressions not considered to be universally recognized.

An example of a multimodal expression, based on the annotation of panic fear (above) and a multimodal expression of embarrassment (below)

Greta in CALLAS Showcases

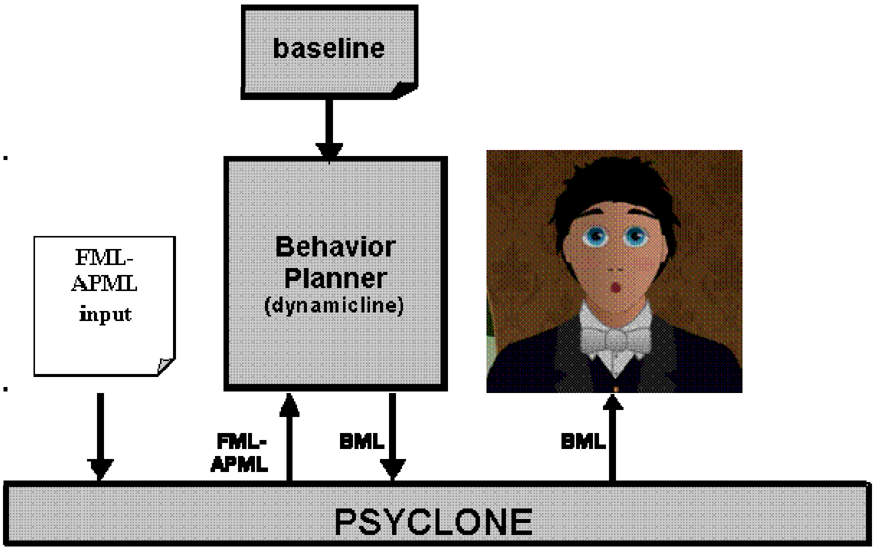
The Interactive Storyteller is done in collaboration with the BBC and Reading University. In this scenario a computer application presents a story content and displays a sequence of still images (with a high emotional impact) related to the story. A web site at the BBC is dedicated to public presentations of news items. Photos accompanied by explicative captions relate an emotional content. A particular story reports images on the Schezuan earthquake.

In the Interactive Storyteller application the interaction takes the form of a guided conversation between the virtual agent and the user. The ECA is a storyteller. Its role is to interact with the user. It first asks the user to comment each displayed image, then provides some explanation. To enhance the emotional experience of the user, the agent shows affective empathy toward the user showing its engagement in the story. The user expresses its opinion about the images. In the background its speech and gestures are analyzed by a system to detect the user's emotional states. For this purpose the MKS Keyword Spotting [1] and Gesture Expressivity Recognition [2] software provided by CALLAS partners are used. Next, the agent explains what the image is about.

It uses various nonverbal signals like emotional facial expressions and gestures to emphasize the message and guide the emotional reactions of the user. Detected information about user's affective state is used by the system to influence the agent's affective behavior.

When the ECA begins conveying the scene s_{i+1} it expresses the same affective state as measured from the user in the previous segment of the story s_i and then, throughout the duration of the scene, it gradually changes its expressed emotional state to the s_{i+1} target affective state.

(Continued on page 6)



Greta's architecture in the MUSICKIOSK showcase

[1] Jérôme Urbain and Thierry Dutoit, Faculté Polytechnique de Mons, Belgium

[2] George Caridakis, Stelios Asteriadis and Amaryllis Raouzaiou, Image, Video and Multimedia Systems Laboratory, National Technical University of Athens, Greece

(Continued from page 5)

Behavior Control System in Music Kiosk

Greta system is modular. Processes are connected with each other by specific languages of representation, namely FML-APML, BML and FAP-BAP of MPEG-4. In the music kiosk scenario, the physical appearance of Greta is not used. Rather characters designed by XIM can be used. The system of Greta used in the Music Kiosk scenario is shown in Figure 4. It takes in input an input file written in FML-APML. It corresponds often to emotion tags.

The output of the system is a list of BML tags with their timing information. BML tags can be 'smile', 'raise eyebrow', 'pointing gesture', etc. These tags can be sent to the animation player of XIM to control the animation of 2D characters.

ENGINEERING INGEGNERIA INFORMATICA

CALLAS @ eNTERFACE 2009

By SILVIA BOI

The CALLAS Consortium has the pleasure to announce its participation at **eNTERFACE '09,** offering the opportunity to share and validate the CALLAS results by people not directly involved in its project developments but interested in the Multimodal Interface research areas.

Two challenging projects are being proposed to session attendees and jointly developed during the four week period from the 13th of July to the 7th of August 2009.

***The AVLaughterCycle project*:** which aims at developing an audiovisual laughing machine, capable of recording the laughter of a user and responding to it with machine-generated laughter linked with the input laughter. The intention is that the initially forced laughter of the user will progressively turn into spontaneous laughter.

The project is being led by three CALLAS consortium partners: THIERRY DUTOIT Dutoit (Faculté Polytechnique de Mons), CATHERINE PELACHAUD (TELECOM Paris-Tech) and AMARYLLIS RAOUZAIOU (ICCS-NTUA).

Laughter is an important element in human communication: it conveys information about our feelings and helps to cheer us up. It has, moreover, the potential to elicit emotions in its listeners, to have a healthy effect, and especially to be one of the best medicines against stress: laughter therapies, "yoga" sessions or groups are emerging everywhere, especially in these times of crisis in which people need to be able to relax .We can cite the laughter chain launched on Skype, Laughter Clubs or the World Laughter Day. In addition, due to the growing interest in virtual agents modelling human behaviour, there is a need to enable these agents to perceive and express emotions. Laughter is clearly an important cue for understanding emotions on the one hand, and manifesting certain emotions on the other.

From an engineering point of view, laughter has not been subject to detailed study. Due to its large variability, it is hard to acoustically describe its structure and only a few systems are able to detect laughter, and laughter synthesis is, to date, inefficient.

Emotional Conversational Agent called Greta

The Interactivity of an affective puppet, a project: which aims at integrating a variety of affective multimodal components developed in CALLAS and at fine-tuning the multimodal and affective interactive loops of a computer animated puppet.

Animator using human glove

The project is being led by three CALLAS consortium partners: STEFANO ROVEDA (Studio Azzurro), MASSIMO BERTONCINI (ENGINEERING Ingegneria Informatica) and GIULIO JACUCCI (Helsinki Institute for Information Technology).

The work is based on Neoclide, a virtual puppet developed by Studio Azzurro, a leading art and technology group worldwide , and a partner in CALLAS. The scenario to be developed foresees that the pre-recorded animation (voices, behaviour, effects) of the puppet is played in real-time, according to interactions with the spectators, whose movements, facial expressions, voices and other signals will be analysed to derive overall emotional states and other semantic aspects that will trigger the animations of the puppet. The puppet acts by respecting a variety of ethical guidelines, communicating about information and science, providing a narrative as a creature from another world, captive inside the screen, willing to learn about the human world. This scenario covers the area of new forms of theatre with stronger visitor/spectator participation by focusing on extending the concept of performance in time: live puppeteering extended into interactive scripted puppeteering, live performances augmented with real-time affective effects, live animation without the presence of the performer. Spectators can implicitly or explicitly provide utterances (gestures, facial expressions, texts) that become part of a repertoire of utterances used by the puppet. This concept has been used before by an artist (his installation has been hosted in the Interactive art Exhibition called ParticipART at the MART). This is a state of the art theme in new form/formats of interactive performance arts. There are two parts in future work to be further developed in this application that give opportunities to integrate many components, such as the real-time affective loops interacting with Euclide and the provision of utterances from spectators to be reused.

The CALLAS Consortium gratefully acknowledges the European Commission co-funding of the CALLAS project N. 34800
under the FP6 Programme, that made this publication possible

The **CALLAS Consortium** is composed by the following members:

- Engineering Ingegneria Informatica S.p.A. - Italy (Co-ordinator)
- VTT Technical Research Centre of Finland
- British Broadcasting Corporation - United Kingdom
- Studio Azzurro Produzioni S.r.l. - Italy
- XIM Ltd. - United Kingdom
- Digital Video S.p.A. - Italy
- Humanware S.r.l. - Italy
- NEXTURE Consulting S.r.l. - Italy
- University of Augsburg - Germany
- Institute of Communication and Computer Systems / NTUA - Greece
- University of Mons, Faculty of Engineering, - Belgium
- University of Teesside - United Kingdom
- Helsinki Institute for Information Technology, HUT and UH - Finland
- TELECOM Paris-Tech - France
- Scuola Normale Superiore di Pisa - Italy
- University of Reading - United Kingdom
- Fondazione Teatro Massimo - Italy
- Human Interface Technology Laboratory - New Zealand

The CALLAS Editorial Team:

MASSIMO BERTONCINI, IRENE BUONAZIA,
ANNAMARIA DE SANTIS,
DAVID FUSCHI, CATHERINE PELACHAUD and
ANTONINA SCUDERI