EMOTIONAL INTERFACES IN PERFORMING ARTS: THE CALLAS PROJECT

Massimo Bertoncini – CALLAS Project
Engineering Ingegneria Informatica, R&D Lab
Via San Martino della Battaglia 56,
Roma
Italy

massimo.bertoncini@eng.it

Irene Buonazia – CALLAS Project
Scuola Normale Superiore di Pisa
Piazza dei Cavalieri 7
Pisa
Italy

i.buonazia@sns.it
http://www.callas-newmedia.eu/

Abstract - CALLAS project aims at designing and developing an integrated multimodal architecture able to include emotional aspects to support applications in the new media business scenario with an "ambient intelligence" paradigm. The project is structured in three main areas: the "Shelf", collecting multimodal affective components (speech, facial expression and gesture recognition); the "Framework", a software infrastructure enabling the cooperation of multiple components with an easy interface addressed to final users; and three "Showcases" addressing three main fields of new media domain: AR art, Entertainment and Digital Theatre, Interactive Installation in public spaces and Next Generation Interactive TV.

INTRODUCTION – THE CALLAS PROJECT

CALLAS - Conveying Affectiveness in Leading-edge Living Adaptive Systems - is an Integrated Project founded by the European Commission within the 6th Framework Programme Information Society Technologies priority, in the strategic objective Multimodal Interfaces (2.5.7). The project started in November 2006 and will end in May 2010. The project consortium is composed of universities and private research laboratories working on multimodal applications, together with artists, broadcasts and theatres, involved as final users [1].

MULTIMODAL AFFECTIVE INTERFACES: OBJECTIVES AND DOMAIN

In everyday life, human communication combines speech with gestures, movements, and non-verbal expressions: each of those communication channels is affected by emotions. Taking in consideration the role of emotions and affectiveness is therefore fundamental to enrich naturalness also in human-machine interaction and communication. The CALLAS project face the challenges of implementing innovative affective interfaces to comprehend emotional input within the domain of interactive media. Affective and emotional interfaces are generally concerned with the real-time identification of user emotions to determine system response. They usually rely on Ekmanian emotions such as anger, fear, sadness, enjoyment, disgust and surprise.

The domain of interactive new media, such as interactive narratives, digital theatre or digital arts, involves different ranges of emotions on the user's side, some of which correspond to responses to aesthetic properties of the media, or characterize the user experience itself in terms of enjoyment and entertainment. To identify these ranges of

emotions, more complex articulations of modalities are required across semantic dimensions as well as across temporal combinations. For instance, input from emotional language and paralinguistic speech (laughter, cries) must be categorized as indicators of user attention and must be integrated across interaction sessions of variable durations, instead of analyzing a single emotional status in real-time.

The first scientific CALLAS objective is to advance the state-of-the-art in Multimodal Affective Interfaces, by creating new emotional models, able to take into account a comprehensive user experience in Digital Arts and Entertainment applications, and by developing new modalities to capture these new emotional categories. The main technological research (developed by the universities and research labs participating in the project) consists in the development and integration of advanced software components for semantic recognition of emotions. Those components will be available through a "living" repository, called the CALLAS "shelf". On the other side, the project (mainly with the effort of the software and engineering companies of the consortium) aims at establishing a software methodology for the development and the engineering of a "framework" for Multimodal Interfaces that will make their development accessible to a larger community of users (even without a deep understanding of the theories of Multimodality), represented in the consortium by theatres, broadcasts and digital artists.

Finally, the effectiveness of CALLAS approach will be validated developing research prototypes in the domain of digital media, arts and entertainment.

In recent years, new media has developed largely in terms of richness of digital contents (combining text, video and sound) and of technical sophistication. On the other hand, emerging technologies, such as ubiquitous computing, augmented and virtual reality, human-computer interaction, and context and location-awareness, are making possible a paradigm embracing users' natural behaviour as the centre of human-computer interaction. Most New Media are actually interactive, and rely on digital content for which user interaction plays a central role. The domain of digital cultural content (digital theatre, mixed reality arts, ubiquitous systems supporting interactive storytelling and TV) is specially challenging, involving a wide range and combination of sophisticated users' emotions and feelings. This particular domain chosen by CALLAS imposes to advance the understanding of emotional interaction, taking into account also non-Ekmanian emotional categories.

CALLAS COMPONENTS: THE SHELF

The Shelf consists of a dynamic pool of advanced multimodal interface technologies, selected taking into special account efficiency and robustness, in order to guarantee a consistent performance for many contexts and scenarios, specially for the use in uncontrolled "production" scenarios, while many technologies are developed and tested in controlled settings. CALLAS shelf components include:

- Emotional Speech Recognition: combines keyword spotting in utterances with information about the emotional state of the speaker, according to correspondences with a list of Ekmanian and non-Ekmanian emotions (component specially developed by Faculté Polytechnique de Mons);

- Emotional Natural Language Understanding (developed by University of Augsburg): includes acoustic as well as linguistic features, relying on a corpus-driven approach;

- Sound Capture and Analysis (component developed by VTT): maps on emotional patterns speech, surrounding sounds (music, crowd cheering), to guarantee natural and adaptive interaction with physical and virtual environment, as well as in the creation of MR/AR environments;

- Video Feature Extraction: extracts contextual and emotional information about users, environment and media from video streams combining audio and visual information. The component especially analyses video streaming on wide spaces, tracking speed, direction and quantity of movement of items in the space;

- Gesture and Body Motion Tracking (developed by VTT): provides information about body movement and gestures interpreted as thresholds to different emotional states. The tracking, especially focussing on hands movements, will be performed with different sensors positioned from upper limb to the whole body;

- Haptic Tracking component is a 3D haptic tracker for virtual environment navigation, based on interpretation of force/tactile feedback (developed by Humanware); it will be further developed into Wearable Interfaces for Motion Capture, embedding different miniaturized transducers for gesture recognition and motion tracking;

- Multimodal Interpretation of User Experience (developed by University of Teesside) researches on the emotional categorisation of the user experience, aiming at defining a new paradigm for investigating emotions in multimodal interfaces;

- Affective Multimodal Interpreter / Facial Expression Recognition extracts expressivity from gaze detection, facial features (measured through coordinates of interest point in the face), and gesture recognition (measured through head-hands coordinates). Such components, operating on high-resolution images of frontal faces and signals coming from many sensors, are developed by ICCS. The output of such researches should be an Expressivity Synthesis, able to generate, from image sequences, sensors, history and personality details, an expressive model of user's behaviour, to be performed by ECA.

- Emotional Natural Language Generation component (to be developed as a research output of the project by University of Augsburg) is responsible for generating natural language without disregarding the affective aspects of a conversation. It is based on an annotated corpus consisting of sentences that present typical expressions used in a conversation. The corpus has to be annotated with categories and topics that the sentences are about and also with the emotional state that they denote.

- Affective Music Synthesis (developed by University of Reading) component aims at enhancing musicality and music expression of virtual actors according to the user's mood, making users' experience of sound and music less mechanical.

- Emotional Attentive ECA (Embodied conversational agents, based on achievements of ECA developed by University of Paris8), will investigate on three core capabilities of ECA in emotional and social context: emotional communication through gesture, facial expression, gaze, body; emotional expression in a social context by blending or masking emotions; modelling perceptual-attentive social behaviours that are a basis for interaction, such as mutual, joint and shared attention.

THE CALLAS INTEGRATED APPROACH: THE FRAMEWORK

The aim of the CALLAS project is to develop a system able to combine emotional components and features in new modalities, enabling different modes of integration, as required by various applications, offering pre-assembled, re-usable, and semantic fusion

components.

The CALLAS Framework is being designed as a software infrastructure that will allow a number of Shelf components to work together to build specific end-users applications in the field of Digital Art and Entertainment. Most of the Shelf components are able to gather not only what a spectator asks to the system or decides to communicate to it, but also information related to his/her emotional state. The CALLAS framework aims at collecting all the partial displays of the affective involvement of spectators, at merging them together in order to really deduce what the audience is feeling and, as a consequence, at producing a proper affective and multimodal response. What makes CALLAS unique is the combination of active and passive modalities. In CALLAS, the emotional aspects of the interaction are elicited from a semantic representation of what is conveyed through both active and passive modalities. This means that in CALLAS the fusion is performed at the semantic level and that the multimodality is adopted to perform a new human computer interaction model, based on the emotional involvement of the audience.

One of the main aims of the CALLAS framework is to provide an intuitive metaphor suitable for non-technical users (mainly artists) willing to adapt and repurpose the CALLAS high-level components or their combinations. Moving from the analysis of the state of the art of many projects in the same field and cooperating with OI and CHIL projects [2], a large number of requirements have been analysed. The development of the framework will move from identifying the first aggregation schemas and developing the first composition components, clustering only a subset of components; then the framework will refine and extend integrations, powering performances and adding new functions, according to specifications coming from the first prototypes (showcases).

After the first analysis a Blackboard Pattern has been identified as a possible solution for a suitable metaphor, easy to use for end users eventually without special technological expertises in multimodal interfaces.

As a plug-in architecture able to glue together Shelf components, the CALLAS Framework is a software infrastructure making life easier for digital art and entertainment application developers. When developers decide to design and implement a multimodal interactive application, they can use a suite of open-source and interoperable toolbox and software subassemblies to save time in the application development. With this aim, the CALLAS project also promotes Technology Transfer, in particular towards SMEs operating in the new media Digital Arts and Entertainment sectors, and in close sectors, in which the CALLAS model can be replicable.

THE SHOWCASES

The effectiveness of the CALLAS approach in pursuing the afore-mentioned objectives will be validated developing significant research prototypes (or "Showcases") in some major fields of Digital Arts and Entertainment. The development of showcases will complement other research activities throughout their life cycle, by providing artistic requirements and context and user data; finally public installation of showcases will contribute in raising awareness about the CALLAS technologies through the setting up of a community of application providers and final users.

Traditionally, human-computer interaction has concentrated on the usability of applications, designed to enable users to perform a specified task as effectively and efficiently as possible: applications were primarily regarded as tools. In contrast,

designers of affective interactive applications rather focus on users experience, where applications are able to analyse and render emotions as part of an interactive system.

CALLAS, addressing at the field of new media for digital art and entertainments as a particularly complex environment for affective interaction between users and applications, identified three specific domains as main showcases, although other showcases could be developed during the project.

Augmented Reality for Art, Entertainment, and Digital Theatre: E-Tree

In recent years, many artists are working on performances and digital installations based on the individual or collective user / audience feedback in real time. CALLAS showcase in this field intends to stress the involvement of emotional experiences and participation. As examples, the participation of both spectators and actors can be enhanced during theatrical performances by developing applications able to modify in real time virtual set-design, light-design or music according to the affective status of actor and audience, for instance detecting paralinguistic expressions (cheers, buzz) from the theatre environment, or identifying gaze and facial expressions of (some) spectators. Such scenario imposes to solve requirements as multiple face detection, use of devices in the dark and without conditioning the perception of the show by spectators.

E-Tree − Emotional Tree - is a CALLAS showcase specially developed by University of Teesside. It integrates an Augmented Reality (AR) environment with CALLAS affective input components; it demonstrates how on-the-fly detection of the mood and the affectiveness of people involved in the AR installation can improve the naturalness of the experience. The AR installation is based on an original concept of the digital artist Maurice Benayoun: a virtual tree dynamically displays growth and evolution reflecting the perceived affective response of spectator(s). The showcase can support the exploration of "affective feedback loops" where a participant's response to dynamic artwork determines changes and development that occur within the artwork, to which the participant then also responds.

The development of the tree is driven by a dimensional affective model, which represents the combination of affective input that has been received in the showcase. This model has three dimensions: Pleasure, Arousal and Dominance, which run on a scale of -1.0 to +1.0. Pleasure and arousal are similar to more popular models of valence and arousal, but the additional dominance dimension helps to distinguish between otherwise similar dimensions that differ in the centre of control (such as fear and anger). Input from each component is translated into a 3D value, combined with the existing model values. The size and rate of affective input determines how quickly the model takes on a new value.

The changing state of the model represents the overall "mood" of the interaction, and with no input, the model slowly moves towards a neutral state, leaving the E-Tree as a record of the history of affective input. The development of the E-Tree is controlled by rules that define branch growth and branching. Rules are chosen using weighted probabilities and parameters that are determined by the affective model. Positive values will assign a higher probability to growth and branching, while negative values assign more probability to no branching, slower growth, and off-axis branching and growth angles. Additionally, the parameters of existing rules are adjusted by current affective input. This means both the current "mood" and the mood that existed when a branch was created contribute to the look of the tree.

At this phase of the project, the showcase has been developed integrating audience inputs coming from movement, gesture and facial expression recognition, and with keyword spotting (In English language), with multiple users.
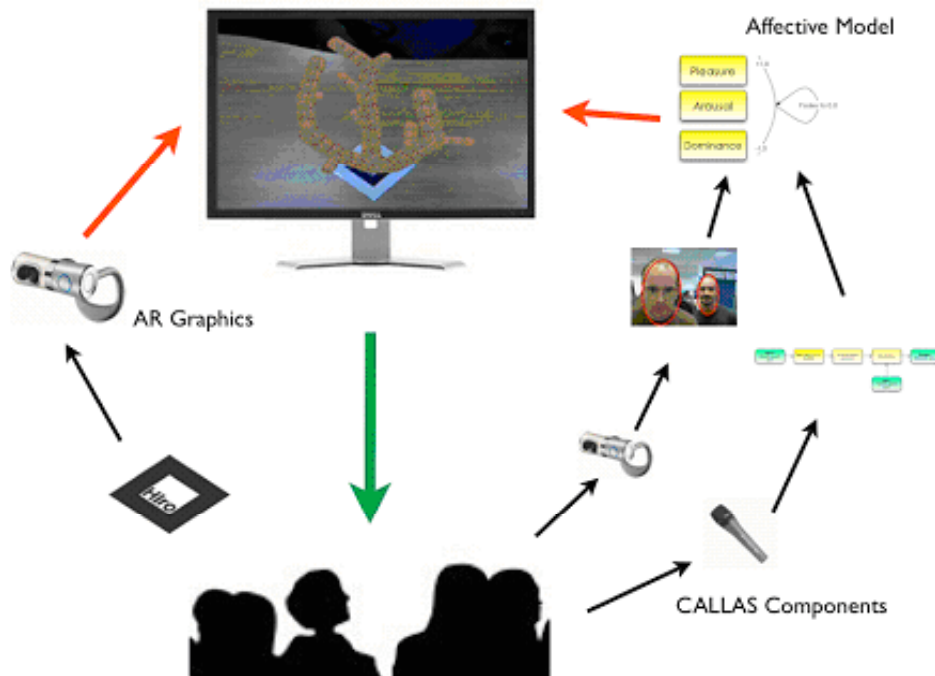

Figure 1. E-Tree showcase model

Interactive Installations for Public Spaces: Puppet wall

This showcase demonstrates multimodal affective applications for public places, enhancing the experiences of visitors or local groups during festivals and events, or in contemporary art museums, thus, changing the way public spaces are perceived and letting people re-configure the spaces they inhabit and visit. The concept is based on the idea to see the general public as a protagonist in new media, turning from passive consumers to active performers and creators. Innovative applications, enabled by the novel CALLAS interface technologies, allow users to animate media in real time using multimodal and emotional inputs.

Puppet Wall transfers the concept of puppet theatre into a digital domain with some novel characteristic, such as multimodal interaction, where the system considers spoken, gestural and bodily input; involvement of emotional intelligence, because detected emotional states of a user dynamically change the system; creation of flexible content by users, which are not tied to any fixed configuration of characters, objects and backgrounds, but can manipulate them and import their preferred objects; collocated co-creation, thanks to the participation of multiple users dialoguing between themselves.

With these properties Puppet Wall wants to provide the users an interesting, easy to use interactive installation that can be setup in any appropriate public setting.

The Puppet Wall is an interactive system consisting of a large touch-screen that enables people to collaboratively create and act out stories using either pre-created or their own media content. Some characters, which are customizable by users, act in the virtual environment with dynamic objects.

At this stage of the project, the multimodal affective interaction is achieved by three major inputs: hand movements are detected with the movements of a MagicWand, a wand with single LED of different colour as light sources, which are tracked in 3D with a pair of Firewire cameras, allowing users to control characters on stage, by moving and rotating them. Users can also manipulate directly puppets touching the screen. Finally users' speech is detected, more precisely its emotional indices. Gestural and auditory inputs are used by the system to manipulate the set and aspects of the characters.

The showcase allows the analysis of performative interaction loops to and from human behaviour and computational mechanisms (for instance how embodied expressions and utterances of users can be used to trigger affective animations that further affect the users and elicit new expressions). Therefore, the showcase joins together emotional input and direct action of the user, giving the opportunity to study how people understand emotional cues, how play with emotions (for instance over-reacting), or make inferences about the emotional states of the co-participants. In the future, additional components from CALLAS shelf will be added.
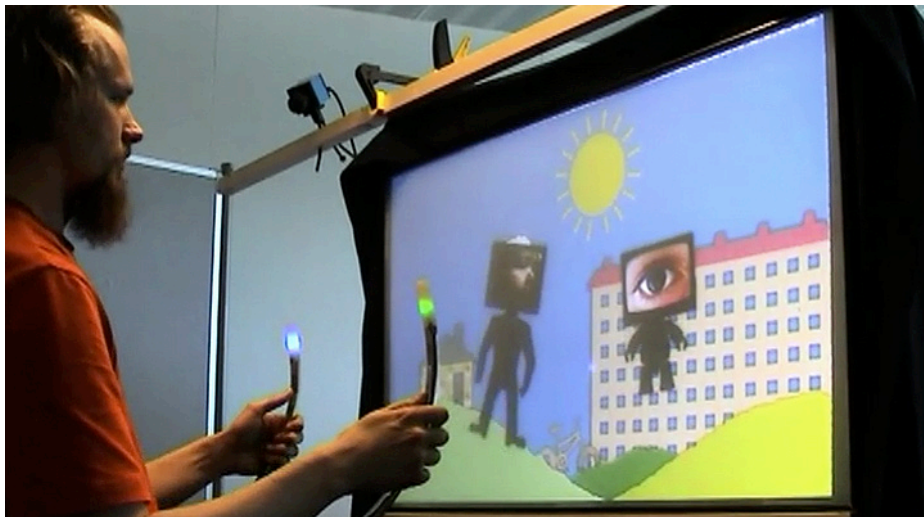

Figure 2. Single user controlling two characters using the first prototype of the PuppetWall

Next-Generation Interactive TV: Affective Interactive TV Installation

This showcase aims at providing a proof-of-concept of a next generation interactive television home platform, which utilises multimodal input to infer higher-level user inputs of affective state. The home platform will make use of an ECA to convey affective content related to both the desired outcome of the broadcast content and the user's perceived viewing experience. The core of the system consists of an interactive existing HTN-based storytelling engine, applied to a plot within the genre of comedic horror (a young lady is flat-sitting for a friend; it turns out that strange things are happening inside the flat), very suitable to appeal to and elicit derivatives of the most basic emotions: fear and humour. The interactive storytelling engine is integrated with an affective-enriched Embodied Conversation Agent (ECA) as co-spectator with expressive behaviour that matches the intention of the story contents. At this stage of the project, the system comprises an interactive narrative using traditional plan-based generative techniques, which is able to create situations exhibiting different levels of

tension or suspense (by featuring the main character in dangerous situations). Aside the evolving plot, an ECA is used to exaggerate the emotional value of a given scene so as to make it more visible to the user. User interaction is achieved through affective input devices to detect gestural actions (symbolic and deictic) and a multi-keyword spotting system detecting emotionally charged words and expressions.

Figure 3. The Affective Interactive TV Installation

References

[1]   The CALLAS consortium is composed by the following partners: Engineering Ingegneria Informatica SpA. – IT; VTT Technical Research Centre of Finland; British Broadcasting Corporation - UK; Metaware SpA. – IT; Studio Azzurro – IT; XIM Ltd – UK; Digital Video SpA. – IT; Humanware – IT; NEXTURE Consulting srl – IT; University of Augsburg – DE; Institute of Communication and Computer Systems - National Technical University of Athens – GR; Faculté Polytechnique de Mons – BE; University of Teesside - UK; Helsinki Institute for Information Technology – FI; Université Paris 8 – FR; Scuola Normale Superiore – IT; University of Reading – UK; Fondazione Teatro Massimo – IT; Human Interface Technology Laboratory - NZ.

[2]   OI – Open Interfaces Project, http://www.oi-project.org/; CHIL -Computers In the Human Interaction Loop, http://chil.server.de/servlet/is/101/.