Issue 02 Emotional IT Fellows

January 22,2009 Volume 2, Issue 2

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

The CALLAS Multimodal Affective Interfaces: unleashing a great economic potential for Digital Art and Entertainment

By MASSIMO BERTONCINI (ENGINEERING Ingegneria Informatica)

Digital Media or New Media (such as interactive television, podcasting, iPod, weblog) are increasingly being perceived as a dynamic two-way participative process asking for an improved and more *affective* human-human and human-machine **interaction**.

The emerging motto in this new era is "*User in the loop*": from being merely passive "consumers", users are dramatically changing their roles to being "active members" of a networked and participated community fully immersed in a pervasive technological environment, available anywhere, anyplace and anytime.

Emotions are always at the heart of successful traditional media (films, novels, music…), therefore the opportunity to interact with **New Media** by means of **emotion-sensitive interfaces** is expected to become a key enabler to allow a more effective many-to-many two-way human-human and human-machine communication and interaction.

New media content and application developers are required to deal with strong competitive pressure to reduce costs, increase production efficiency, yet enhance the user experience. The impact of piracy and file sharing puts pressure on raising quality and offering new levels of engaging interactivity.

Consequently, content creators are seeking ways to differentiate their products with innovative interfaces, allowing users to be fully engaged with a totally innovative, unforgettable experience, able to affect all their senses. Affective Multimodal Interfaces can become one of the most significant and compelling ways to strengthen innovation and competitiveness within the field of Digital Art and Entertainment. In particular, Affective Multimodal technology is **pushing** Process Innovation at two levels:

- Emerging technologies are changing the way in which the same things are made
- New services/applications are emerging (mobile tv, internet TV, Digital Art, Digital Media, Co-operative Geo-based Videogames,…)

As the CALLAS technologies become available in the course of the project lifetime, they are expected to generate two different, yet complementary revenue streams.

(Continued on page 2)

(Continued from page 1)

Revenue streams

1. **multimodal interactive applications** for developers to adopt affective interfaces, which will have **lower costs**,
2. creation of **new services**, with **increasing revenues.**

The **primary line for CALLAS commercial exploitation** is to offer in an **open-source** way the CALLAS Framework and provide consultancy to vertical application developers, service providers, producers, event organisers and network operators to create new interactive entertainment environments, accessible content and e-training applications.

The market for the CALLAS Framework will be mainly constituted by:

- ***CALLAS customers***, players (artists, technology providers) which aim at developing multimodal applications
- ***CALLAS final users***, organisations/institutions which are interested in utilising such multimodal applications (local authorities for festivals, theatres).

By applying software engineering principles, as well as an open source development model, CALLAS will make it possible for new media developers to integrate affective, multimodal interfaces into their products more **rapidly** and at **lower cost**.

Currently, the use of such technologies would require specialised research and discrete proprietary interfaces without the benefit of a synchronised, integrated framework approach.

The benefits of an open source licence are:

- The free availability of the Framework and the architecture enables every interested user to be part of the CALLAS system with low investment costs.
- New or improved functionalities can be added by the user community, also after the lifetime of the project.
- The CALLAS system can be ported to new platforms which are not supported by the project or which cannot be foreseen at the moment.
- The CALLAS results will be sustainable as the maintenance can be done by the user community as well as by companies.

The CALLAS Framework could also be used for developing applications in different domains, such as Call centres, travel industry, toys/robotics appliances, criminality detection. Such reusability will be ensured by the large effort which the CALLAS consortium will put in practice to build a large community of potential users of the technologies which the CALLAS project will make available during the project lifetime (the C^3-CALLAS Community Club).

The **second business stream** with a strong expected impact for CALLAS technologies will be the **generation of new revenue sources** due to the **creation of new services** (e.g. iTV applications). According to a BBC estimation, in about 5 years time from now, nearly 100% of digital TV programming will have an interactive element or content, which will be made available for truly interactive platforms, such as mobile devices, PCs, and games consoles. This is where innovation is required that is important to the iTV application developers and broadcasters. Primarily, the BBC is keen to re-capture the teenager audience, especially 14 - 20 year olds. This demographic category has largely drifted away from conventional broadcast content., which is considered to be no longer relevant to them, as they seek richly interactive, 'high octane' entertainment, with a strong element of social binding.

Essentially, a gradual culture change in TV audiences is occurring. According to BBC estimations, at present about 60% of the audience has little need for richly interactive broadcast content, while in 5 years time, that figure will have dropped to 30%. CALLAS represents a strategic leverage for sustaining the creation of new services.

For commercial broadcasters and other public content providers, this unveils opportunities for new revenue streams. For example, there is a potential revenue stream regarding the use of an emotionally sensitive ECA interface. The ECA on interactive TV could, for example, provide a much more tangible and entertaining front-end to Electronic Programme Guide (EPG) data than is currently available. This opens-up the prospect of a tiered data subscription service, where the user perceives that they are subscribing to the service provided by the ECA, rather than the data itself.

A subset of significant yet potential business cases is listed for both the two main expected business streams.

Cost Reduction Business Cases

1. SMEs developing digital interactive installations (e.g. Studio Azzurro, Humanware)
2. Traditional theatres (Teatro Massimo) using multimodal applications complementing physical scenographies
3. Local authorities, Digital Arts Museums setting-up interactive live performances ("white night" events, Open-space Installations, Interactive Installations in Museums or Local Districts, etc.)

Revenue Increase Business Cases

1. Large Service Providers enlarging their market to Digital Entertainment Business (Engineering)
2. Companies providing new services or tools in the growing interactive TV market (DigitalVideo, XIM, BBC)
3. Spin-offs based on providing advanced mobile services in festival (VTT, SNS)
4. "Theme Parks" using Mixed Reality Installations

(Continued on page 13)

One's Own Soundtrack: Music Affectivisation

by ATTA BADII, ALI KHAN and DAVID FUSCHI

(Intelligent Media Systems & Services Research Laboratory (IMSS), University of Reading)

The ability to interact with a subject depends to an extent on the level of empathy that can be reached. This holds in most contexts, but is particularly relevant in the case of Human Machine Interaction.

A successful film, opera, theatre piece, album or any other multi-media object bears an associated aural-memory. There are songs or tunes that characterise our memories of a certain event (be it happy, sad, pleasant or unpleasant) and this is because rhythms and audio information are extremely powerful conveyors of emotions, memories and feelings.

If it is true that often a picture is worth a thousand words, we could say that a melody is worth a full sensorial memory, so strong, and complete, it involves and activates all senses (olfactory and visual aspects of the environment). Audio has also often been used to support therapy in many contexts given its recognised

(Continued on page 4)

(Continued from page 3)

effect on people's mood. Thus it would be justified to say that if one could create one's own soundtrack and dynamically adapt it to the current mood; this would be a valuable result per se.

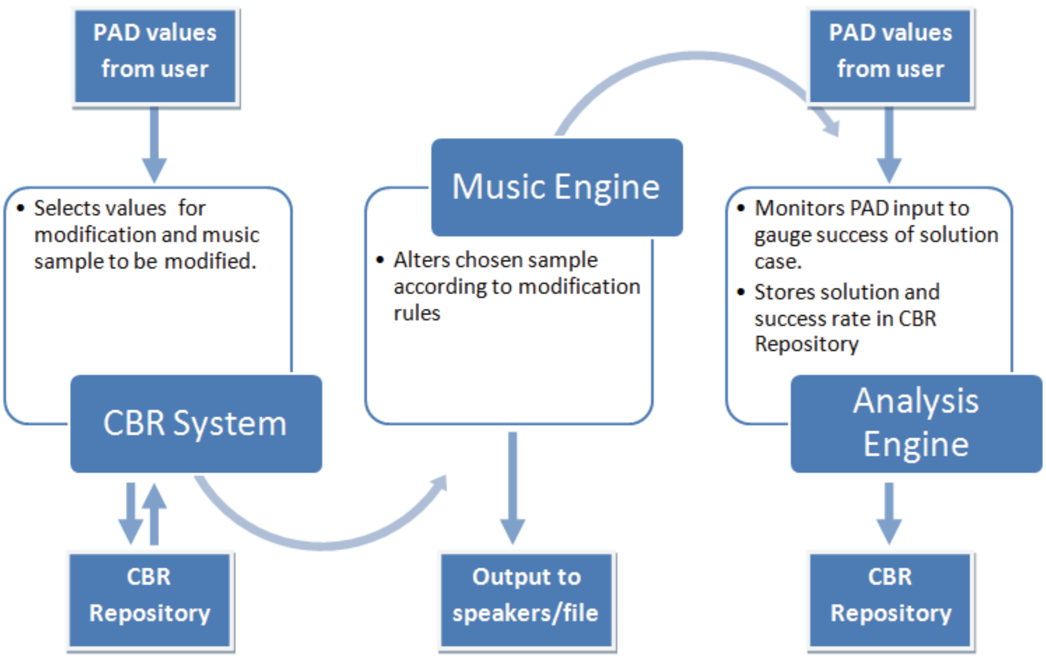
Affective Music Synthesis

In CALLAS we are currently developing a shelf output-component called Affective Music Synthesis (AMS) that enables the dynamic interactive modification of the characteristics of given music on the basis of the user's mood. In more detail the component is able to process a MIDI file altering some key characteristics (pitch, tempo and velocity) in response to the changing mood of the user (in terms of its Psychological Correlates as may be expressed, for example, in the form of Pleasure-Arousal-Dominance Model or for short PAD).

The current output-module is one of the two sub-components of the Affective Music Synthesis shelf output-component which we are currently developing. The overall component will be able to work on/off-line to either simply affectivise a collection (or a set) of musical pieces or to generate a musical piece from a repository of pre-recorded loopable samples that could be further affectivised either individually or as an ensemble once arranged together.

Our system is based on improvisation using cognitive models, case-based reasoning and fuzzy values acting on close-to-affect-target musical notes as retrieved from the case-based repository per context. It modifies the music pieces according to the interpretation of the user's emotive state represented by the Pleasure-Arousal-Dominance (PAD) model as computed by the emotive input acquisition components of the CALLAS framework.

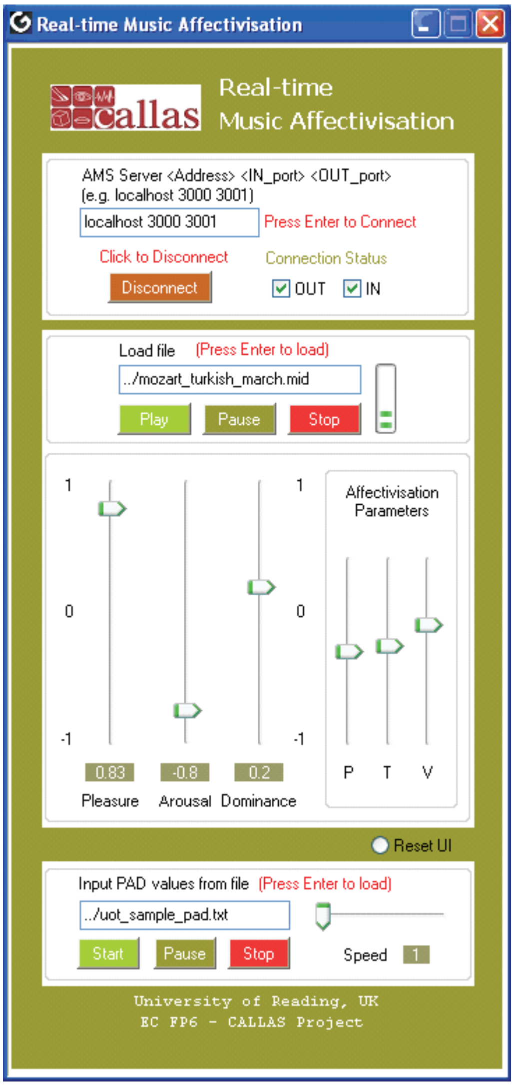
The PAD model representation of the user's emotive state characterises the criteria for the music affectivisation process. Using previous instances of affective adaptations from a case repository, a selection of music samples are adapted according to a user's mood. The PAD values are used by the case-based reasoning subsystem for selection and retrieval of this previously stored contextually suitable case in the case-based reasoning repository. Using combinations of positive and negative states for the affective dynamics along three affective axes (pleasure, arousal and dominance); the following octants of temperament space [1,2] are used as reference emotive states of the user: Exuberant, Bored, Dependent, Disdainful, Relaxed, Anxious, Docile and Hostile. This provides a level of interactivity that makes way for an interesting environment to experiment and learn about expression in music.



Real-time Music Affectivisation

The Affective Music Synthesis consists of a sub-system (developed in Pure data – Pd [4] with a graphical user interface based on GrIPD [3]) known as Real-time Music Affectivisation (RMA).

This sub-system takes as input a MIDI sequence and a stream of PAD values to output an affectivised version of the original MIDI sequence that reacts on-the-fly to varying emotive input. The RMA graphical user interface contains three vertical sliders representing Pleasure, Arousal and Dominance which may be set manually or by streaming PAD values from a file (e.g. a file that has recorded data from an input acquisition module).



In addition, the interface also contains three smaller vertical sliders representing pitch, tempo and velocity. These reflect the affectivisation parameters as they are computed and sent by the server. Upon a change in the user's mood i.e. the emotive input (represented by PAD) is communicated to the Affective Music Synthesis server which computes corresponding affectivisation parameters (pitch, tempo and velocity) on-the-fly using a case-based repository containing the octants of temperament space [1, 2] mapped to the affectivisation parameters of the music piece.

This mapping results from surveying a number of users and their feedback in terms of labelling a certain modified piece of music with a mood according to individual perception. Each time the server receives data on the emotive state of the user; it retrieves the corresponding affectivisation parameters from the repository, if exactly matched; else it computes the parameters by looking at the closest two matches.

The retrieved/computed parametric data is then sent back to the real-time sub-system which modifies the MIDI sequence accordingly. Ongoing work includes integration of our system with the CALLAS framework in order to receive an integrated emotive state from various live shelf-input components of CALLAS.

It is intended to make this system configurable in terms of PAD source so that it facilitates the user to choose whether to receive input PAD manually from the interface sliders, stream it from a file, or acquire it directly from a live input component (such as Facial Expression Recogniser or FER) via the framework.

MPEG-SMR

Future directions for our system includes incorporation of the MPEG Standard for Symbolic Music Representation also known as MPEG-SMR [5] to enhance the parameters of affectivisation of music (currently three parameters provided by MIDI are used: pitch, velocity and tempo).

(Continued on page 6)

(Continued from page 5)

Variable Mapping of PAD Model to Affectivisation Parameters

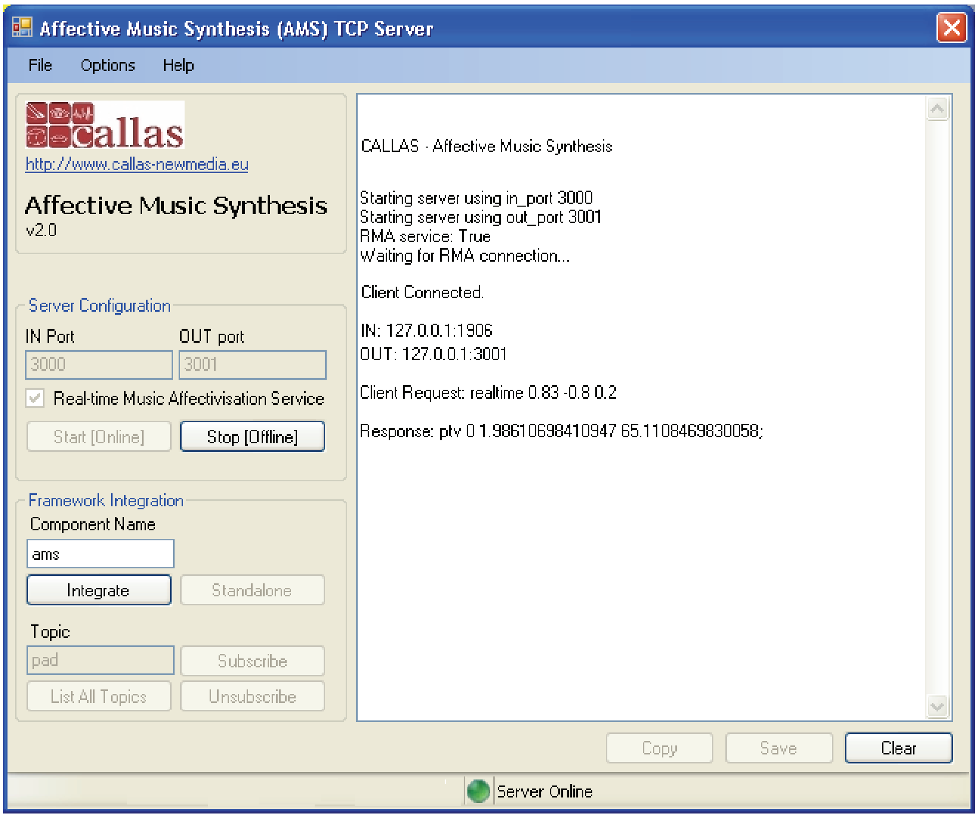
Another interesting development on the agenda involves the addition of a feature in our component to use alternating mappings of emotional models with affectivisation parameters. The system may then be setup for emotive input to be mapped in a directly or an inversely proportional fashion to the affective output, hence enabling the system to either resonate with or counter the user's mood. To make things more interesting, an automatic threshold switch for alternating between the two mappings could also be used. This threshold point could either be statically or dynamically computed by maintaining a history of user moods and affectivisation services rendered to use in a pre-specified time window.

Affective Music Synthesis Improviser

Currently, our system affectivises and modifies music in context of the emotive input representing and informing of the user's mood. An Affective Music Synthesis Improviser add-on may be developed to synthesise music using a repository of short loop-able samples such as beats, rhythms, melodies etc. Using a look-up table, this system would determine the various music characteristics to suit a particular user mood. Such a system would involve compiling together various pre-recorded samples together to form a short loopable piece of music. This piece could then also be streamed to the real-time subsystem for real-time affectivisation based on the emotive input that drove the sample selection process.

Together with the improviser add-on and the real-time music affectivisation sub-system, AMS would be able to remove, add and/or update new patterns on to the synthesised music as the user's mood changes from one state to another.

For example, if a mood change requires the tempo to be increased (upbeat), the real-time sub-system could increase the tempo and the improviser add-on could update the beat pattern being used resulting in an enhanced affectivised output; conversely the improviser could react with a downbeat rendition sympathetically.



Our future work will build on the achievements to-date to develop:

- Two types of emotive-state-to-affective-output mappings (directly- and inversely-proportional) including a manual or an automatic trigger/switch.
- An Affective Music Synthesis Improviser to compile affective music from a repository of pre-recorded short loopable samples.

REFERENCES

[1] Mehrabian, A. (1987). Eating characteristics and temperament: General measures and interrelationships. Springer-Verlag, New York.

[2] Mehrabian, A. (1991). Outline of a general emotion-based theory of temperament. In J. Strelau and A. Angleitner (Eds.), Explorations in temperament: International perspectives on theory and measurement (pp. 75-86). Plenum Press, New York.

[3] GrIPD – Graphical Interface for Pure Data. Available on WWW: http://crca.ucsd.edu/~jsarlo/gripd/

[4] Pure Data. Available on WWW: http://puredata.info/

[5] MPEG-SMR. Available on WWW: http://www.interactivemusicnetwork.org/mpeg-ahg/

Towards a user-friendly gaze estimation system

By STELIOS ASTERIADIS (Institute of Communication and Computer Systems / NTUA)

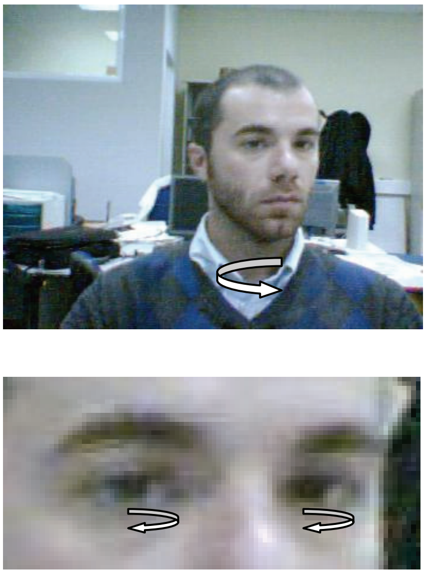
Automatic gaze detection plays an important role in many scenarios and applications within Human Computer Interaction environments.
Driver alertness estimation, gaming environments, learning environments are only some of the application domains where already available products exist, or intense research is being conducted in many institutions around the world.

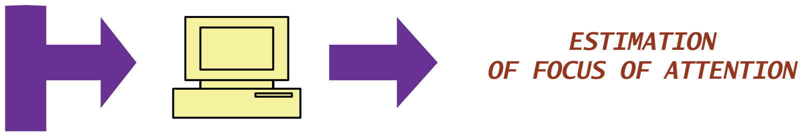
The problem of estimating the focus of attention of a person can be split in two major categories: Estimation of the pose of his/her head and estimation of the eye gaze.

There have been proposed in literature a serious amount of techniques, concentrating, either to both problems or addressing them independently, depending, usually, on the application. For example, mouse controlling is usually faced as an eye-tracking problem; however, under such circumstances, there is always the need to use hardware dedicated to such applications. Such hardware, in the best case, involves infrared cameras, located somewhere under the screen of the user.
The advantage of using infrared lighting is that ambient lighting plays only a minor role in processing the video sequences, while these methods make use of the so called bright pupil effect. Alternatively, some more intrusive systems make use of small high resolution cameras placed right in front of the eyes of the user.

The advantage of such methods in comparison to the previous ones is that they do not necessitate preprocessing, in the sense of using algorithms for localizing eyes within a whole frame.

(Continued on page 8)





(Continued from page 7)

However, the disadvantage of these techniques is that they require specialized hardware, which might not be accessible very easily.
For head pose estimation, some of the most typical techniques make use of helmets equipped with magnetic sensors. Although the accuracy of these techniques is high, they usually tend to be quite disturbing and, of course, again, such equipment might not be accessible to everybody. The use of only the head pose (leaving eye gaze apart) is widely known for research in the field of extracting knowledge from meetings. In these scenarios, the head movements of the participants (together with vocal data, many times) are processed and automatic decisions regarding roles are takes.
In the above methods, one could also add those methods that require cameras but, still, some initial knowledge of the characteristics of the setup or the camera(s) used should be known a priori. This kind of techniques also, usually can detect either head pose or eye gaze separately and, in the latter case, the user should stand as still as possible.
Furthermore, not a lot of methods exist for inferring head pose and eye gaze together, with one system, especially in an un-intrusive and un-calibrated environment, as it is difficult to extract gaze information from eyes in scenarios where the user moves freely and spontaneously.

In this field, this is the main focus of attention of our research: we are conducting research in estimating user focus of attention, combining information both from head pose and eye gaze. Our research takes into serious consideration the desire to keep our methods independent of controlled lighting conditions, spontaneous movements of the user, user-identity. That is, we make research on systems serving real-time needs, and not requiring that users have or feed the system with any kind of knowledge or training. To this aim, we use mono-camera systems that do not require high resolution. It is worth to notice that our experiments are run with the use of a simple web-camera, and have been conducted in in-office conditions; that is, we used normal lighting, with un-controlled background, while a multitude of people have tested it without needing the system to be tuned to any kind of conditions or users.

Our experiments have shown that, based on our findings, we can proceed with reasoning on user state estimation based on input stemming from head pose and eye gaze. To this aim, we have trained systems for extracting information in real time conditions. The extracted information is relevant to the user's engagement to the displayed content in a human-computer-interaction scenario.

Public Broadcaster as Modern Day Story-teller

By MARC PRICE

(BBC)

Can interactive TV ever be anything more than pressing the red button?

We believe that it can. The tricky bit is knowing what it would do, how it would work, and perhaps most importantly, why we would want to do it..

Why do audiences interact?

Interaction is essentially a form of communication. From conception through to death, we humans spend our time passively and actively engaging with our environment, in a bid to make sense of our world and place within it.

Humans are essentially learning machines. Learning is a natural process which normally occurs through exploration and play – a two way communication process which is nominally known as a relationship.

At a basic level, learning might consist of understanding how to fulfil basic needs such as eating or understanding basic gravitational phenomena.

As we grow, our learning quickly grows to encompass relationships of increasing complexity. The need to interact, therefore, is simply an inherent behaviour; a need to establish relationships of increasing complexity with all aspects of our world.

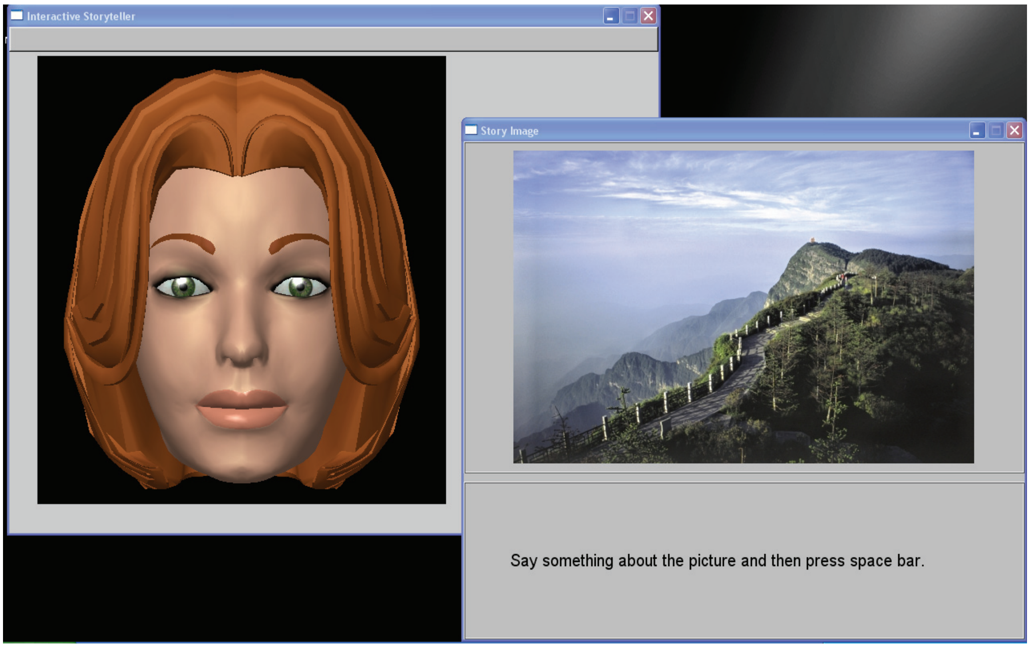
Pressing a red button clearly does not satisfy that need.

Perhaps we need a story?

A core part of the learning process is the sequence of affective states that are felt over the course of a relationship. Over time, this complex knowledge can be drawn upon as reference experience: once you have burned your hand on a hot coal, you know not to touch things that are hot. Moreover, knowledge tends to be context independent and transferable to others through higher-level relationships.

A powerful way of doing this is through story-telling – a tradition that pre-dates the written word and yet is very familiar to the broadcaster.

A story is not simply a sequence of events. A story is a journey of memory and imagination. It is often conveyed in such a way so as to allow the audience to experience the journey from the mind's eye of one of the story characters. The audience, to a greater or lesser extent, experiences the sequence of affective states that are felt by the character over the course of the various relationships within the story.



(Continued on page 10)

(Continued from page 9)

This is a learning which the audience is then able to utilise as with any other experience.

The skill of a story-teller to convey the felt experience is therefore highly crucial. In live oral story-telling, the story-teller is able to read the affective state of the audience. The story-teller is thus able to improvise accordingly to bring the audience through the desired sequence of affective states. We could consider this as a form of 'closed-loop' story-telling.

With books, film, and other traditional broadcast media, the story is invariant, and by analogy would therefore be 'open-loop'.

The same is true for present-day 'interactive TV'. With the use of the affective technologies provided within the CALLAS project, we hope to continue our work towards closing the loop.

Our ultimate goal is a new form of broadcast media that would be a kind of 'interactive TV' where we close the loop in the story experience.

The general idea is that the home platform continuously senses the instantaneous affective states of the viewers in the home. In real-time, it locally modifies the story content - much the same as the bard did in the oral traditions – in a bid to carry the audience through the desired sequence of affective states.

Watch this space to see how we do it!

REFERENCES

[1] George Lakoff, Mark Johnson, *Metaphors We Live By*, Chicago University Press.

[2] Milton H Erickson, Sidney Rosen, *My Voice Will Go With You*, W.W. Norton

Mind or body: Which comes first in emotions?

By SYLWIA HYNIEWSKA

(TELECOM PARISTECH)

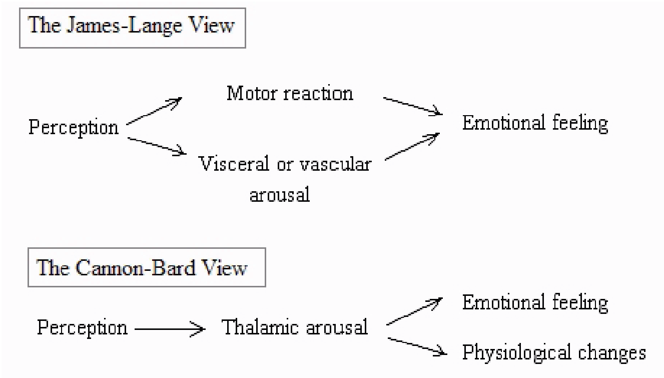
In an emotion elicitation, the body and the mind are implicated. Is it from our body reactions that we know we are experiencing an emotion or is it from our cognitive evaluations of the surrounding context?

The definition of emotions is still debated today by researchers. One controversial aspect of emotions lies in the body-mind distinction. Can we define emotions as specific patterns of physiological changes or rather as an evaluation of those changes in a specific context? In other words, what defines emotion, the body or the mind? Which comes first – the experience of emotion or physiological arousal?

We will look here on works by pioneers from that domain.

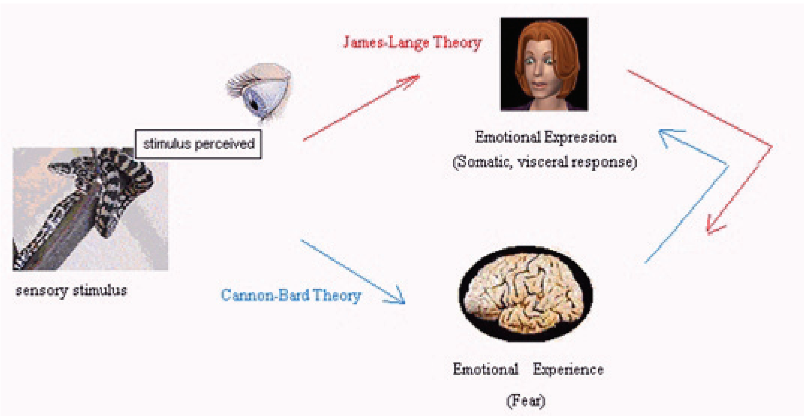
In his "Principles of psychology", James (1890), a psychologist and philosopher, who established the first laboratory of psychological research in the United States at Harvard, suggests that "bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur IS the emotion" (emphasis in original; James, 1890). According to this theory, the visceral and the motor reactions to a stimulus are prior to the felt emotion; more exactly the emotional experience is based on the physiological changes. Thus, "we feel sorry because we cry, angry because we strike, afraid because we tremble" (James, 1980).James emphasised that both the visceral ("gut") reactions, such as a churning stomach, and the overt bodily motor reactions, such as trembling or striking, are the basis of emotion (Carlson & Hatfield, 1991). Lange, a physiologist, had a similar perception of emotion, however he emphasised vascular changes, such as blood pressure changes, as central to emotion (Carlson & Hatfield, 1991). Thus, this approach has been named the James-Lange theory.

This idea that emotion is the perception of bodily changes has lead to the search for specific physiological patterns that would be specific of particular emotions. James-Lange theory states also that there are neither specific brain circuits nor regions involved in the experience of emotion (Davidson et al., 2000). The idea that there are no specialised emotion brain regions has been questioned by Cannon, a physiologist, and Bard, a psychologist.



James-Lange and Cannon-Bard differentiated perception of emotion

They argue that the physiological changes are non-specific. The viscera are relatively insensitive structures and therefore cannot provide the subtle and complex kinds of information needed to differentiate diverse emotional states. Very different emotional states are associated with identical visceral changes. According to Cannon, the physiological changes are also too slow to be accounted for the rapid emotional state changes (Davidson et al., 2000). The viscera not being a possible source of information for the emotion elicitation and differentiation, Cannon concluded the brain and more specifically the thalamus is the control centre for emotional behaviour. It is the activation of the thalamus that produces, according to Cannon, the emotional experience and the bodily changes associated with it.



James-Lange and Cannon-Bard comprehension of emotion: a different order in the elicitation of emotion

Cannon and Bard have demonstrated in their experiments that there are specific neural circuits involved in the expression and experience of emotion (Bard, 1928, 1929; Davidson et al., 2000). Emotions seem to be closely tied to the brain organisation.

Today this question of mind-body dichotomy driving emotion elicitation is still not answered. Nowadays researchers try to answer this issue by new technology in neuro-imagery such as fMRI (functional Magnetic Resonance Imagery).

REFERENCES

[1] Bard, P. (1928). *A diencephalic mechanism for the expression of rage with special reference to the sympathetic nervous system.* American Journal of Physiology, 84, 490-515.

[2] Bard, P. (1929). *The central representation of the sympathetic system.* Archives of Neurology and Psychiatry, 22, 230-246.

[3] Carlson & Hatfield, J. G., & Hatfield, E. (1991). *Psychology of emotion.* New York, NY: Harcourt Brace Jovanovich College Publishers

[4] Davidson, R. J. , Jackson, D. C., Kalin, N. H. (2000). *Emotion, plasticity, context and regulation: perspectives from affective neuroscience.* Psychological bulletin, 126, 6, 890-909.

[5] James, W. (1890). *Principles of psychology.* Vols. 1 and 2. New York: Dover Publications.

MusicKiosk in Santa Cecilia

By LORENZO SUTTON, ROBERTO GRISLEY and PAOLA PACETTI

(Accademia di Santa Cecilia)

The CALLAS MusicKiosk is an installation at the MUSA, Museum of Musical Instruments of the Accademia Nazionale di Santa Cecilia in Rome. It allows people to create musical stories based on their emotional expressions. The kiosk reads visitor emotion from video and audio inputs and uses this data to influence the animation and the music. CALLAS partner XIM has developed the kiosk using the CALLAS Framework.

At the end of October, XIM visited the museum in Rome to install the hardware for the kiosk. The kiosk was installed in the museum's Forum, a poly-functional room in the museum.

Two Apple Mac Mini computers are used to provide the kiosk functionality: one to handle input (a Windows machine running the CALLAS input components and framework) and one to handle the graphical and musical output (a Mac OS machine with Flash and Ableton Live). To harmonise with the museum's daily activities and exhibition needs the Mac OS machine can also be used as a regular museum kiosk machine by simply switching the user account. A large television screen, high-definition USB camera and a professional microphone were also installed. Remote access was enabled on the two computers to allow XIM to make software updates and capture log files for evaluation purposes from their London offices.

The Accademia Nazionale di Santa Cecilia is one of the most ancient musical institutions in the world. Founded in 1585, it has evolved into a modern academy and a music foundation with an international repute.

The Accademia Nazionale di Santa Cecilia is also a centre of excellence for outstanding music education. It hosts up to 250 concerts a year focussing on alternate seasons of symphonic and chamber music, and other themes.

In 2008 the Accademia inaugurated its Museum of Musical Instruments (the MUSA), where more than 100 instruments of its collection are on display, with examples dating back to 1500.

The museum is hosting the CALLAS MusicKiosk where it will be accessible to visitors over the coming months. XIM plans to update and enhance the kiosk with exciting new features that exploit the CALLAS Framework, and to use the kiosk as an evaluation tool to provide user feedback to the CALLAS consortium.

In Next Issue:

Interactive Opera for Teatro Massimo

Audio localization combined with face detection

GRETA

and much more, stay tuned!

(Continued from page 3)

Transferring CALLAS Multimodal Interfaces Technologies to the Market

There are currently many R&D institutions involved in the overall multimodal interfaces domain, as well as plenty of dynamic and innovation-led SMEs.

Unfortunately, due to the high costs necessary for developing and implementing a multimodal interactive application, so far the multimodal interface market has been very risky and uncertain.

There are significant barriers which need to be overcome for players aimed at developing multimodal applications. This is particularly true for **SMEs**.

They presently experience high entry costs and significant barriers for entering this market, such as the difficulty to undertake long-term basic research programmes, to access state-of-the-art technologies in a timely way and without incurring high costs, to utilise risky technologies in a really fragmented and niche market.

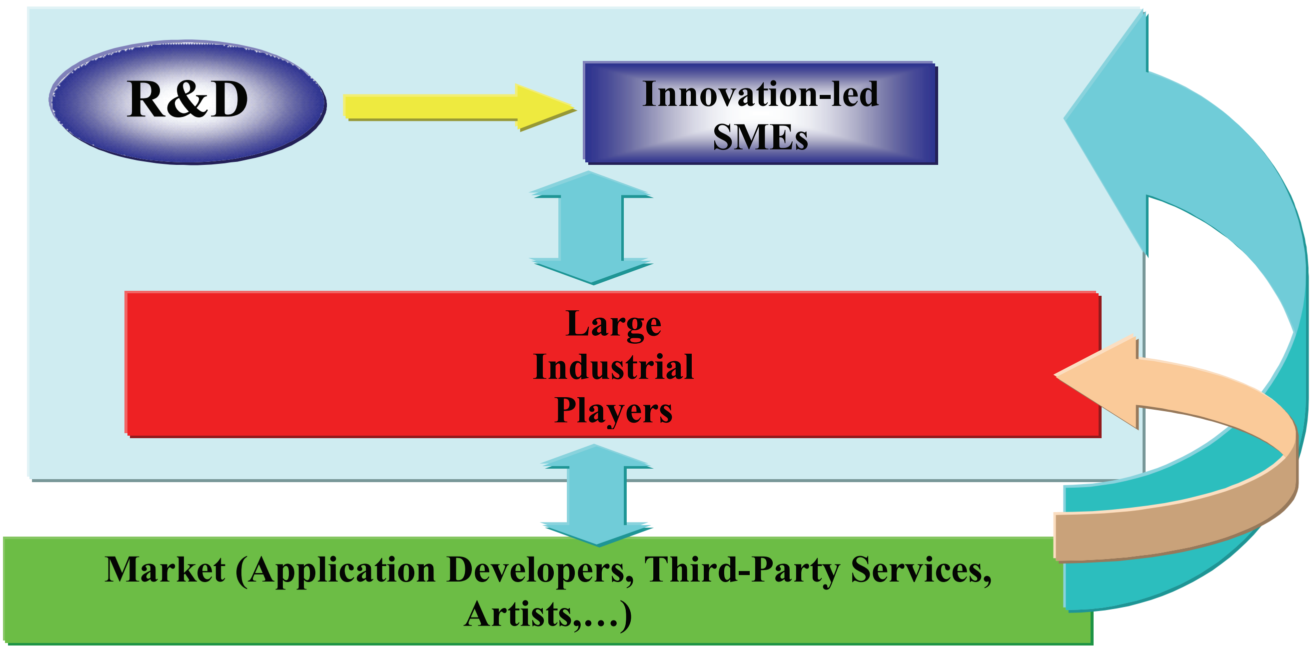
On the other hand a significant number of specialist SMEs have been included in the CALLAS Consortium (6). Their inclusion is a key factor in the project activities, because of their characteristics of flexibility and quick reaction to trends and movements such as the aggressive and innovative market of New Media requires.

Thanks to the cost savings deriving from the usage of the CALLAS technologies, the market risks for marketing Multimodal interfaces, even those emotion-sensitive, are becoming more acceptable, especially in the Digital Art and Entertainment domain.

More than this, in the CALLAS project an **"Incubator" Model** for transferring technology is going to be put in place, in which a Large IT & Service Provider (Engineering) will have a "hub" role for leveraging the research and the innovation to the market.

Thanks to the capacity of the large company to do even medium- and long-term investments, the SMEs will be given the opportunities to adopt and drive innovation to the market.

(Continued on page 14)



Business case N. 1

Cost Reduction in Studio Azzurro Live Performance

- *Description* : large re-use of CALLAS components and/ or sub-assemblies in the preparation and setting-up of a digital interactive opera.
- *Analysis of estimated costs of two real cases* : **Michael Nyman** opera "Facing Goya" and Neither opera in **Stuttgart Staatsopera.**
- *Key assumption* : CALLAS will reduce costs and time for designing and implementing **High-quality** multimodal applications for artistic mixed-reality digital installations, since powerful and suitable software engineering techniques will be used.
- *Conclusion* : A ***20%*** average cost reduction is expected in the production of a digital opera.

	Facing Goya - Nyman		Neither - Opernhaus Stuttgart	
	Expected Costs	Using CALLAS	Incurred Costs	Using CALLAS
Technical Inspection	5.000	5.000	3.000	3.000
Project - Direction & storyboard	40.000	40.000	35.000	35.000
Production: research, shooting, editing, interface development & software integration	79.500	41.500	125.000	73.000
Reharsal in Studio	22.000	22.000		
Reharsal in loco, Equipment, Set up, Première, Set down	61.000	60.000	63.000	62.000
Performance	32.000	32.000		
Foreseen Average Costs Per Installation (Number of installations per year 15)		1.500		1.500
Total	**239.500**	**202.000**	**226.000**	**173.000**
Cost Reduction %		**0,16**		**0,23**

Business case N. 2

Revenue Increase in Engineering Digital Entertainment Market

- *Description* : increase the market share for ***industry, telecommunication and services*** business line (currently 22%) and XALTIA (the multimedia company of the Engineering Group) through penetration in the digital entertainment market.
- *Key financial assumptions* : growth rate for Engineering in last years of 17%, as in the corporate accounts; analysis of revenue and market share in last three years ; extrapolation of data considering a changing in the market share, thanks to a larger penetration in the digital media one (despite a decrease of the financial one).
- *Conclusion* : an expected value creation for 21 millions of Euro after 2007 (*average)* in term of revenue increasing and of 1,06 millions (*average)* in term of net profit.

	2002	*2003*	*2004*	*2005 (Estim.)*	*2006 (Estim.)*	*2007 (Estim.)*	*2008 (Estim.)*	*2009 (Estim.)*
Finance	114	94	118	138	161	172	188	213
Public amministration	93	121	143	167	195	231	264	294
Industry, telecom & services	49	61	75	87	102	134	176	228
Total revenue and income	256	276	335	392	459	537	628	734
Market share								
Finance	44,40%	33,90%	35,10%	35,10%	35,10%	32,00%	30,00%	29,00%
Public amministration	36,40%	43,90%	42,60%	42,60%	42,60%	43,00%	42,00%	40,00%
Industry, telecom & services	19,20%	22,20%	22,30%	22,30%	22,30%	25,00%	28,00%	31,00%
Revenue increase with CALLAS		12	13	13	15	32	42	52
Revenue without CALLAS		12	13	13	15	17	20	24
Value creation		**0**	**0**	**0**	**0**	**14**	**21**	**28**
Value creation percentage		**0,0%**	**0,0%**	**0,0%**	**0,0%**	**2,7%**	**3,4%**	**3,8%**

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
- University of Paris 8 - France
- Scuola Normale Superiore di Pisa - Italy
- University of Reading - United Kingdom
- Fondazione Teatro Massimo - Italy
- Human Interface Technology Laboratory - New Zealand

Happy New Year to all CALLAS emotional IT fellows

The content of this Newsletter can be reprinted without permission provided that it is done in full

The CALLAS Editorial Team:

MASSIMO BERTONCINI, IRENE BUONAZIA, CATHERINE PELACHAUD, ANTONINA SCUDERI and DELPHINE TONGLET