Issue 01 Emotional IT Fellows

October 15,2008 Volume 1, Issue 1

A CALLAS Newsletter

The CALLAS Project:
Conveying
Affectiveness in
Leading-edge **L**iving
Adaptive **S**ystems

In this issue:

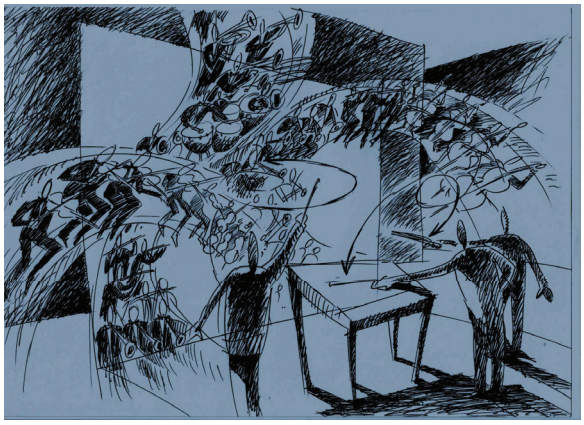
Visit us at:
www.callas-newmedia.eu

Contact us:
info@callas-newmedia.eu

Emotions and Art

by ANTONINA SCUDERI (NEXTURE Consulting srl)

Art and Technology have always proceeded side by side: new technological discoveries have very soon been experimented in the domain of arts, and, on the other hand, the creativity of artists suggested desiderata which technology later made possible.



This picture is a courtesy of Studio Azzurro

Research, debate and achievements in the domain of cooperation between arts , technologies and aesthetic applications of technologies are nowadays so huge that a library can hardly contain them. But what CALLAS is addressing is to come back to one of the very first aims of each art form: deal with emotions.

Emotions by the artist, emotions by who receives (through any sense) artworks. And, most of all, emotions spread from the interactive loop from creators to audience back to creator, in the pattern that, already in the Sixties, Umberto Eco called "open work".

In the current interactive world, art also has to be interactive if it wants to continue to be able to convey a message - as Arjen Mulder stated in the recent book "Interact or Die!" (awarded by MEDIA.ART.RESEARCH as best contribution in the field during the last Ars Electronica Festival in Linz).

More precisely "“art systems” as expressions of interactive art are emerging as a challenging area of convergence with affective computing. Understanding art is always emotional, the message that the artist communicates to each of us is absolutely unique and merges with our beliefs, mood and personal feelings, which blended together build the artistic experience.

The enrichment of artistic message through interaction implies the capacity to reflect the way the art is perceived and to modify the way it is proposed, tailoring the result to the spectator senses and engaging it directly as a co-author.

Exploration of these boundaries is pioneered by artists that love to get their hands dirty with IT programming and digital devices or

(Continued on page 2)

(Emotions and Art: continued from page 1)

by teams where artistic sensibility and technological competence are present and support each other.

Yet several examples have shown the power of virtual and augmented reality and multimodality to capture and render emotions. CALLAS wants to move a step forward, making life of artists easier, by setting the premises to escape from the burden of the overhead of the technology, offering new art professionals easy ways to engage the public and to reflect its emotional patterns through feedback loops.

This picture is a courtesy of Studio Azzurro

Follow us in this emotional trip to the future, where CALLAS artists, scientists and researchers are measuring themselves against the objective of preserving artistic freedom through a structured approach where complexity is hidden and all you will have to do is to configure and invoke a functionality you need.

Technology, please stay hidden!

by DIEGO ARNONE (ENGINEERING Ingegneria Informatica)

In the Art and Entertainment domain, the software applications aiming at collecting the emotional state of the audience and able to emotionally react have to face several quite complex issues. Human-to-human communication conveys emotions through both verbal and not verbal interaction: body movements, hand and head gestures, facial expressions. All they convey more than words, can express anger, fear, surprise, sadness and also interest, boredom, excitement.

How to allow a computer to understand, for instance, what the audience feels during a live performance? How to distinguish emotions by simply gathering tone of voices, facial expressions, eyes movements? Also, what a human says can somehow differ from what their hands are expressing, even giving a non conventional meaning to the uttered words.

In CALLAS we are investing in the expertise of academic institutions, ICT companies and artists in researching a solution both for the extraction and the synthesis of emotions in multimodal human-computer interaction.

Input components, innovative but not bound to be expensive or high-tech devices, are being developed to extract and remove ambiguity from emotions through complex algorithms. Powerful output components, able to create sounds, speech, augmented reality that reacts emotionally to the spectators are provided.

All these modules address the above-mentioned open issues and are packaged and made available on a rich repository, called the CALLAS Shelf.

Extracted emotions are processed to affect the behaviour of artistic applications at the output side. Fusion and interpretation of emotions are required and CALLAS encapsulates the related algorithms in (often ad hoc) components, always addressing the requirement of reusability.

But due to their peculiarities, the CALLAS components are often coded in different programming languages and run on different operating systems. So, an integration issue emerges and a Software Engineering solution is required.

CALLAS proposes a "glue", a Framework, whose implementation allows an easy aggregation of really heterogeneous software modules by providing a proper set of APIs for a semantic intercommunication. But it's not enough. One more tessera in the CALLAS mosaic: a graphical interface designed for artists not confident with technology and lacking in computer programming.

So, we are developing an easy-to-use Integrated Development Environment (the CALLAS Framework IDE) allowing a designer to aggregate different components in a multimodal application able to gather emotions of the audience and to emotionally respond. Complex theories are encapsulated and technology is made invisible.

Behind human emotional activity

by STEFANO ROVEDA
(Studio Azzurro)

The scope of our research in CALLAS is to use advanced interactive technology to elicit emotion activity on humans.

This could mean to solve problems such as:

- Finding solutions for a real-time (or near to real-time) procedure to create corpora continuously during normal module's execution, to create a large, context-based and subject-specific database of expressions.
- Design tools dealing with large databases for further definition of events to be classified and recognized dynamically, introducing more definitions while operating.
- Move from relative to absolute approach, defining "normal" or released position on each subject feature (posture, voice, face mimic,) as a reference point, without training or manual intervention.
- Extract relative movements for a single feature (e.g. mouth movement will be influenced by head pose and body moves).
- Compare the "shape" of each single gesture or mimic element and compare it with a library of similar gestures of same subject or all subjects to extract qualities parameters for that specific gesture

Fields of activities include interactive installations for public spaces and digital theatre.

We observe during the first field trials human behaviour while experiencing highly interactive environments as well as simpler technologies. Different levels of complexity can be observed due to external conditions that influence behaviours and due to the number of people involved.

As an example, during interviews with lighting set all around his seat, the subject will limit body movements to head and arms only, rarely moving back and forth on the seat. Having the troupe all in line behind the camera and white panels all around the unique subject attention point was the camera lens / eyes of the director. That condition will enhance face mimic and head pose during dialogue.

When a group is sharing an interactive space, different activities overlap each other. Space design and interactive rules may vary the body movements of the subjects. Special events in the interactive plot may cause synchronous reactions by the subjects. Moreover, elements of the group may communicate with each other, beyond the digital environment stimulation.

As conclusion of different field trials, we can say that we are dealing with very complex phenomena where different activities are overlapped and/or melted with each other. Moreover, humans are very well trained to understand and perform emotions and tools will be expected to have similar performance. I gathered from the scientific literature some basic concepts on that specific field that could be helpful to the researchers of the consortium in order to evolve their tools.

Human activities

I believe that first of all we have to distinguish the typologies of human activities in which emotion takes part and we shall consider how these activities interact, playing a main role in the communication process.

As cognitive science states, we can distinguish three categories of human activities: the sense motor one, the emotional and the rational. Looking back at the evolution of the species as well as the growth of an individual, we can gather that the first activity that evolves is the sense motor one (form the birth to the first tooth for human beings). The correlation between the signals sensed and the muscle activity is certainly common to all the animals. We can distinguish the reflexes among all the sense motor activities. Here the motor activity reacts immediately without any rational involvement. For example, we immediately close our eyelid to protect our eye from a foreign body.

Unwilling movements are another motor activity. For example, the beats of our eyelids indicate our emotional condition. So when we're relaxed, the beats are only to make wet our orb, while an increased frequency indicates stress and emotional disease.

Other motor activities are part of behavioural schemes that involve many micro-actions according to learned sequences. A scheme, once learned, is considered as an inseparable unit, though it can be expressed in different ways.

These ways are pointers of different emotional conditions. When we close our eyes with relaxed eyelids and in the meanwhile we raise our head, we perform a behavioural scheme widespread in Southern Italy meaning "No" or "I

(Continued on page 4)

(Behind human emotional activity: Continued from page 3)

don't know" or "I don't want to reply". The variants include an aggressive attitude, conspiracy of silence, a suggestion of trust and the joke. Moreover, the motor activity can also be meant.

Reason acts and coordinates muscles: we can close our eyes if someone we trust asks us to do it and if we are under proper conditions.

The act of closing our eyes can be characterized by many different synchronous signals overruling its meaning. For example, when we squeeze our eyes, we can assign different meanings to the act in relation with how long we keep our eyes closed, how deep we squeeze, or other mimics accompanying the eye's. We're talking about micro-signals forming a general appearance that man has learned to distinguish during his millenary evolution.

The typology that characterizes every individual interferes with these movements and micro-signals making things more complicated. For example, puffed eyes or extra-long eyelashes can modify the morphology of the act and of the related signals. It means that expressions may be considered in a relative scale and there is no way to determine absolute models of behaviour satisfying all humans and their different cultural contexts.

Moreover, the perception of a micro-act, like the closing of your eyes, has to be read in relation with the context, the individual and considering the other signals and acts at once or anyway correlated over time.

Considering the complexity and the superimposition of the motivations which generate every single micro-act, further work should be done to disarticulate the complexity in lower level events adding quality parameters (or "shape analysis") to single expression acts, exploring the possibility to comprehend behaviour at an higher level through data fusion techniques.

Emotional activity

Always considering the parallelism between ontogenesis and phylogenesis, the second period of the life of a man is ruled by the emotions.

This period starts from teething (when the individual is about 7 years old) and goes to puberty (when the individual is 14 years old).

Many scientists think that this is the period that is mostly attacked by the contemporary culture.

The abuse of strong emotions from the young age, caused by technologies of representation (cinema and television), makes the emotional life of the contemporary man altered, somehow less intense due to continuous shocks. Contemporary western culture is strongly rational-based leaving little space for motion and emotion.

The tendency to over stimulate the intellect oppresses the sense motor and the emotional growth of the individual.

The three activities (sense motor, emotional and rational) live and evolve interweaved.

About the three seven-year periods we talk about "dominated period": one activity dominates another. For example, language seems to be the first rational activity. It appears when we are about 1 year old much before the time when rational activity will become predominant. On the other hand, sense motor schemes may be learned by grownups (e.g. driving).

Emotional activity orients and motivates motion and rational activity, as well as shaping the other two activities. Due to an emotional state my rational activity and my gestures and mimic will operate accordingly. The quality of my gestures will show up my emotional state. Human emotional activity, according to recent studies, is responsible for decisions.

While rationality helps us in analysing problems, emotion is responsible for the synthesis that leads us to choose one strategy rather than another one. It is well known that for some individuals it is hard to make a decision, as they swing between the different solutions without making a choice.

Human activity originates and expresses itself throughout each part of the body. The western scientific community localizes each activity in a specific area of the body, as happens with knowledge expressed through the spoken language.

Cognitive activities are more localised in the head while emotion seems to be more related to the womb and the breast. The motor activities are instead spread all over our body. From the Chinese medicine perspective, these statements seem to be reductive because the body is a unique complex and interrelated organ.

From this point of view, all the body is involved in every expressed or unexpressed activity, as in a reverberation and feedback process. According to this holistic approach, it is not possible to consider the emotional and

other activities as separated from each other.

The existence of conflicting signals tells us that the activities are not always on the same wavelength and synchronous. It means that we can say a sentence and contradict it with our mimics. Sometimes things that happen can bring to a change in emotional condition, influencing the following actions. Some emotions like crying or laughing are common to every human being. Some others are related to the cultural context and it may happen that the same expressions have different or opposite meanings among different cultures.

Anyway, the same basic or innate emotions will be performed through the filter of the cultural knowledge of the subject, and may vary the meaning due to subtle differences while performing.

In his neuro-cultural theory, Elkman believes that the cultural factor is responsible for the rules dominating the emotional expression.

Aside the cultural domain where the subject is grown, each subject has its individual and unique features also in the body language, voice and mimic.

All activities we may observe in a subject should be related to its typology.

Physiognomic and temperament aspects should be considered to extract relative aspects of expressions.

The Tree of life

By STEVE GILROY
(University of Teesside)

One of the aims of CALLAS is to facilitate the development of compelling digital artworks that are enhanced by the incorporation of affective capabilities. Interactive Digital Arts installations rely on increasingly sophisticated input modalities, also taking advantage of the integration of sensors in Virtual Reality (VR) and Augmented Reality (AR) systems. In addition to supporting more natural interaction, this creates an opportunity for analyzing audience reactions using recent developments in affective computing, such as those leveraged in CALLAS.

This resulted in the production of an artistic brief supporting the investigation of affective interfaces in AR Art installations, the Emotional Tree, or e-Tree.

In this installation, a virtual tree is animated to reflect its growth, which is influenced by the perceived emotional responses of the spectators, as interpreted in a dimensional model of affect.

The tree exhibits sophisticated growth patterns in terms of branching, speed of growth, branches orientations and branches motion.

Furthermore, growth cycles are not monotonic, and parts of the tree can fade as well as a consequence of negative responses, still resuming growth when negative influence cease. This creates the basis for a rich interactive experience, in which the installation induces a feedback loop, its reactions to perceived user attitudes eliciting new user's responses in return.

In this context, emotions are the content of a metaphorical dialogue between the installation and the spectator.

The use of AR offers several advantages for this type of installation: it preserves the natural physical space of the installation, it imposes no limit, in principle, on the number of users who can participate, and the visual display does not interfere with the sensors used to capture spectator's response.

(Continued on page 6)

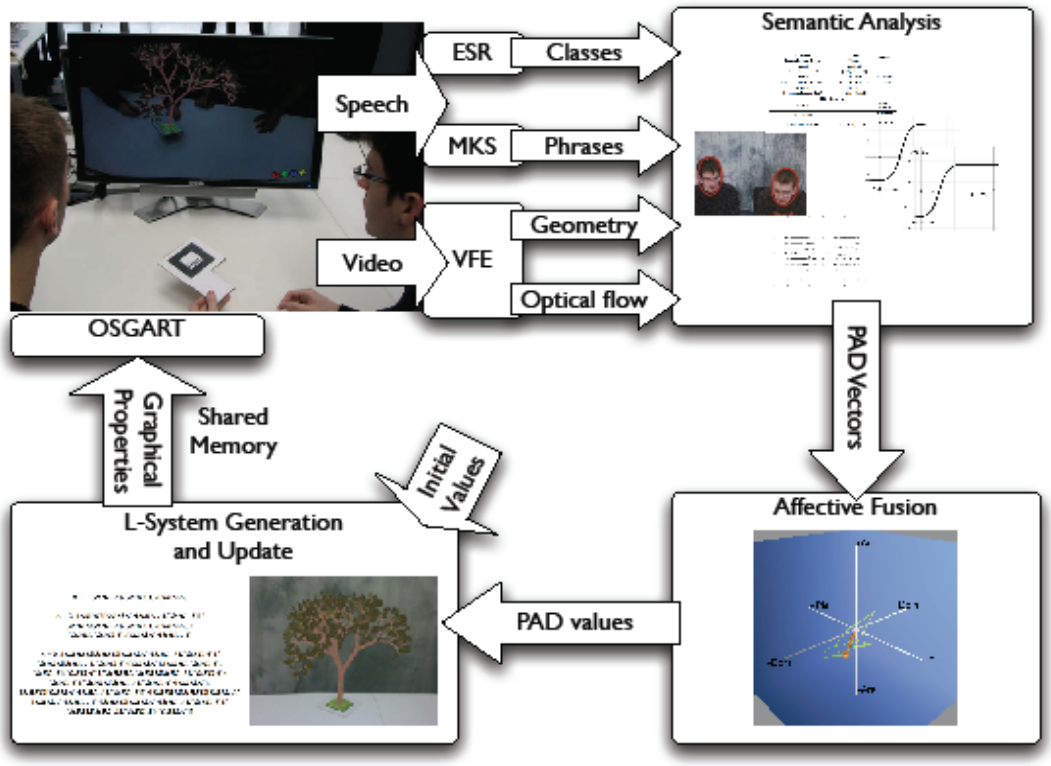
(The Tree of Life: continued from page 5)

The e-Tree installation incorporates a comprehensive affective model of user experience, beyond the recognition of a limited set of discrete emotions, which is based on the Pleasure-Arousal-Dominance (PAD) model [1]. We have adopted the PAD model as an affective representation of emotion within the system largely because, as a dimensional model, it was well adapted to the exploration of affective input for which emotional categories were not readily available. E-Tree captures audience reactions using multimodal affective input components provided by various members of the CALLAS project. All of the affective information captured from user responses is combined to give a single overall representation of the current affective state within the PAD model, also taking into account context and history, through the previous values of this affective state.

For the visual appearance of e-Tree, the naturalistic tree structure is defined by an L-system, its growth is governed by rules which are modulated by the real-time interpretation of the spectators affective state. The graphical implementation makes use of a marker-driven AR system that displays the tree situated in the environment of the participants. This is implemented using OSGART[2], which extends the OpenSceneGraph[3] 3D graphical framework to support AR applications, utilising ARToolkit [4] for marker detection and tracking. Participants can directly interact with the installation by manipulating the markers and seeing the tree react appropriately on a magic mirror video display, either a large (30" or larger) monitor or projection onto a wall to incorporate group participation.

Speech and video input is processed by three software components, Emotional Speech Recognition (ESR), Multi-Keyword Spotting (MKS) and Video Feature Extraction (VFE). These components identify relevant multimodal features of the input that will be mapped to appropriate affective and aesthetic terms. ESR produces affective classification of speech utterances, while MKS identifies specific keywords and phrases. VFE detects and tracks faces in the video frame and produces geometry information of each detect face for each frame of video. In addition, it calculates the optical flow of successive frames. This is used as a characterisation of the movement of spectators.

Analysis of affective features may involve transformation of feature properties (e.g., calculating the area of a face, identification of underlying semantic concepts of spoken phrases) or temporal analysis (e.g., calculating average optical flow over several frames, tracking faces).



Affective interpretation is achieved by mapping properties of feature analysis to a vector of PAD values representing a point in the emotional space of the PAD model. The affective output of all the components is combined over time in an affective multimodal fusion process where PAD representations for each input modality are weighted and summed to give a target composite value that represents the perceived current affective reaction. This is then combined with the current mood of the installation to give an overall representation of the new affective state of the installation.

These PAD values are sent to the L-system generation component that maps each dimension of the PAD emotional model to parameters controlling the growth and appearance of the tree. The thickness of branches, size and colour of leaves and droop of the tree are all directly dependent on the current values, and reflect the immediate affective state of the installation.

The growth and branching of the tree serve to record a history of the user experience, as it changes over time. The emotional aspect of the e-Tree is thus that it grows in a way that reflects its perception of the user response.

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

[2] J. Looser, R. Grasset, H. Seichter, M. Billinghurst. 2006. OSGART - A Pragmatic Approach to MR. In Industrial Workshop at ISMAR 2006, Santa Barbara, California, USA, October, 2006.

[3] H. Kato, M. Billinghurst. 1999. Marker Tracking and HMD Calibration for a Video-Based Augmented Reality Conferencing System. Proceedings of the Second IEEE and ACM International Workshop on Augmented Reality (IWAR 1999), San Francisco, California, USA, pp.85, October 1999.

[4] Burns, D., Osfield, R.. 2004. Open Scene Graph. Proceedings of the IEEE Virtual Reality 2004 (VR'04), p. 265.

Let the children play

by LAURENCE PEARCE
(XIM Ltd)

The CALLAS MusickKiosk is an installation that allows young people and adults to create musical stories based on their emotional expressions.

The goal of Musickiosk has been to build a publicly accessible installation that allows users of all ages to enjoy the creation of music and to explore how emotional expressions can be interpreted into a musical performance.

The kiosk portrays a story that is related to the historical musical instrument library of Santa Cecilia in Rome, where the kiosk is to be hosted. A young musician is lost, trying to

find the orchestra who are about to perform using historical instruments. As he hunts for the concert, he enters different recital rooms, each containing groups of musicians. As the user watches and comments on these rooms, the boy, the musicians playing in the room and the music being produced all respond to the user's perceived emotional expression.

Finally the boy finds the concert room, enters and takes his place.

This interaction uses two inputs: user speech and facial expressions taken from a microphone and webcam respectively.

These are interpreted into emotional representations, which

(Continued on page 8)

(Let the children play: Continued from page 7)

are used to affect the 2D animation within the story.

At the same time, different layered musical phrases are queued by means of a MIDI gateway and an audio engine (Ableton Live), which in turn provides the timebase for the animation so that the music rooms in the animation are synchronised with the music.

The first functional prototype of MusicKiosk has been completed, and following further refinement and user validation, will be presented at Santa Cecilia Auditorium later this year.

Electro Emotion

By LASSI LIIKKANEN (Helsinki Institute for Information Technology)

A crucial aspect of emotional experiences and their displays is their tremendous context dependence.

As a CALLAS showcase intended to address public spaces, our concept Electro Emotion takes the contextual reliance as a challenge. In previous research people have collected and analyzed emotional displays in largely artificial settings, with the exception of professional broadcast media corpora.

In Electro Emotion, we place a novel installation into a public place to attract people to explore their expressive potential in a shared social environment. Our application embeds ideas of cross modal interaction. In this approach we build on expressive features of sound and human gesturing which are transformed into a new kind of interactive experience that not only makes visible what you say but also makes audible how you move!

The interactive installation Electro Emotion is a tool to investigate performative interaction that utilizes several state-of-the-art components made by CALLAS Shelf developers.

The showcase itself is developed by a multidisciplinary design team working at the Helsinki Institute for Information Technology HIIT in Finland.

The capital area of Finland will also witness the first real

installation of the system during 2009.

Once operational in the wild, we expect that will be able to gather a heterogeneous and colourful corpus on multimodal, real-world emotional displays.

Our goal is not only to explore the dimensions of emotional displays in public environments but also inform the developers of public information systems about the opportunities of retrieving emotion-related information using new kind of interaction techniques.

Welcome to the CALLAS Club

by IRENE BUONAZIA
(Scuola Normale Superiore di Pisa)

Are you an artist working with new media to create thrilling experiences of aesthetic communication?

Or are you a developer willing to participate in the most innovative trend of research in human-computer interfaces?

Do you work in a company developing software or devices for the new media world, from audio & video interactive production, to Augmented Reality environments?

Are you organizing events, in theatres, museums, public spaces, exhibitions, cities, where people have to play the main role?

Then you should join C³ – the CALLAS Community Club!

If you believe interactive interfaces have to address all the channels of human communication – emotions included! -, in the CALLAS Community Club you will find a proactive network of researchers and creative users, have first access to innovative technical frameworks, find partners for new challenging creative projects and be involved in CALLAS events and training sessions.

This picture is a courtesy of Studio Azzurro

Registering at http://www.callas-newmedia.eu/index.php/the-cs-the-callas-community-club.html, you can introduce yourself in the C3 dedicated web site: describe your work and projects, share your knowledge, questions and interests, became part of a leading-edge research project.

In Next Issue:

……..about active audience participation to performances;

…… about public broadcasters as modern story-tellers;

……..News from the C³;

……...CALLAS at ICT 2008;

and much more, stay tuned!

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

The CALLAS Editorial Team:

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