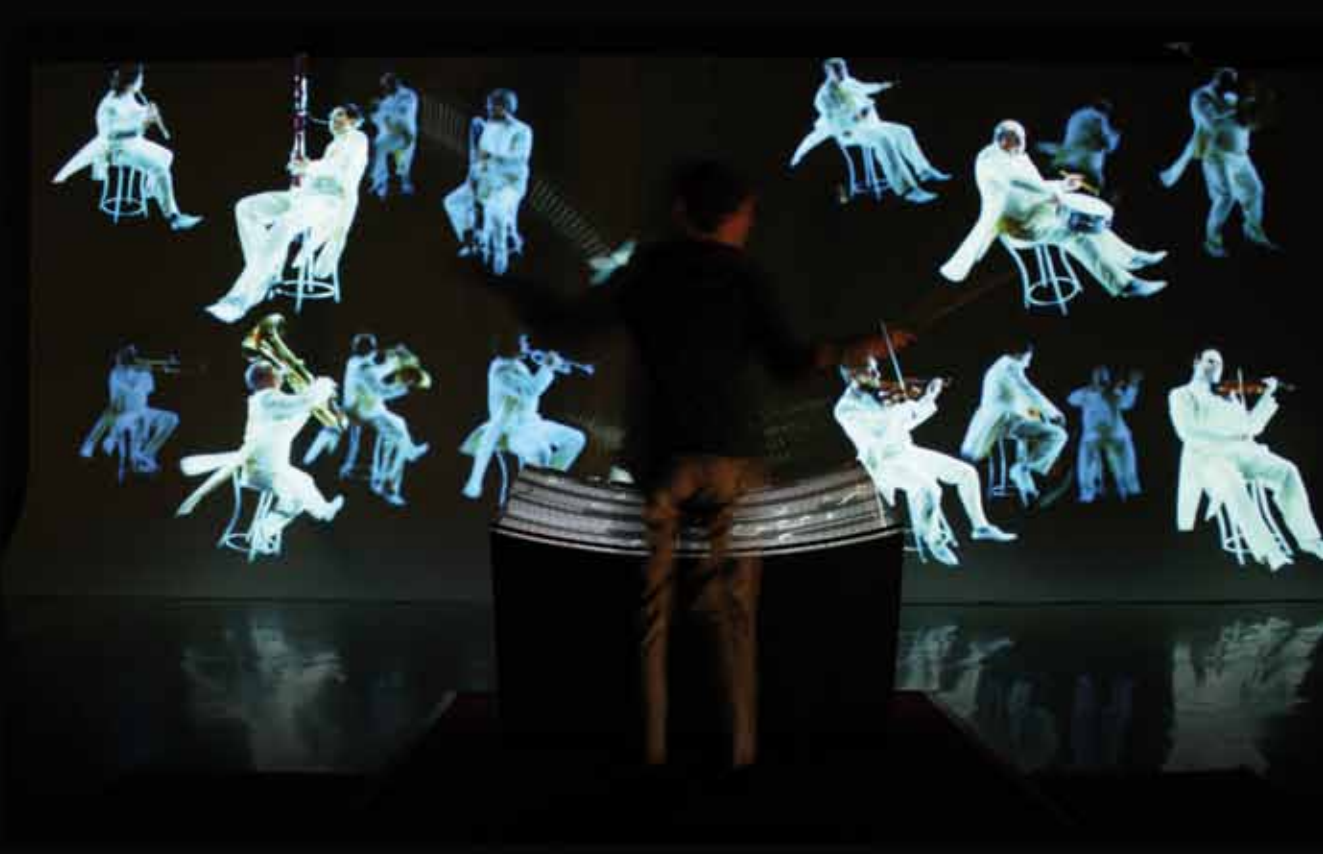


callas

conveying affectiveness in leading edge living adaptive systems

A challenging adventure…
…still going on

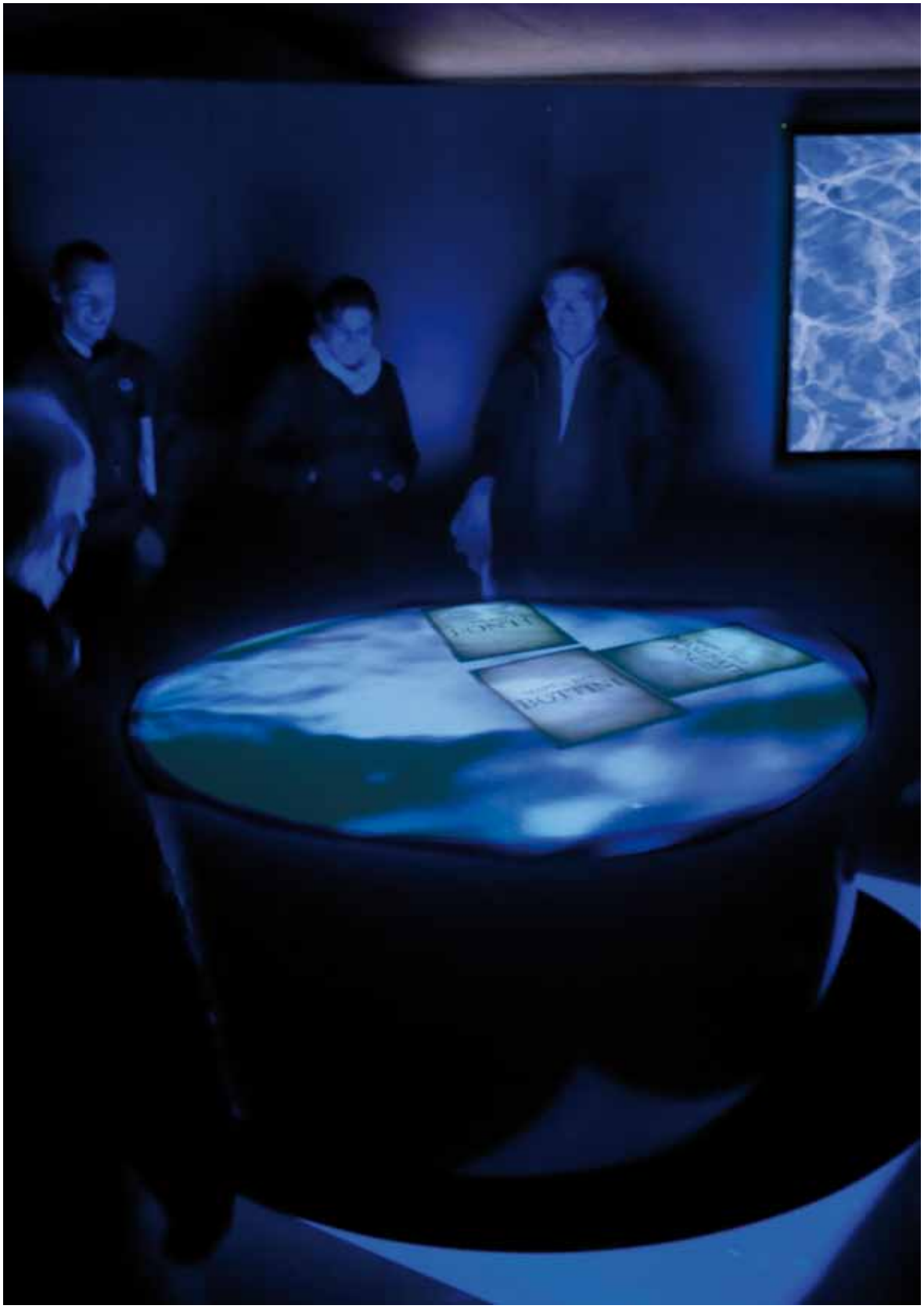
Index

CALLAS Concept

The CALLAS project has developed a totally new concept for human-computer interfaces, in which "multimodality", "emotions" and "affectiveness" are key aspects for enriching the naturalness of "human-to-human" and "human-to-machine" interaction and communication.

The originality of the Callas approach lies in:

* Bringing the software engineering approach centre stage in the multimodal affective interfaces, allowing developers in this field to reuse and integrate existing components.
* Providing a new real-time semantic approach to affective multimodal fusion.
* Increasing the robustness and further validating a significant set of software components for "detecting multimodal input", "rendering multimodal output" against unimodal and, above all, "synchronizing multimodal corpora".
* Developing multimodal interactive applications in Digital Media domain as "proof-of-concept" of the proposed integrated approach.
* Developing an innovative "hub"-based sustainability model aimed at fostering a significant market penetration of Affective Multimodal Interfaces.



Summary of CALLAS Project

The CALLAS three years and half research project has been the flagship initiative of European Commission on the Affective Multimodal Interfaces for Digital Media, ranging from Art to Culture and Environment.

Human beings naturally communicate in a multimodal way combing different "senses" like gesture, movements, speech, nonverbal expressions. Emotions and affectiveness play a key role in enriching the naturalness of human-human and human-machines communication and interaction.

However traditional human machine interaction is normally based on common devices such as keyboard, mouse, which are not designed to allow computing systems to understand and express emotions.

Hence computers are not able to naturally communicate with humans if they do not have the ability of emotion processing and rendering.

Affective Multimodal Interfaces are at the crossroad of Affective Computing and Intelligent Interaction research fields. They are expected to allow computers and artificial systems to handle emotions and affect, and, accordingly, to make it available a more natural and user-centric human-computer interaction paradigm.

Unfortunately despite a lot of significant advancements have been achieved in the Affective Multimodal Interfaces domain, much work still need to be done in order to bridge the gap among computational systems and human emotions.

The CALLAS project delivered significant outcomes in the Affective Multimodal Interfaces field in all the three working areas in which has been structured, namely the "Shelf", the "Framework", the "Showcases". Thanks to its scientific advancements and technology outcomes, the CALLAS project strongly contributed to increase the maturity of Affective Multimedia Interaction technologies in the tailored fields of Digital Media, which, by the way, represent a promising application domain for these kind of technologies, since here the interaction of the user with technology is traditionally very rich from the emotional point of view.

Main significant yet original project outcomes are:

* **An Open Source plug-and-play environment of Reusable Software Components**

  For the rapid prototyping of multimodal applications, reducing complexity for programmers, artists and practitioners (the CALLAS IDE).

* **A Library of Software Components: "the CALLAS Shelf"**

  For the emotional processing of single and multimodal inputs (speech, gesture, gaze…).

* **An Innovative Approach to late (or Semantic) Affective Multimodal Fusion**

  Built on the top of the Pleasure-Arousal-Dominance (PAD) model and related mapping from recognised input, which allow input components to be adaptively combined and fused together in order to detect the user's emotional state in real time.

* **A Broad Range of Affective Multimodal Applications**

  Generating an affective output dynamically linked to the detected user state, which have been developed on the top of the CALLAS IDE and Framework, as proof of concept of the effectiveness of the CALLAS approach.

* **An Innovative "Hub" based Model for Technology Transfer.**

  Of the research lab prototypes into mature yet sustainable applications in Digital Media, and of the developed innovation toward the market, primarily represented by innovation-led SME for laying the foundations for effectively reusing CALLAS technologies in different application fields, ranging from car driver affective status monitoring, to adaptive personalised -learning systems, assistive technologies for social inclusion of disadvantaged people (elderly, disabled), to the gaming systems and related consoles.

An integrated approach

CALLAS is based on a three-layered working areas.

The CALLAS Shelf: a library of multimodal components developed and made available from the Consortium partners.

The CALLAS Framework: a plug-in architecture for the interoperability between the shelf components, allowing multimodal applications developers to combine them.

The CALLAS Showcase: experimental applications developed using the CALLAS Framework to demonstrate how successful the CALLAS technology is in conveying affectiveness and augmenting people experience in different interactive spaces.

Callas shelf

The CALLAS Shelf is a set of software components aimed at capturing basic users' emotions during multimodal interaction, process them and providing a real-time emotionally-enriched response to the interacting user.

The CALLAS Shelf was meant as an open workbench in which, thanks to a specific quality control methodology developed within the CALLAS project, new mature, robust and effective off-the-shelf third party components could be included into the Shelf. In particular a continuous technology watch and market scouting process, led by the consortium academic and industrial partners, has been implemented aiming to be constantly updated and aligned with the current state-of-the-art of the multimodal components existing in the market or released by leading-edge research centres, and, eventually to include suitably emerging interesting components. For example a Virtual Baton component from Studio Azzurro partner was successfully integrated and made available to the CALLAS Shelf and Framework, even if not specifically included into the original Shelf.

Here below there is a list of the most significant affective multimodal input and output components which have been made available to the CALLAS Shelf.

CALLAS inputs shelf

Multi-Keyword Spotting (MKS)

This component developed by University of Mons partner, aims at recognizing spoken utterances inside a predefined list of expressions, called "grammar". The component is speaker-independent: i.e., it has already been trained with a large number of speakers to be able to deal with different voices and does not need the user to train the recognizer on his own voice to be efficient. The reason for using a grammar is to limit the number of recognition errors while concentrating on the most relevant expressions for the application. The drawback is a limitation of the recognizable expressions: only the utterances defined in the grammar list can be outputted by the recognizer.

The primary objective of this Shelf component is to recognize spoken utterances, not emotions themselves. Recognized utterances can be used for various purposes in applications, the most obvious one being to serve as a command/instruction to drive the application into a new state. Nevertheless, spoken utterances convey emotional content, both in the choices of words as in the way to pronounce them.

The affective content conveyed by word/utterances themselves can be inferred after using MKS, thanks to, for example, PAD mappings of the grammar utterances, as it has been done in CALLAS applications. The component has been tested under a variety of different conditions. Test performed for CALLAS, using the CALLAS-developed grammars in shared-office conditions with a headset microphone always showed a recognition rate around 90%.

Emotional Speech Recognition based on Acoustic features: "EmoVoice"

Developed by University of Augsburg partner, is a comprehensive toolbox for speech emotion recognition which makes use only of acoustic features (no word features), to be independent of other components. However it can be combined with a speech recogniser and a text-emotion recognition component to exploit also the meaning of words.

EmoVoice includes all essential tasks of speech emotion recognition, which are audio segmentation, feature extraction and classification. For all three tasks, systematic analyses have been carried out in order to identify suitable methodology that is capable to work in real-time. Audio segmentation means finding appropriate units of emotions in the speech signal coming in continuously from the microphone. While also other units can be used, the most suitable units for real-time emotion recognition were found to be based on voice activity detection that is segments with no break of acoustically measured voice activity within. During feature extraction, relevant acoustic features for emotions are calculated. In EmoVoice, these features are based on pitch, energy, spectrum, cepstrum, duration and voice quality and are generated by applying statistical functions such as mean, maximum etc. to these acoustic measures. With this method, a high number of features

is generated from which the most relevant ones for a given task or application can be automatically selected. The last step, classification, maps a feature vector onto an emotion class. Currently, two classifiers are integrated into EmoVoice: Naïve Bayes and Support Vector Machines (SVM). While the former is faster and very robust on natural emotions, the latter has been reported in the literature to achieve higher accuracies. EmoVoice was so far tested in several respects. On generally available databases, offline recognition rates of more than 80 % could be achieved for 7 acted emotions (Berlin Database of Emotional Speech) and more than 50 % for 4 natural emotions (FAU Aibo Emotion).

Emotional Speech Recognition based on Linguistic features: "EmoText"

Developed by University of Augsburg partner, makes use of linguistic information to recognize emotions from text by statistical or semantic analysis. The **statistical analyzer** of EmoText makes use of lexical, stylometric, deictic and grammatical features. The lexical feature sets hold frequency counts of occurrences in the current turn (rsp. the 7 previous ones) of the words in the lexicon. The sets differ in their lexicons: all words occurring in the corpus (2033) or subsets of different sizes of the most frequent words. In the stylometric feature sets, there are 730 features from letters, word length, word digrams, standard deviation of word length, and sentence lengths in words, see (Forsyth & Holmes, 1996); (Kjell, 1994); (Ramyaa & Rasheed, 2004). The deictic feature sets consists of 530 features from the frequency of demonstrative determiners, demonstrative pronouns, time references, place references, third person forms, as well as stopwords. The grammatical features refer to specific linguistic patterns that have been characteristic of the expression of emotions, such as interjections, exclamations or repetitions. The **semantic analyzer** of EmoText makes use of the Stanford parser and the SPIN parser.

First the Stanford parser a probabilistic natural-language engine featuring an order-dependent parser, is applied to dynamically tag and lemmatize words and to detect the syntactic structure of input sentences. Then a semantic parser for spoken dialogue systems (SPIN) parses this representation into emotion categories. SPIN is a rule-based framework that uses order-independent rules for detecting predefined patterns of words in the analyzed texts. In addition, EmoText includes a hybrid approach to emotion recognition in order to take advantage of the synergetic effect of statistical and semantic approaches. EmoText was extensively tested with a variety of corpus including both longer texts and shorter dialogue utterances, overperforming the majority of these tests in several test conditions.

The Audio Analysis Component (AAC)

Developed by VTT partner, aims for detection of different audio cues, which could be significant for multimodal fusion in order to detect the audience state. This component takes in live audio input via microphones connected to a personal computer and outputs the class

decision and frame power of an audio signal. The component currently classifies the audio stream into 9 different sound classes: silence, speech, music, variable and constant noise, laughing, whistling, applauding and clapping (applauding by a few persons only). The rational behind this component is that emotion recognition related to audio is typically done directly from speech signal using prosodic features as in EmoVoice.

Other nonlinguistic vocalization, for example laughs and cries, can give emotional cues of behaviour. Although these outbursts can be present in many emotional states; i.e. laughter can be caused by humour, anger or anxiety.

When we consider CALLAS environment (digital media and artistic installations) we might draw simpler conclusions of what kind of human sounds will imply to certain cues from affective state of the audience, i.e. applause and whistling is approval, booing and whistling implies to negative feedback.

Hidden Markov Models have been used as classification model. With this regard one important requirement for live audio analysis is a short response time. Computational costs should be as low as possible. HMM-classifier is a trade of between performance and the computational cost. Fusing simple rule based classified with HMM-classifier improves performance without increasing the response time considerably.

Video analysis is a widely researched area and it is used in many application fields, such as surveillance, multimedia content analysis and medical imaging. In multimodal interaction the main areas of video analysis have been concentrating in gesture recognition, pose/gaze recognition, head tracking and additionally with the latter, facial expression recognition.

Video Feature Extraction (VFE)

This component developed by VTT partner aims at extracting information from people participating in an event or installation, providing input for further affective analysis. The approach is to use face detection for counting and tracking people, as well as orientation of the head for head movement information. Quantitative movement analysis can also present the level of interest or enthusiasm of the audience. VFE component finds and tracks human faces, and calculates the optical flow from a live video feed from a webcam. The component is divided into several parallel threads. Face detection is performed by a Viola-Jones boosted cascaded classifier trained to spot forward-looking faces. The detection is performed in two phases using two separate Haar cascades. The first classifier scans the entire frame using a Haar cascade provided in Intel's OpenCV computer vision library, which also implements the Viola-Jones detector. The second classifier is trained using a frontal face database assembled from two databases of facial images: The *Database of Faces* by AT&T and *The Color FERET Database* by NIST. After the first detector produces a set of what it thinks are faces, the second classifier is used to sweep these locations and their immediate surroundings again for confirmation. This

way most of the – already few – false positives from the first round can be eliminated without noticeably affecting the real positives, and as the second sweep is performed only in limited areas of the frame, the impact on the component's performance is not too harsh, as opposed to performing the analysis on complete frames with both cascades.

Gesture expressivity analysis and synthesis has been increasingly attracting attention from research areas such as affective computing, pattern recognition and psychological and behavioural studies. As a result the corresponding research community has been very active drifting along the way the technology industry in novel either theoretical or application oriented fields. Gesture expressivity has been well studied from a synthesis point of view but rarely has been confronted from the analysis viewpoint. Additionally usually manual annotation was incorporated to evaluate perceived gesture expressivity. Automatic extraction of expressivity features was rarely encountered and usually featured full body movement. In the majority of cases features are extracted to be used as input to emotion classifiers and are not studied as expressivity features per se. Although gesture expressivity features applicability to emotion recognition is unquestionable, their importance goes far beyond mere input features. Moreover full body movement is not the case for many HCI scenarios and the absence of gesture expressivity features standardization for hand gestures has become a necessity in the wide research area of affective computing. The formal definition of gesture expressivity features was the missing link for many architectures involving expressivity. This formal definition of gesture expressivity features was the missing link for many architectures involving expressivity.

Video-based Gesture Expressivity Features Extraction

This component, developed by ICCS partner, is used to detect and track the user's hands and extract and transmit expressivity features' values. Hand detection is robust, since skin and motion information are fused while the expressivity features extraction algorithm has been proven to be accurate by an analysis-synthesis scenario and user evaluation . The localization of regions of interest in the approach of the described component is achieved by detecting moving skin regions. Head detection is performed by a very well known algorithm (Viola-Jones) and skin sample is extracted for building a skin model that will be used for hand detection. Skin information is further enhanced with motion information and Kalman filtering ensures noise removal and smooth hand trajectories.

Wii-based Gesture Learning Environment component (WiiGLE), developed by University of Augsburg partner, is used to classify hand movements in the three-dimensional space based on the analysis of acceleration data from Nintendo's Wiimote controller. For this purpose, a general classification process pipeline has been implemented that allows to record training corpora of arbitrary gestures, train classifiers, and then use these classifiers for online recognition of gestures. The Wiimo-

te uses accelerometers to sense its movements in 3D space. The controller is able to connect via Bluetooth to a common PC. The acceleration data is gathered for each direction (x: left/right, y: back/forth, z: up/down) with a sampling rate of 100Hz. To allow for fast and simple use of the Wiimote in a number of different applications, the WiiGLE environment has been developed and make it available to the CALLAS Shelf. It allows defining arbitrary gesture classes for an application, selecting features for the classification task, training and comparing classifiers, and using it as the classification component of an application.

Behavioural Cues

Can be detected using camera and some image processing, or with any accelerometer-equipped sensor device held or worn by the target person. Traditionally accelerometers have been used for measuring simple motion, and more recently it has become a widespread subject of interest to map motion and gestures directly into control input of various processes, the Wii console being the most prominent example today.

Accelerometer-Based Gesture Recognition

This component, developed by VTT partner, it becomes possible to extract semantic, emotional cues from the way the user is holding and moving any mobile phones equipped with an accelerometer as a sensor.. The aim of the component is to map types of movement defined by expressivity parameters (e.g. graceful/fast tempo) to different emotions, not to do direct emotional gesture recognition. By themselves the extracted features can be used to measure for example the level of excitement of the user, but rather than to determine exact emotions, the goal of the component is to extract expressivity parameters that when combined with other modalities, such as audio analysis, can be used to improve the precision of determining the mood of the user significantly. The decision to offer technology for recognizing gestures from mobile phones as input devices in addition to Wii controllers was based on the simple fact that more people have accelerometer-equipped mobile phones on their persona at any given time than they do gaming accessories. A mobile phone is a device that everybody possesses and in the future it can be a multipurpose tool even for accessing gesture-driven multimodal interfaces. This encourages the development of gesture/body motion tracking with the mobile phone as a sensor. This makes it possible to create future public installations, where people can use their own phones as gesture input devices.

Inertial Platform Embedded in Human Glove

The Human Glove is an innovative device with related software component. It was developed by Humanware partner with the purpose of investigating if and to a what extent kinematical data like acceleration, velocity and jerk of movement of the forearm/arm could be effectively used for implicitly detecting emotion information. In particular the main objective of

this device and its related software platform, consisting of accelerometers, gyroscopes and magnetometers, is to allow to use the data both for tracking the arm/forearm movement (integrating kinematical data to get the relative position of the hand from the torso) and for analyzing the user's activity (directly from kinematical data or differentiating the acceleration to obtain the motion jerk), where the distance of the hand from the body, the hand speed, the hand acceleration and the hand jerk can be used to detect implicitly communicated affect. Moreover, these data can be used to map the emotions in a PAD space and fuse them with other modalities.

Gaze Detection and Head Pose Estimation

This component, developed by ICCS partner aims at estimating the Focus of Attention of a person sitting in front of a Computer web-camera, in real time. It takes into account the Facial Features coordinates detected in order to estimate gaze directionality. Translations of the eyes middle points give indicates regarding head rotation, while the fraction between the inter-ocular distance and the vertical distance between the eyes and the mouth is always monitored. If the fraction is found to be restricted within certain limits, no rotation is decided but the user is supposed to be conducting only translational movements. Also, the eye centre movements, with regards to the coordinate system defined by the four points around the eye, gives indicates regarding eye gaze directionality. The above is averaged for both eyes, so eye gaze is estimated. The above metrics are normalized with the inter-ocular distance in pixels and, thus, are scale independent.

Facial Feature Detection

The Facial Feature Detection is a core part for analysis of both Head Pose-Eye Gaze and Facial Expression Recognition Components. It provides normalized coordinates of 19 facial features: Left/Right eye centres and corners, eyelids, two points on each of the eyebrows, nose centre and four points around the mouth. Facial feature detection is realized on a two-stage level: At the start-up of a frame sequence, after face detection, facial features are detected using Distance Vector Fields. This method has been chosen as it takes into account geometrical information of features' area, thus, it has a large degree of independence from natural lighting conditions; a key factor that has been taken into account for this particular system. Following this, facial feature extraction is executed on a frame-by-frame level, using a two-level Lucas-Kanade algorithm. Lucas-Kanade is a widely known and extensively used tracker in Computer Vision, as it can track successfully highly detailed features

Facial Expression Recognition

This component, developed by ICCS partner provides decisions regarding emotion from expressions. The output can be an emotion or a quadrant.

The emotions extracted are the six Ekmanians ones: Anger, Disgust, Joy, Fear, Surprise, Neutral Expression. The other possible output is a quadrant. The component uses the emotional dimensional model and gives as output quadrants of a coordinate system defined by two axes: Passive-Active emotions and Positive-Negative emotions.

Smart Sensor Integration (SSI)

The SSI is a framework for multimodal signal processing in real-time, developed by University of Augsburg partner. It allows the recording and processing of human generated signals in pipelines based on filter and feature extraction blocks. In this way standard sensors, such as microphone, camera, or Wiimote are turned into "smart" sensors, which do not any longer deliver raw measurements, but present information in a form that meets the requirements of the application as effectively as possible. Originally dedicated as a framework to jump-start the development of multimodal online emotion recognition (OER) systems, it particularly supports the machine learning pipeline in its full length and offers a graphical interface that assists user to collect their own training corpora and use them to obtain personalized models. It also suits the fusion of multimodal information at different stages including early and late fusion.

Low Level Multimodal Fusion

This component, developed by ICCS partner, is the integration of the input of different sensors at an early stage of processing. The component exploits the short-term memory function of recurrent neural networks in order to integrate different kinds of information within a modality, e.g. when recognizing emotions from facial expressions by combining features from various parts of the face. This component includes three different processing modules; a simple neural network based on the back-propagation learning algorithm, a simple recurrent neural network based on Elman's prototype, and an alternative recurrent network based on the algorithm of real time recurrent learning. In all three modules, the user is able to setup the network's structure for the specified problem, perform a training procedure to determine the network's connection weights and, also, test the resulted network's performance against previously unseen data.

The interpretation, understanding, and fusion of the multimodal sensor inputs were a major goal of the CALLAS project. Of particular interest is the mapping between gestures / body movements and emotion, and their relation to other modalities, such as affective speech and mimics. However, multimodal data corpora with emotional content are rather rare. Actually we were not able to find one, which contains all input modalities needed for our analysis. To this end, we decided to set up an experiment, which allows us and others to build such a corpus from scratch. Doing so, we did not only come up with an abstract idea of a suitable setting, but also created the necessary tools, which allow everybody to repeat the experiment in the same or a similar manner. This way partners at different locations were able to contribute to the same data cor-

pus, which helped us to enlarge the database. Captured modalities include speech and facial expressions but the focus is on hand gesture expressivity. Thus, this is the primary modality and is recorded using three methods: bare hands, Nintendo Wii remote controls and data gloves. Such a setup allows for a multimodal affective analysis and potentially provides quantitative parameters for synthesis of systems affectively aware and able to convey affect, such as Embodied Conversational Agents. Additionally, comparative studies of gesture expressivity based on different recording techniques could be based on the introduced corpus. Cross cultural affective behaviour issues are also incorporated since the experiment was performed in three countries i.e. Greece, Germany and Italy.

Callas Output Shelf

Affective Embodied Conversational Agent (ECA)

From the output side a comprehensive work concerned the improvement of Greta, an Affective Embodied Conversational Agent (ECA), made available by TCOM partner. Greta is a general purpose use, modular, 3D embodied conversation agents that works in real-time. It is able to convey its communicative intentions using both verbal and various nonverbal communicative channels. Greta can talk to the user and contemporarily display facial expressions, gestures, gazes, torso and head movements. It may use nonverbal cues during the talk, for example, to accentuate the verbal content, to disclose some cognitive processes (by using performatives), or to signalise her emotional state. Greta's architecture follows the SAIBA framework that defines functionalities and communication protocols for ECA systems and the MPEG4 standard of animation. The agent system is optimised to be used in interactive real-time applications.

It was developed originally within Humaine and MagicSter EU projects, and its architecture was completely redesigned within the CALLAS project in a full modular way in order to implement a near real time version. The research work focused on the improvement of expressive qualities; as well as the stability and efficiency.

Affective Music Synthesis (AMS)

This component, developed by University of Reading partner 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.

The real-time music affectivisation of AMS consists of a generic sequencer and generative music synthesiser for producing and synchronising percussive rhythms and melodies. It uses generative minimalistic Brownian melodies sequenced with rhythm tracks of multiple instrument sounds and musical constructs (there is a variety available, wind, string, percussion, auto, mute) to arrive at the resultant music piece. This resultant is affectivised based on user and reviewer feedback by sub-

jecting it to pitch, tempo and velocity modifications before delivery. This flexible track and pattern arrangement leads to enhanced mood support with gradual and perceivable changes. When music is composed by AMS, continuous PAD representations of user mood are input into AMS in real-time which are resolved by Case Based Reasoning and the solutions of the retrieved cases are passed to a Real-time Music Affectivisation module developed in Pure Data. This module uses the solutions to the varying emotive input provided by CBR to compose and affectivise music that is fed back to the user on-the-fly. Each time AMS receives emotive state data, 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 sequence accordingly

Affective Music Synthesis provides for a stage where improvisation using compositional algorithms and affectivisation logic in real-time creates an ambient sound experience for the users. The lack or absence of dependence on an active user input and interaction in the composition process and its usage of continuous user affects i.e. emotive information for affectivisation, sets it apart from other works in the past. It uses multiple modalities and musical constructs to arrive at the resultant e.g. percussion sequencing with a generative tune overlay.

The CALLAS IDE (Integrated Development Environment)

The CALLAS IDE consists of both the software Framework and the related Authoring Tool as a graphical user interface for supporting intuitive design of multimodal interactive applications

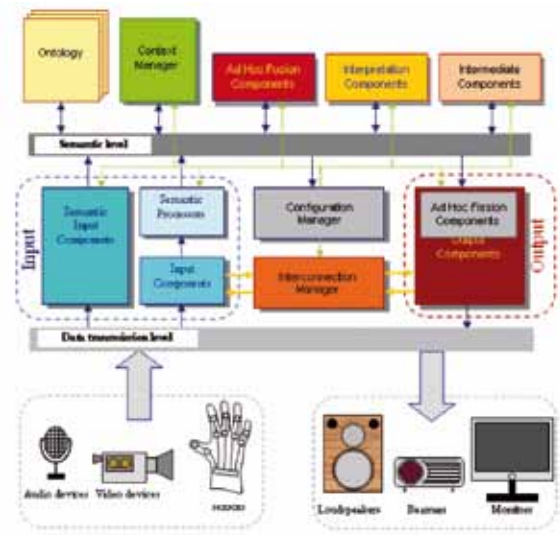
The CALLAS IDE has been designed as a software infrastructure that will allow Shelf components to be differently combined to develop specific affective multimodal end-users applications in the field of Digital Art and Entertainment.

The CALLAS Framework is finally a software infrastructure making it easier the life for digital art and entertainment application developers. Thanks to the CALLAS Framework developers will not start from the scratch any time they decide to design and implement a multimodal interactive application, but they can use a suite of open-source and interoperable toolbox and software sub-assemblies for saving a lot of time in the application development.

Despite the availability of more robust multimodal input software components, as well as higher-accuracy and more mature devices such as gaze trackers, touch screens, and gesture trackers, very few affective multimodal interactive applications so far have gained benefits of these advancement. One reason for that is the high cost in time for implementing multimodal interfaces by starting from scratch. However, when software engineering principles and methodologies are properly implemented, like in the case of the CALLAS project, a large part of the software in an affective multimodal system can be reused, allowing developers to rapidly prototyping new application at lower costs.

One of the main aims of the CALLAS framework was to provide an intuitive metaphor suitable for nontechnical users willing to adapt and repurpose the CALLAS high-level components or their combinations. The Blackboard Pattern has been used as a solution for a suitable metaphor, easy to use for end users eventually without special technological expertises in multimodal interfaces.

The CALLAS Integrated Framework is a software "glue" whose concrete implementation provides a proper set of API for a semantic intercommunication, aimed at integrating heterogeneous modules often coded in different programming languages and running on different operating systems. Components can be native, if they access the framework by invoking directly its API, compliant, if they implement a well specified integration protocol, or external, when an ad hoc integration solution is needed. Thanks to the adoption of the Blackboard paradigm and to the implementation of a proper Configuration Manager, the aggregation of components can change either over time or on the basis of data exchanged. So, the resulting applications are able to gather emotions of the audience and respond in an emotional-aware way by relying on mul-



timodal interfaces that can evolve during the application lifetime.

The CALLAS Framework over performs other existing integration framework for multimodal interfaces in many significant features:

* None of them manage the emotional states of the audience, whereas the CALLAS Framework must provide fusion and interpretation algorithms whose main goal is the emotion recognition and interpretation.
* This approach allows designers and artists to overcome the limitations of the pipelining approach by modeling applications as finite state machines where each state corresponds to an aggregation of components. With this regard a key functionality made available is the dynamic runtime configuration of the final artistic installation, that must evolve over time depending on the information circulating among components.
* The CALLAS Framework allows the integration of a wider range of components provide that they are able to communicate through OSC (Open Sound Contro) over TCP/UDP connections, rather that integration could be possible only if they provide a well defined API.
* The CALLAS Framework target includes also artists that are often not confident with technology, whereas all the other existing framework tailors applications developers.
* Applications developers are mainly programmers, while the CALLAS framework target includes also artists that are often not confident with technology.

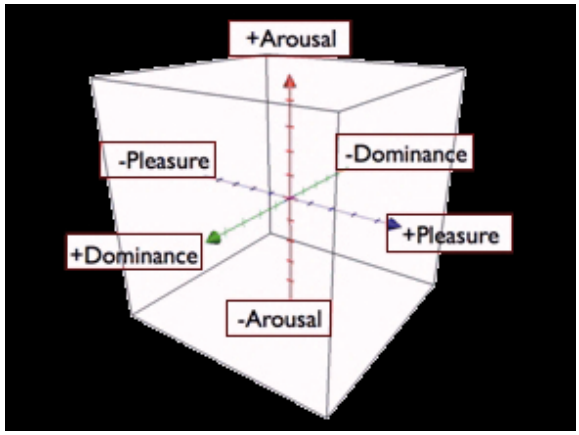
The CALLAS Affective Multimodal Fusion

An innovative Affective Multimodal Fusion approach has been developed, extensively validated within the CALLAS project, and has been finally made available as an additional component to the CALLAS Shelf and Framework for being reused in other application scenarios.

The originality of the CALLAS approach to the multimodal fusion lies in performing fusion at (late) semantic level, by integrating common meaning representations derived from different modalities into a combined final interpretation. In particular, the implemented fusion process is able to integrate disparate descriptions of emotional states and other affective input into a blended affective response which is continuously variable over the time. This differs from the available discrete methods for affective fusion, which are essentially aimed at combining co-occurring affective inputs to increase confidence in a classification of emotional state at any given time.

On the contrary the CALLAS affective fusion methods build on the PAD (Pleasure-Arousal-Dominance) dimensional emotional model, which maps emotional tendencies and responses along three dimensions: pleasure-displeasure, arousal-non-arousal, dominance-submissiveness.

Inputs received from different modalities are processed by associated components to produce separate PAD vectors and/or interest values for each component. Thereafter, outputs of all the components are combined using weighted sum to resolve PAD state of the overall system.



The continuous nature of PAD model is the key for modelling intermediate state of affect which may not have an a priori labelling. By using such model, the direction of the vector represents the type of emotion, while its length indicates the intensity. The affective experience can be described by the path of this vector.

The CALLAS fusion approach has undergone a further round of evaluation, incorporating a range of physiological measures of emotional state as ground truth (corrugator EMG, zygomaticus EMG, Heart Rate, GSR), as well as questionnaires on interactive experience. 30 pairs have been evaluated, using measurements from both member of a pair where possible. These have shown positive results, especially in the correlation between zygomaticus EMG and perceived pleasure.

CALLAS Showcases

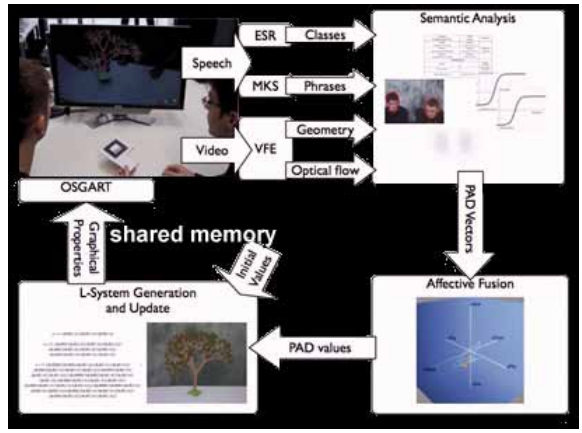
Several real life application prototypes have been made developed for the Digital Media And Entertainment domain as proof-of-concept of the effectiveness of the CALLAS integrated approach.

"E-tree"

An Augmented Reality artistic installation, in which the behaviour and the appearance of a digital tree artwork dynamically change along the time, depending on the affective state of the interacting users. The objective of this showcase was to demonstrate the integration of an AR environment with CALLAS affective output components to create an art installation which instantly react to the emotional states of the interacting users by generating an emotional responses. The artwork is a naturalistic representation of a tree whose growth and evolution reflects the perceived affective response of spectator throughout an interactive session. Spectators reactions are captured using emotional speech recognition (EmoVoice), multi-keyword spotting (MKS) and video feature extraction (VFE) input shelf components.

The appearance of E-Tree is achieved through the definition of graphical objects within a formal grammar known as an L-System. This grammar is transformed into a 3D scene graph within the OSGART AR graphical framework.

The latest E-Tree prototype defines this gram-



mar in external files that can be modified by the artist or adjusted to suit different environments in which the installation is employed.

The general structure of the tree is achieved by applying transforming rules at regular intervals that decide where new branches will be formed. The branches and there accompanying leaves are defined both by a predefined set of rules, but also parameters that modify these rules and control internal "growth" processes that animate the creation of branches and leaves. These parameters are derived from changing PAD values and externally specified control values that can again be modified by the artist.

There are two temporal variables that control changes in E-Tree, the general passage of time as each branch grows and changes, and the generation of a branch, i.e. how many times a rule has been recursively applied. In general, the addition of a new branch to an existing

apex is only dependent on the preceding elements in a rule, so that branch growth is not dependent on generational timing. Once a branch or leaf completes its initial growth animation to the values specified in the appropriate rule, the simulated internal processes can cause the branch or leaf to grow further, wither and die, be pruned or fall off, and subsequently regenerate. The maximum number of generations allowed (the more generations the more complex and potentially realistic the tree) is limited by PAD values, so that a more "positive" tree has the potential of many branches and leaves, while a "negative" affective input will cause the growth and branching of the tree to be limited.

The branching rules of the tree are defined stochastically. That is, there is a set probability that a particular rule will be applied to a potential apex or branch transformation. E-Tree defines distinct sets of rules whose probabilities are determined by current PAD values. The rules strike a balance between upward growth of the tree and lateral branching. A low amount of Pleasure and Arousal will favour minimal branching, higher values of Pleasure will cause a greater amount of branching, and higher Arousal will favour upwards growth.

Common Touch

The Common Touch Showcase aims at demonstrating multimodal affective applications for group-based interaction in public places. Common touch means, figuratively, the ability to appeal to ordinary people - is a large

interactive wall that is meant to be installed in a street as the window of a shop. Common words of an incomplete sentence are displayed in a conspicuous way on the screen. It appeals to the walking people-by, who lightly touches the wall and notices that the touch not only reveals hidden words, but also makes another sentence appear. Seeing this person interacting with the wall, other passers-by stop and begin to touch the new sentences to reveal hidden words. They are soon overwhelmed by slogans and trying to touch them all. To drive the type of slogans displayed, the affective expressions of the group - derived from multimodal tactile, video and audio inputs - are exploited by the installation like in an advertisement or a charismatic oration. Both the crowd and the installation contest for the control of the display.

This showcase is a large interactive wall that is meant to be installed in a street as the window of a shop. Common words of an incomplete sentence are displayed in a conspicuous way on the screen. It appeals to the passer-by, who lightly touches the wall and notices that the touch not only reveals hidden words, but also makes another sentence appear. Seeing this person interacting with the wall, other

passers-by stop and begin to touch the new sentences to reveal hidden words.

Besides explicit touch interaction revealing hidden words, multimodal implicit inputs are used to monitor user affective response to the installation. The system analyses user speech in terms of emotions and predefined sequences of words (through the MKS and ESR components), group size (through the face recognition component VFE) and a number of simultaneous touches on the touch display surface, whereas the goal is to map received signals to a PAD model. A system internal PAD state controls which sentences are displayed to the user. Sentences are chosen based on PAD value assigned to them beforehand manually using self-assessment manikin method

The Common Touch is to be exhibited in autumn 2010 at ENSAD during La Nuit Blanche, a major artistic event in Paris was designed and developed during m37-m42. The showcase makes full use of the Framework as well as the Teesside fusion component, together with EmoVoice, MKS and VFE components. It now integrates MultiTouch a new CALLAS component. The scenario has been validated with high visibility artistic partner (Ecole Nationale Superieure des Arts Decoratifs - ENSAD) that targets an exhibition in autumn. The showcase has been tailored to illustrate contributions on affective interaction in semi-public settings.

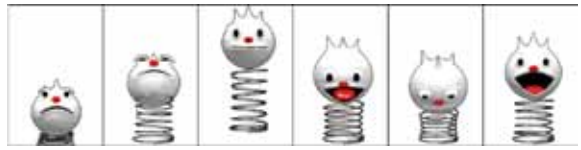
"Galileo to Hell"

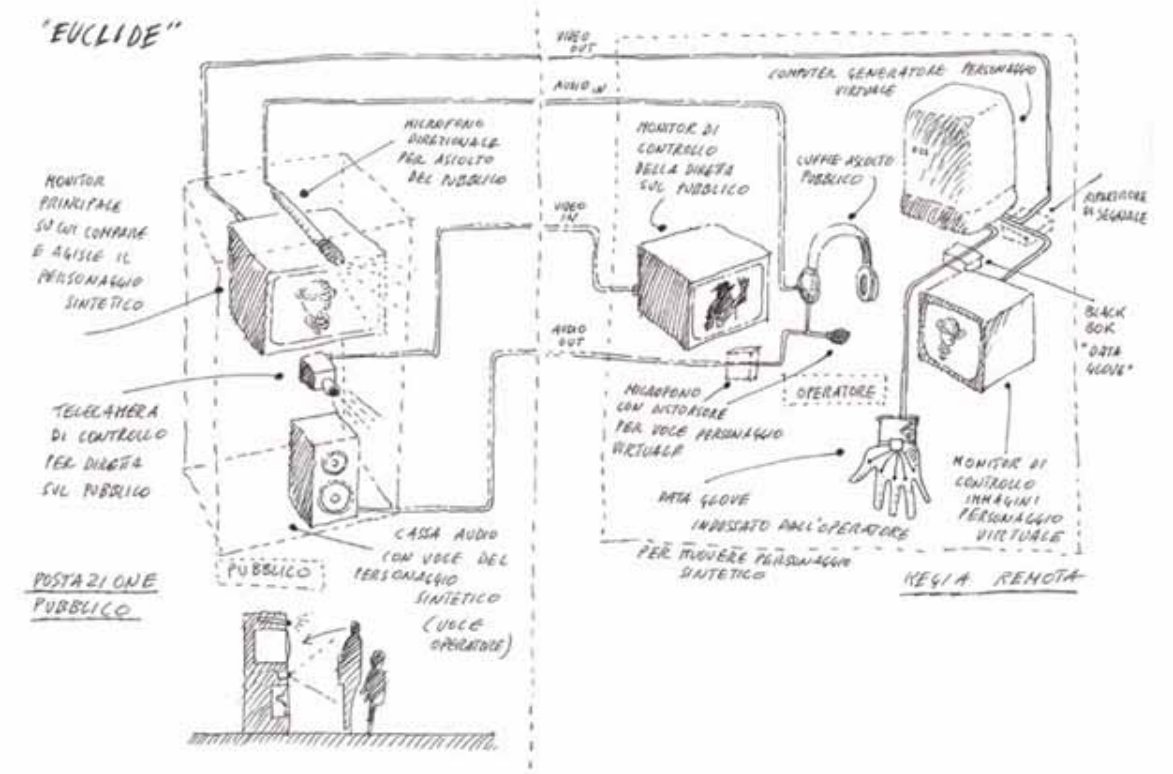
"Galileo to Hell" is another exploration on the group-based interaction in a real life the-

atrical setting was the theatre show, realised by Studio Azzurro partner and performed on July 2008 at Teatro degli Arcimboldi (Milano, Italy), in which the whole stage becomes an interactive "sensible" environment where the spectators are allowed to interact emotionally with the installation through their behaviour The EmoVoice and VFE Shelf components have been used in this application.

Euclide

Developed by Studio Azzurro partner, is an exploration into interactive, affective puppetry and demonstrated 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 reconfigure the spaces they inhabit and visit.

Spectators movements, facial expression, voices are detected through Smart Sensors Integration SSI, Video Feature Expression, and Multi-KeyWord Spotting input shelf components and an overall emotional state of the interacting user is dynamically computed. The puppet response and animation are triggered by the detected emotional states, based on a subset of pre-recorded animation issues (voices, behaviours, effects). This showcase was demonstrated at Science Museum in Naples and was particularly suitable for groups of young people.

MusicKiosk

The MusicKiosk implemented by XIM partner, is a museum installation for young people mu-

sic edutainment, which engages young visitors of the Museum of Musical Instruments at Accademia Nazionale Santa Cecilia in Roma, in creating different music compositions, supported by animated affective characters, depending on the detected user affective state.

The primary objective of the MusicKiosk showcase, was to demonstrate the capabilities of the CALLAS Framework internally (within the CALLAS Consortium) and externally (to the general public) in an entertainment product.. The concept behind the MusicKiosk developed out of a previously-established working relationship between XIM and the Accademia Nazionale di Santa Cecilia (Rome). Working from a historical image depicting a concert held in honour of Queen Margherita, a fictional storyline was developed concerning the boy violinist in the image who is late for a concert. The human player "guides" the boy through various backstage rooms of the concert hall where rehearsals are taking place, before finding the stage at the start of the concert. This amusing scenario provides opportunity for the boy's face (depicted in cartoon form) to mirror the estimated emotion of the human player's mood as detected in facial expressions, vocal intonation and emotive keywords. A small vocabulary of keyword-recognition is used so that the human player can control the MusicKiosk, for example directing the boy from room to room by calling out the door numbers, or asking the system to restart. Additional interest was added to the MusicKiosk by introducing a second character using a stylised cartoon version of the boy's violin: the human player may choose whether to be the boy or the violin at the start of each session. Once the human player has chosen a character, there is a choice of "rooms" to enter. Each contains different musical styles and different music is played. The music changes depending upon the Ekmanian Expression detected in the player's face and the amount of interest (Arousal in the P-A-D model) detected in the player's voice. As greater interest is detected, so more musicians play. The MusicKiosk was presented to live audiences of schoolchildren who visited the Accademia Nazionale de Santa Cecilia on educational visits. The children enjoyed interacting with the MusicKiosk and were able to try it out themselves. A simple (anonymous) survey was carried out among the schoolchildren to gauge their reaction to the system. It should be particularly noted that the MusicKiosk is designed to be controlled by the spoken words and facial expressions of random participants: no "voice programming" or expression training is required either for the participant or the Kiosk, and the entire system can be used by members of the public with no prior training and very little guidance.

Interactive Opera

This interactive opera showcase, developed by Digital Video partner, is a public installation set up in Teatro Massimo (Palermo, Italy), empowering children to actively interact and affect the behaviour of the opera characters, as rendered within a cartoon interactive opera.

The goal is to present animated cartoon stories, depicting the most famous Opera masterpieces (like "Il Barbiere di Siviglia and several

others), in order to let the audience directly interact with animated characters: users can control characters' moods and animations with emotions expressed by their face and the tone of their voices and characters will act correspondingly, showing animations that represent a set of Ekmanian emotions, such as fear, anger, joy, boredom, sadness. The partnership between the Teatro Massimo di Palermo and Digital Video led to an Interactive Opera exhibition in the theatre, prepared for young children visiting as school students, joining the experience of a new way to learn characters and stories regarding the world of the Opera.

Three virtual characters play inside a cartoon environment that reproduces the stage of the aria and each one of them is controlled by a person who can use a webcam pointed at his/her face and can change the character's mood changing accordingly the face expression or the tone of the voice, recorded with a microphone.

An audience of school students in Teatro Massimo revealed interest and amusement for the showcase. Children enjoyed playing with characters and teachers gave us advices to improve the edutainment experience. Based on the audience feedback gathered from the exhibitions, Interactive Opera's stories will have to be more complex and articulated, with emotional behaviour changes that will have to modify significantly the course of the story and make the narrative context more non-linear. Further development of Interactive Opera has also involved the integration the showcase with the XIM MusicKiosk in order to share a common communication platform with the CALLAS Framework, thus making Interactive Opera be compliant with other CALLAS Shelf input components and enrich the users' possibilities for more multimodal interaction with the animated environment.

Affective Interactive Storyteller

Was developed by the BBC partner as a short demonstrator aimed at providing a much richer interactive experience in a next generation TV context, simply pressing the red button can give. In doing this, the target was on the core business of the public broadcaster: telling stories.

The Affective Interactive Storyteller is a computer application that behaves like a live storyteller, in the sense that it brings the audience through a desired sequence of affective states over the course of a story, in order to more effectively convey the underlying message. It presents story content with the aid of an ECA and a sequence of still images. Each image in the sequence represents a scene or segment of the story. Interaction takes the

form of a guided conversation between the ECA and the user: the ECA invites the user to comment on the first image in the sequence; then the ECA tells the story for that specific image; then the image for the next scene is presented and the ECA once again invites the user to comment; and so on. The spoken (and gestural) input from the user is analysed for cues of their affective state, and this is used to direct the way in which the ECA conveys the next segment of the story - the idea being that the user is guided through a sequence of affective states appropriate to the story arc.

The CALLAS Marketable Assets and the Sustainability Model

A strong added value of the CALLAS project was on the:

* investigation on potential business models aimed at laying the foundations for turning on research prototypes into mature applications, eventually in domains with full or at least partial overlapping with the primary target CALLAS application domains,
* identifying in the Open Source CALLAS IDE and the Shelf Components the main concrete marketable outcomes of the project.

Despite there are currently many R&D institutions involved in the overall multimodal interfaces value chain, as well as plenty of innovation-led SMEs, unfortunately due to the high costs necessary for developing and implementing a multimodal interactive application, so far the multimodal interfaces 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. The currently experience high 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 utilize risky technologies in a really fragmented and niche market. Thanks to the cost savings deriving from the usage of the CALLAS IDE and framework, the market risks are becoming more acceptable also in the Digital Media and Entertainment domain.

All of this will be made possible thanks to the adoption of a Incubator-based Business Ecosystem Model, which will be able to foster the technology transfer from Innovation-led SMEs and large companies towards the market, in which Large IT&Service Providers will play the fundamental role of "hub" for leveraging the risks and approach easily the market. Thanks to the envisaged capacity of the large companies to do even medium- and long-term investments, the SMEs will be given the opportunities of adopting and driving innovation to the tailored market.

Potential Application Scenarios: A glimpse into the future

More than this both the CALLAS Sustainability Model and the full modularity of the CALLAS approach and technologies unveils the possibility for being replicated in other yet partially overlapping application domains.

Entertainment

The majority of our daily interactions with the physical words requires us to operate in a three-dimensional space and handle three-dimensional objects with different shapes, sizes and forms. Thanks to the continuous advances into the sensing technologies the interactions will move away from being purely surface-bound (like in the traditional case of computer or mobile phone displays), and involve people's movement and physical objects above or in front of the display. The directness and ease of use of current multitouch interactive surfaces already highlights the promise of the natural (multimodal) user interfaces where the only experience the user needs to start interacting is their real life experience. Such emerging technology trends, matched with affective multimodal interfaces, lay the foundations for completely new entertainment applications, like search engines and gaming.

Natural Control Interfaces for Gaming

Nowadays a lot of innovation is emerging into the gaming domain. The traditional game play practice of using an arcade joystick and two push-buttons has been progressively replaced by an intuitive experience offered by hand-held controllers like Nintendo's WiiMote and Nunchuk, Sony's PlayStation Motion Controller or Activision's RIDE skateboard and other emerging platforms implementing computer vision, speech analysis algorithms (Microsoft Project Natal) and embedding physiological sensors (Wii Vitality Sensor) able to real time facial expression, hand and body gesture tracking, and biofeedback analysis. Standard human movements and interaction with physical objects are the significant way of interacting with digital content in all of these controllers. As a result, players can interact with the game characters and environment in new augmented ways, using nonverbal cues; e.g. smiling or gloating when they win or shouting in desperation in the sight of a fleet of enemies.

Analyzing, capturing and synthesizing player experience in both traditional screen-based games and augmented- and mixed-reality platforms has been a challenging area within

the crossroads of cognitive science, psychology, artificial intelligence and human-computer interaction in the last years. New gameplay modalities could enhance the importance of the study and the complexity of player experience. Artificial and computational intelligence could be eventually used to synthesize the affective state of player characters, based on multiple modalities of player-game interaction. Multiple modalities of input can also provide novel means for game platforms to measure player satisfaction and engagement when playing, without necessarily having to resort to post-play and off-line questionnaires. For instance, players immersed by gameplay will rarely gaze away from the screen, while disappointed or indifferent players will typically show very little response or emotion. Affective Multimodal Interfaces could be used to adapt the game to maximize player's experience, thereby, closing the affective game loop: e.g. change the game soundtrack to a vivid or dimmer tune to match the player's powerful stance or prospect of defeat; maximize frustration by increasing the number of gaps in a platform game. From the point of view of non-player characters, an injured or frustrated opponent will look down when facing defeat, informing the users about its status, much in the way a human opponent would be expected to.

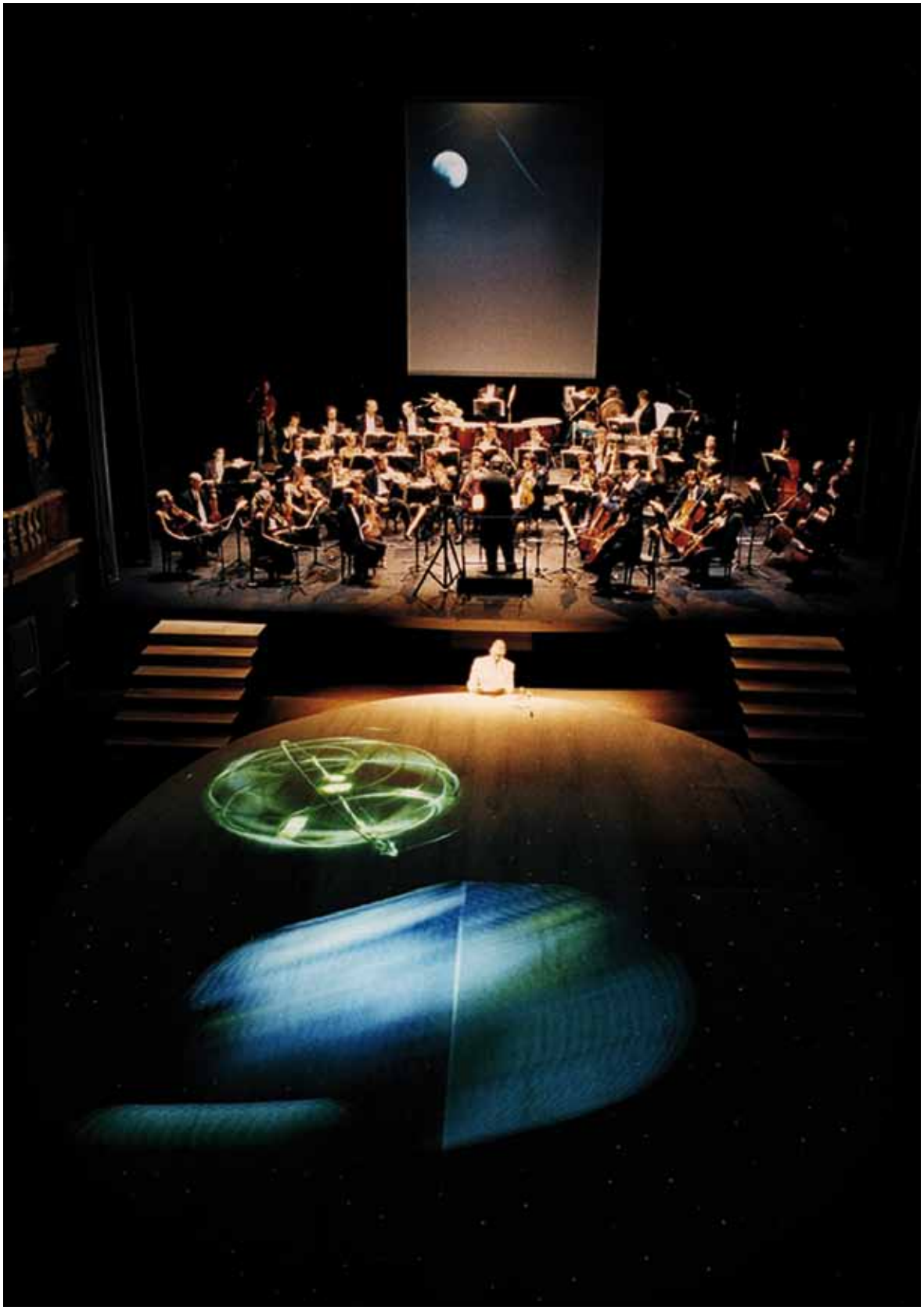
Affective Multimodal Interaction for Search Engines

This envisaged application scenario is aimed at develop and make it available a next generation multimodal search engine, in which the quality and the effectiveness of the search could be dramatically improved thanks to the further narrowing of the research criteria.

The recent advances in ICT have led to affordable costs of seamless computing and advanced networking technologies, paving the way, in this respect, for the emerging of an Ambient Intelligence environment, where the user is unobtrusively and transparently fully immersed. As an effect, the users can interact anywhere and anytime with such an environment, where the difference between the real and the virtual artifacts do not make sense anymore.

It is now possible for users to rapidly move from a mainly textual-based to a media-based Internet, where rich audiovisual content (images, graphics, sound, videos, 3D models, etc.), 3D representations (avatars), virtual and mixed reality worlds, serious games, life-logging applications, multimodal yet affective utterances (gestures, facial expressions, eye movements, etc.) become a reality.

Consequently, a new generation of affective multimodal and multimedia search engines able to handle at the same time multimedia and multimodal content is going to emerge, in order to fully support the experience in the form of real enjoyment of these media, in the sense of having true interaction with the media. The vision here is for a really context-aware search engine which will be able to collect and use any kind of information or content coming from the user as effective way for further narrowing the search, matching finally what the user actually wants



to search, and ultimately improve the quality of the user experience. Such next generation search would be able to handle specific types of multimedia (text, 2D image, sketch, video, 3D objects, audio and combination of the above) and support affective multimodal interaction (gestures, face expressions, eye movements, nonverbal yet implicit, emotional cues) along with real world information (GPS, temperature, time, weather sensors, RFID objects,), which can be used as queries and retrieve any available relevant content of any of the aforementioned types and from any end-user access device.

Digital Art based on Brain Computer Interfaces

Although the main CALLAS focus was on the usage of multimodal affective interfaces for Digital Art, the recent advances in Brain Computer Interfaces (BCIs) and in the medicine originally born physiological measurements, when matched with affective multimodal interfaces, unlock new yet unexplored possibilities for affect-based self-generated media content generation and for music artworks performing, based on the raw creative ability of the unmediated brain, bypassing the performers' bodies.

A fascinating scenario at the intersection of the advancements in Affective Computing, multimodal Interfaces and Brain Computer Interfaces would be an affective brain-driven orchestra, in which all the members play virtual musical instruments through Brain Computer Interface, and music and video change in time with the performers' brain waves and heart rate.

Single performers could be shown images while her heart rate and skin conductance are being measured. As her mood changes, so does the visual experience: images are blurred and changed in line with the changing biological measures of the conductor. Affective expression of a performance could be controlled by the real-time measured physiological state of a human.

Adaptive Personalised Museums Experience

Affective Multimodal Interfaces represent a very power communication tool for museums and cultural exhibitions allowing them to attracting a wider audience, especially among the younger generation, which not very seldom feels bored during a museum visit. In particular the joint application of advanced personalised and affective multimodal interactive technologies together with effective computer-based tools for analysing how people move around exhibition rooms, and what is her emotional state, pave the way for a new concept of museums which can turn on visitors from mere explores of artworks to active orchestrators of the real experience at hand. Consequently museums would become more effective in getting the exhibition's educational message across. In other words museums, when augmented with affective multimodal interfaces could make a deeper and more lasting emotional impression on the visitor while stimulating a full comprehension

of the artworks on display. Finally we could think technologically-augmented museums as body-driven interactive multimodal narrative spaces. Audiovisual animation or 3D projections concerning the various exhibited materials together with a real time detection (eventually through non-invasive physiological sensors) of the affective and the interest level of the users during the visit, can give the opportunity for next generation museum tours and recommendation systems to create dynamically a personalised tour, fostering at the same time a sense of wonder and genuine curiosity and interests in the museum visitors.

Affective Learning

Affective computing and affective multimodal interfaces could be successfully used to adapt the presentation style of a computer-based tutor when a learner becomes bored, interested, frustrated, or pleased. More than this if we focus on the school environments, "I can't do this" and "I'm not good at this" are common statements made by kids, and more in general by students while trying to learn. Usually triggered by affective states of confusion, frustration, and hopelessness, these statements represent some of the greatest problems left unaddressed by educational approaches. Education has emphasized conveying a great deal of information and facts, and has not modelled the learning process. When teachers present material to the class, it is usually in a polished form that omits the natural steps of making mistakes (feeling confused), recovering from them (overcoming frustration), deconstructing what went wrong (not becoming dispirited), and finally starting over again (with hope and maybe even enthusiasm). Learning naturally involves failure and a host of associated affective responses. Affective Multimodal interfaces are expected to play a relevant role in order to create adaptive and personalised technology-learning systems which are really responsive to the affective state of the learner, and consequently facilitate the child's and the learner's own efforts. Such "companion" would be able to help keep the child's exploration going, by occasionally prompting with questions or feedback, and by watching and responding to the affective state of the child—watching especially for signs of frustration and boredom that may precede quitting, for signs of curiosity or interest that tend to indicate active exploration, and for signs of enjoyment and mastery, which might indicate a successful learning experience.


Palazzo Collicola 宫是一座十八世纪的寝宫，它美丽、光明而又质朴，在这里仍然能够呼吸到另一个时代的气息。

Affective Multimodal Interfaces for Cars Drivers' Safety

It is largely acknowledged the importance of being calm and conscious on the road for safe driving. Many emotional states affect driving negatively. For example emotions like anger and angry are often identified as one of the main causes for accidents on the road. However there are many other emotional and physiological states that should be detected and controlled in order to improve the driver's safety, like sleepiness, frustration and stress. Merging affective multimodal interfaces with physiological measurement could represent an effective way to enhance drivers's safety into the cars of the next generation. A potential application scenario should help drivers in two ways. First a car could be equipped with a wireless wearable physiological sensing device which should measure galvanic skin responses (GSR), heartbeat and body measurement in order to detect the emotional state of the driver. Second, the system could either warn the driver to take precautions on her stat, otherwise, in specific circumstances, it will take precautions itself on her behalf.

Affective Virtual Companions supporting Elderly People Independent Living

Europe is experiencing a clear demographical change in the last years, with an increasing numbers of elderly people, and this of course is creating a considerable societal impact. The subsequent scenario consisted of a lot of elderly people living alone, and/or ageing workforce, spending money for ensuring security. Social and service robotics could be the right answer to effectively deal with these emerging societal problems, since they have a big potential to offer to elderly people to increase the quality of their life. With this regard a lot of research is investigating on domestic robots navigation in indoor environments and in flexible objects navigation, however some important questions have been emerging, such as what are the circumstances in which elderly accept robots, how they will be able to communicate with robots, which are the most important services a small companion robot should provide to elderly people. Matching social robots with affective multimodal interfaces would allow to develop intuitive multimodal affective-aware communication between elderly users and their virtual companions.

Since affect is undoubtedly an important requirement for artificial companions to be capable of engaging in social interaction with human users, affective multimodal interfaces could contribute to answer to the above posed question to a most extent. In such a case affective augmented robots capable of processing and render affective information would increase the motivation and the engagement of elderly people in effective social communication with their virtual companion, and, above all, with their relatives, and among them.

Social Inclusion of Disabled People

People affected by some forms of disabilities can benefit greatly from Affective Multimodal Interfaces, which undoubtedly has a great potential to offer for effectively preventing social exclusion of disadvantaged categories of people. One of the most innovative yet unexplored application scenario is to use affective multimodal interfaces for people affected by autism disorders.

With this regard Affective Multimodal Interfaces could be effectively used as an effective way for improve the social communication of people suffering with autism disorder and enhance their social inclusion. Social-emotional communication difficulties lie at the core of autism spectrum disorders, making interpersonal interactions overwhelming, frustrating, and stressful.

Wearable multimodal affective interfaces, could help the growing number of individuals diagnosed with autism - approximately 1 in 150 children in the United States - learn about nonverbal communication in a natural, social context, as well as to help families, educators, and other persons who deal with autism spectrum disorders to better understand these alternative means of nonverbal communication.

E-Health: Virtual Rehabilitation of Motor Impaired People

Merging Affective Multimodal Interfaces with Brain Computer Interfaces (BCI) holds a great potential for virtual rehabilitation and motor recovery of seriously impaired people, who have lost totally or partially some body functionalities.. A challenging yet fascinating scenario could be to use a suitable combination of affective multimodal interfaces and BCIs for supporting clinical rehabilitation of post-stroke patients. In current neuroscience-based rehabilitation protocols based on mental rehearsal of movements (like motor imagery practice) are considered as an effective way to induce an activation of sensor-motor networks, which were affected by lesions. It follows that an affective-augmented BCI, based on movement imagery, augmented with the image showing of that movement (for example the grasping movement) would be able to provide patients with an early reinforcer sign in the critical phase where there is still no clear evidence of movement recovery. The affective interfaces component of the envisaged system could eventually help patients and therapists as well to detect a real time information on the affective state of the user, which could eventually affect in different way the carrying out of the exercise. For example if the system will detect a prevailing frustration state, it will be better for the therapist not to continue with the planned rehabilitation exercise, but rather it could be more effective to provide patient with a psychological support in order to alleviate her frustration state.

Experiential Marketing

Affective Multimodal Interfaces could have a great potential to offer to companies' marketing departments in a near future. Thanks to these technologies, marketing people could exactly know what customers are doing and the reasons why behind their actions, when accessing the product virtual catalogue by the company web site, or when in front of the physical products in a shop.

A challenging application for experiential marketing would allow to capture customer's facial expressions and gestures' expressiveness from body motion tracking in response to a specific company product, and, subsequently, to extract the customer emotional state. This will help companies to have a clear yet unprecedented insight into the heart and mind of the company's potential customers, to uncover the non-rational influences affecting customers decisions from purchase to engagement, and allows companies to understand how people are feeling about company's brand, products and communication strategy effectiveness.

Consortium

The Consortium is composed of 18 partners from 8 different countries, coordinated by Engineering .

It includes prestigious names from Research and Applied Technology, System Integrators and innovative SMEs, all collaborating to ensure the scientific relevance of its pioneering work and its acceptance worldwide to reach the target of early adoption of CALLAS showcased technology.

Members

* Engineering Ingegneria Informatica SpA - Italy [project Coordinator];
* British Broadcasting Corporation - United Kingdom;
* VTT Technical Research Centre of - Finland;
* Studio Azzurro - Italy;
* XIM - United Kingdom;
* Digital Video - Italy;
* Humanware - Italy;
* NEXTURE Consulting srl - Italy;
* University of Augsburg - Germany;
* Institute of Communication and Computer Systems - National Technical University of Athens - Greece;
* Université de Mons - Belgium;
* University of Teesside - United Kingdom;
* Aalto University- Finland;
* Telecom Paristech - France;
* Scuola Normale Superiore of Pisa - Italy;
* University of Reading - United Kingdom;
* Fondazione Teatro Massimo - Italy;
* Human Interface Technology Laboratory - New Zealand;

Contacts

* **www.callas-newmedia.eu**
* **info@callas-newmedia.eu**

Dr. Massimo Bertoncini
Project Director, R&D Lab
Engineering Ingegneria Informatica
massimo.bertoncini@eng.it

Prof. Elisabeth Andrè
Shelf Director
University of Augsburg
andre@informatik.uni-augsburg.de

Prof. Marc Cavazza
Integration Director
University of Teeside
M.O.Cavazza@tees.ac.uk

Prof. Giulio Jacucci
Showcases Director
Aalto University
giulio.jacucci@helsinki.fi

Diego Arnone
Framework Director, R&D Lab
Engineering Ingegneria Informatica
diego.arnone@eng.it

Antonina Scuderi
Impact & Communication Director
Nexture Consulting
t.scuderi@nexture.it


FONTE DELLA SCALA
MUSEO DELL'ACQUA
Via Martin Caserma Lamarmora
Via Monte Grappa
Via Nazario Sauro
Via Ricasoli
POZZO DEL ROTONE
POZZO DEL ROTONE

Image credits: Studio Azzurro Produzioni S.r.l.

Design: Gianluca Savini

Work partially supported by European Community under the Information Society Technologies (IST) programme of the 6th Framework Programme for RTD – project CALLA, contract IST-034800. The author is solely responsible for the content of this paper. It does not support the opinion of the European Commission, and the European Community is not resoonsible for any use that might be made of data appearing therein.

European Commission Sixth Framework Programme

European Commission

Information Society and Media