



## TABLE OF CONTENTS

### Centre Stage CALLAS

Multimodal Affective Interface Technologies.....pg. 3

### An Integrated Approach

The Shelf .....pg. 5

The Framework .....pg. 7

The Showcases .....pg. 8

### The CALLAS Community

What is it? .....pg. 11

Why join? .....pg. 11

Get involved! .....pg. 11

### Behind the Scenes

The CALLAS Consortium.....pg. 12

Project Governance .....pg. 12

Concept and graphic design: Metaware spa

All the pictures are courtesy of Studio Azzurro. © Studio Azzurro Produzioni srl

Printed by Tipografia Moderna, Navacchio, Pisa, Italy

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









## Multimodal Affective Interface Technologies

People naturally communicate combining gestures, movements, speech, verbal and non-verbal expressions. Within such a framework emotions and affectiveness play a fundamental role in enriching the naturalness of human-human and human-machine interaction and communication.

One of the main challenges for CALLAS is to implement the concept of affective emotional input for interactive media rather than within a traditional interface paradigm. Affective and emotional interfaces are generally concerned with the real-time identification of user emotions to determine system response. They rely most often on Ekmanian emotions such as joy, fear or anger. To enhance the user experience in Arts and Entertainment, CALLAS aims to handle new and innovative categories of emotions, as the well as improve the performance of existing modalities at the input level. This technology will impact a variety of extensions in New Media and business applications, including Augmented Reality for Art, Entertainment, and Digital Theatre, Interactive installations for public spaces and next generation Interactive TV.

However, interaction with new media such as interactive narratives, digital theatre or digital arts involves different ranges of emotions on the user's side, some of which correspond to responses to aesthetic properties of the media, or characterise the user experience itself in terms of enjoyment and entertainment. To identify these, more complex articulations of modalities are required across semantic dimensions as well as across temporal combinations. Firstly, modalities involved range from emotional language and paralinguistic speech (laughter, cries) to categorisations of user attention (suggesting interest or boredom for instance). Secondly, these have to be integrated across interaction sessions of variable durations rather than analysing a single emotional status in real-time.

One such example of integration consists of affective input to interactive narrative, in which the evolution of a baseline plot can be influenced by user reactions to the story unfolding, analysed in terms of overall attitudes (body postures, evolution of user activity, paralinguistic speech).

### **CALLAS is addressing the following high-level objectives:**

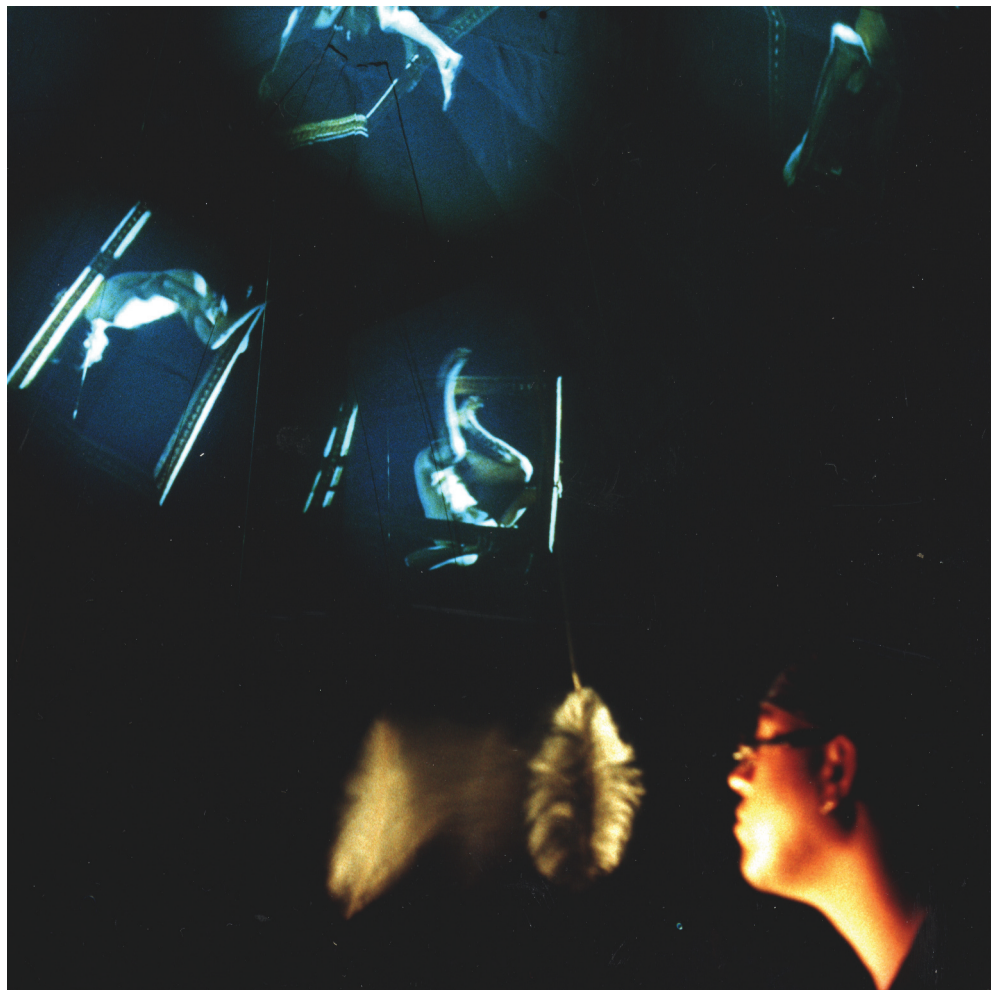
- advancing the state-of-the-art in Multimodal Affective Interfaces by developing new emotional models that will be able to take into account a comprehensive user experience in Digital Arts and Entertainment applications and new modality-processing techniques to capture and elicit these new emotional categories;
- researching, developing and integrating advanced software components, tailored to the processing of individual modalities supporting the semantic recognition of emotions, making them available through a repository, called the CALLAS "Shelf";



## **CENTRE STAGE CALLAS**

- establishing a software methodology for the development and the engineering of Multimodal Interfaces that will make their development accessible to a larger community, i.e. the assembly of a Multimodal interface from individual components that will eventually not require a deep understanding of the theories of Multimodality.

*Studio Azzurro, **Il soffio sull'angelo (primo naufragio del pensiero)**,  
Sala Fibonacci, Università degli Studi di Pisa, 1997*





### The Shelf

The Shelf consists of a dynamic pool of advances in multimodal interface technologies.

A specific methodology has been defined by CALLAS for the software components to be included in the Shelf, with particular focus on efficiency and robustness, in order to guarantee a consistent performance for many contexts and scenarios.

Following this methodology, significant efforts are being carried out to make available a comprehensive toolbox ready to be used in the development of affective multimodal applications. This is necessary because research results in the field are usually developed and tested in controlled settings, and therefore are not immediately suitable for use in uncontrolled “production” scenarios, especially by those users who are not experts in multimodal theory.

The CALLAS Shelf components will offer robust and effective building blocks for applications glued together by the CALLAS Framework plug-in architecture. These components include:

#### Emotional Speech Recognition

Beyond semantic information, i.e. the meaning of words, speech also conveys information about the emotional state of the speaker which can be valuable in order to actually estimate the speaker’s emotion.

#### Emotional Natural Language Understanding

Human speakers do not only express their emotions by means of acoustic, but also by means of linguistic features. As a consequence, the fusion of acoustic and linguistic features plays an important role for the analysis of spoken utterances. This component is based on an integrated method that includes acoustic as well as linguistic features, relying on a corpus-driven approach.

#### Sound Capture and Analysis

Sound analysis provides meaningful information containing speech, surrounding sounds (music, crowd cheering), etc. This will provide input for multimodal user interfaces and guarantee the natural and adaptive interaction with the physical and virtual environment, as well as in the creation of MR/AR environments.

#### Video Feature Extraction

Video streams combine both audio and visual information that contain information about the emotional state, content and context. This will provide modalities in the creation of multimodal and natural interaction in showcase HCI. The adaptive MR/AR environments created in the CALLAS showcases will also benefit from the information provided from video feature extraction. In CALLAS, the work will consist of exploring the utilisation of audio and visual features to extract contextual and emotional information about user,





## **AN INTEGRATED APPROACH**

environment and media, based on audio and speech in a manner which will benefit the showcase.

### **Gesture and Body Motion Tracking**

Gesture and Body Motion Tracking is an important multimodal component in building affective and natural interaction. This Shelf component will provide information about body movement and gestures that can be thresholds to different emotional states. The tracking will be performed with different sensors positioned from upper limb to the whole body.

### **Wearable Interfaces for Motion Capture**

The CALLAS consortium will develop Human-Machine Interfaces (HMIs) embedding different miniaturised transducers for gesture recognition and motion tracking.

### **Haptic Tracking**

Haptic systems improve the multimodal communication by means of force/tactile feedback. In order to improve the user/virtual environment interaction, Humanware will provide a 3D haptic tracker for virtual environment navigation.

### **Multimodal Interpretation of User Experience**

According to the CALLAS objective of providing “new paradigm for investigating emotions in multimodal interfaces tailored to New Media environments”, there is a need for specific research into the emotional categorisation of the user experience. This should open the way to affective interfaces that will be able to react to a wider range of user feelings than what the detection of basic emotions would entail.

### **Affective Multimodal Interpreter/Facial Expression Recognition**

Everyday human-human interaction is mainly performed through speech, facial expressions and body gestures to emphasise a certain part of speech and display of emotions. This Shelf component will enhance existing automatic feature extraction and emotion recognition concepts with the introduction of measures of confidence on both the feature values and the final emotion estimation.

### **Emotional Natural Language Generation**

This component is responsible for generating natural language without disregarding the affective aspects of a conversation. It is based on an annotated corpus consisting of sentences that present typical expressions that will be used in a conversation. The corpus has to be annotated with the categories and topic that the sentences are about and also with the emotional state that they denote, in order to be used later in the application.

### **Affective Music Synthesis**

Computer music usually sounds mechanical. Therefore, it would increase



## AN INTEGRATED APPROACH

the quality of experience if musicality and music expression of virtual actors could be enhanced according to the user's mood, giving resonance to it and resulting in more enjoyment. This component will synthesise and adjust music according to the interpretation of the user's emotion performed by the framework.

### Emotional Attentive ECA

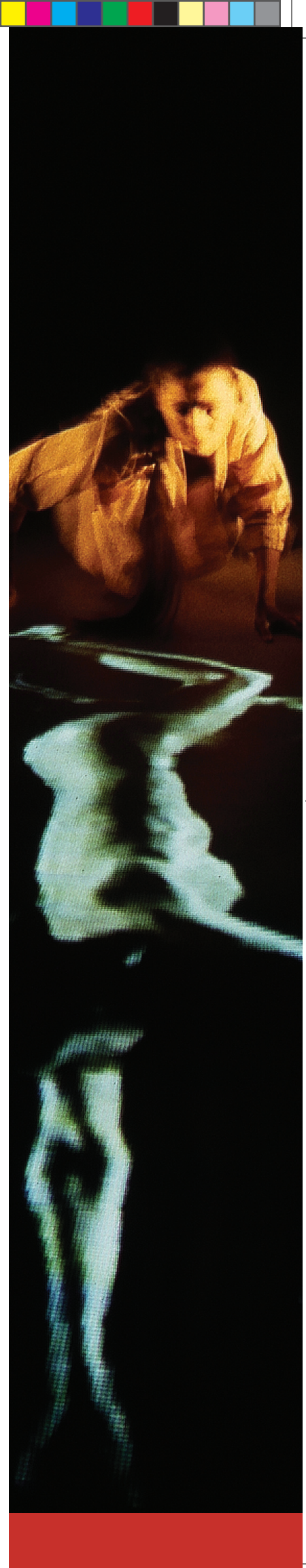
The work here follows two core themes of an emotional ECA and social attention. CALLAS is developing an ECA that is sensitive to and expressive through aspects relating to emotion and attention. The technical approach to these capabilities is split into three primary modules focusing respectively on three core capabilities:

- emotional communication through gesture, facial expression, gaze, body;
- emotional expression in a social context by blending or masking emotions;
- modelling perceptual-attentive social behaviours that are a basis for interaction, such as mutual, joint and shared attention.

### The Framework

The CALLAS Framework is being designed as a software infrastructure that will allow a number of Shelf components to work together to build specific end-users applications in the field of Digital Art and Entertainment. Most of the Shelf components are able to gather not only what a spectator asks to the system or decides to communicate to it, but also information related to their emotional state. On this basis, there are many projects that are studying the process of multimodal interaction (e.g. fusion at data, feature or semantic level) and the extraction and categorization of emotional states, but CALLAS is the first project aiming at understanding how to work with affective input on multimodal fusion. The CALLAS framework aims at collecting all the partial displays of the affective involvement of spectators, at merging them together in order to really deduce what the audience is feeling and as a consequence, at producing a proper affective and multimodal response. What makes CALLAS unique is the combination of active and passive modalities. In CALLAS, the emotional aspects of the interaction are elicited from a semantic representation of what is conveyed through both active and passive modalities. This means that in CALLAS the fusion is performed at the semantic level and that the multimodality is adopted to perform a new human computer interaction model, based on the emotional involvement of the audience.

The CALLAS Framework is finally a software infrastructure making life easier for digital art and entertainment application developers. When developers decide to design and implement a multimodal interactive application, they can use a suite of open-source and interoperable toolbox and software sub-assemblies to save time in the application development.





## AN INTEGRATED APPROACH

### The Showcases

Traditionally, human-computer interaction has concentrated on the usability of applications. That is, the objective was to design applications that enable a user to perform a specified task as effectively and efficiently as possible, where the applications were primarily regarded as tools. In contrast, designers of affective interactive applications rather focus on the user's experience, where these applications are able to analyse and render emotions as part of an interactive system. The effectiveness of the CALLAS approach in pursuing the aforementioned objectives will be validated by developing significant research prototypes (or Showcases) in three major fields of Digital Arts and Entertainment:

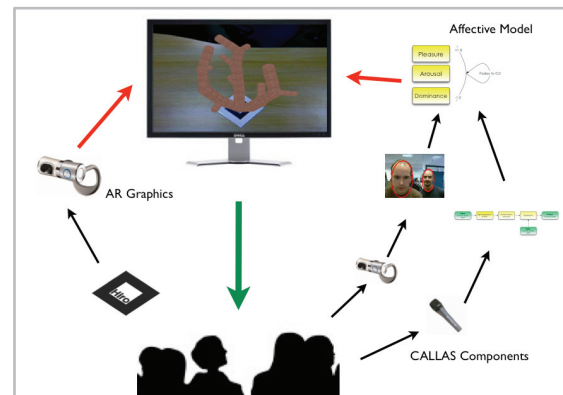
#### Augmented Reality for Art, Entertainment and Digital Theatre

The objective of this showcase will be to demonstrate how on-the-fly detection of the mood and the affectiveness of the people involved in a AR installation or in a live audience (directors, actors, audience) can improve the naturalness of the experience.

E-Tree (Emotional Tree) is a CALLAS showcase that integrates an Augmented Reality (AR) environment with CALLAS affective input components, and displays a dynamically generated piece of digital art—a virtual tree—whose growth and evolution reflects the perceived affective response of spectator(s). The showcase can support the exploration of “affective feedback loops” where a participant's response to dynamic artwork determines changes and development that occur within the artwork, to which the participant then also responds.

The development of the tree is driven by a dimensional affective model which represents the combination of affective input that has been received in the showcase. This model has three dimensions: Pleasure, Arousal and Dominance, which run on a scale of -1.0 to +1.0. Pleasure and arousal are similar to more popular models of valence and arousal, but the additional dominance dimension helps to distinguish between otherwise similar dimensions that differ in the centre of control (such as fear and anger). Input from each component is translated into a 3-dimensional value which is combined with the existing model values. The size and rate of affective input determines how quickly the model takes on a new value. The changing state of the model represents the overall “mood” of the interaction, and with no input, the model slowly moves towards a neutral state, leaving the E-Tree as a record of the history of affective input.

The development of the E-Tree is controlled by rules that define branch growth and branching. Rules are chosen using weighted probabilities and parameters that are determined by the affective model. Positive values will assign a higher probability to growth and branching, while negative values assign



*After an original concept by Digital Artist Maurice Benayoun*



## AN INTEGRATED APPROACH

more probability to no branching, slower growth, and off-axis branching and growth angles. Additionally, the parameters of existing rules are adjusted by current affective input. This means both the current “mood” and the mood that existed when a branch was created contribute to the look of the tree.

### Interactive Installations for Public Spaces

This showcase demonstrates multimodal affective applications for public places. Our aim is to enhance the experiences of visitors or local groups as they visit festivals and events, or contemporary art museums. We believe that a lot can be done to improve public spaces, using novel technologies to change the way they are perceived and let people re-configure the spaces they inhabit and visit. Among the key ideas we hold is that the technologies should be widely accessible and allow people to bring in and create their own content. Demonstrations should provide context and emotion awareness; that is being adaptive to the emotional state or use situation. Creating new kinds of interactive media also affords state-of-art research on affective interaction, which is in many regards an unexplored topic. Finally, the design of novel multimodal applications in co-operation with professionals of various arts is expected to extend our work to fine arts. This can allow involved artists to reach new dimensions of expression with the help of CALLAS technologies.

The prominent application of this showcase is PuppetWall. It is an application derived from the puppeteering metaphor, intended to transfer the concept of puppet theatre into a digital domain with some novel characteristic:

- multimodal interaction; system considers spoken, gestural and bodily input;
- emotional intelligence; detected emotional states of a user can be used to dynamically change the system;
- flexible content; users are not tied to any fixed configuration of characters, objects and backgrounds, but can manipulate them and import their preferred objects;
- collocated co-creation; the system is designed for multiple users that establish dialogue between themselves.

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

### Next-Generation Interactive Television

Showcase 3 will provide a proof-of-concept of a next generation interactive television home platform, which utilises multimodal input to infer higher-level user inputs of affective state. The home platform will make use of an ECA to convey affective content related to both the desired outcome of the broadcast content and the user’s perceived viewing experience. The objectives are:

- Identify relevant input modalities and expressive features to be used within them.
- Integration of CALLAS components and ECA with existing HTN-based interactive storytelling technology, so that the ECA acts as a co-spectator, with expressive behaviour that matches the intention of the story content.







## AN INTEGRATED APPROACH

- Generation of appropriate affectively-driven interactive story content.
- Evaluation of user behaviour in this setting and using the results to refine the concept to maximise user experience.

The core of the system consists of an interactive storytelling engine integrated with an affective-enriched Embodied Conversation Agent (ECA) as co-spectator. User interaction is achieved through gestural actions and speech. Affective input is based on Emotional Speech Recognition as well as keyword spotting, and from the gesture as symbolic and deictic.

The co-spectator functions as a means of conveying desired affective state through its reaction to dramatic events in the story content. The affective state of the viewer is inferred from speech and gesture. The content follows the (pseudo-emergent) interactive plot within the genre of comedic horror - the interactive storytelling engine allows us to define the interactive plot in terms of potentially conflicting goals of entities. It involves a young lady who is flat-sitting for a friend. It turns out that strange things are happening inside the flat.

The incentive behind the use of comedic horror is to appeal to and elicit derivatives of the most basic emotions: fear and humour. These states are probably the easiest to automatically discern. Moreover, they are known to be the most easily elicited states from entertaining filmic drama.

The interactive story makes use of devices (props and events) which are typical of the horror genre, and which border on the cliché in order to mock and hence evoke an atmosphere of humour. The devices, however, also have a sinister twist providing the opportunity to rapidly switch from humour and facilitate mild variants of fear or disgust.

Thus, we have a scenario that has been constructed to not only entertain its users, but also to provide a basis for fundamentally testing the affective technologies involved. Each of these showcases will encompass application scenarios each related to the use of affective, multi-modal interfaces in new media domains which will address both short and long term perspectives.



## What is the CALLAS Community?

The CALLAS Community is a global and collaborative aggregation of users including digital art professionals and researchers, individuals, organisations, SMEs (small and medium enterprises), as well as any individuals interested in the latest trends in New Media and affective interface technology and applications.

To be more precise, CALLAS Community members are engaged to be involved in all phases of the project, from specification and evaluation to collecting valuable user feedback. Through Community research and collaboration, the Consortium can explore the potential of its primary user basin and tailor specific services to fit specific users needs. Members are encouraged to maintain a personal homepage, autonomously post papers and publications, publicise and attend current events, participate in the CALLAS training programme, share ideas on current trends, new technology and applications and keep up to date with project initiatives and developments.

## Why join?

The benefits of the Community are mutual for both the project and the members who actively participate. The Community is collaborating to provide CALLAS with a primary user base which helps direct technological developments within the project and have direct access and initial adoption of CALLAS technologies and solutions.

The choice of open source software and open standards reflects the open character of the project. By registering as members of the CALLAS Community, software developers and researchers will find an advanced, lively, friendly environment to be updated on the state of the art on New Media and affective interfaces. Application developers, providers, and final users will gain insight into one of the most rapidly developing fields, and influence a large-scale, exciting project meaning to change the way we all look at the future of Digital Media, Art and Entertainment.

Focused events are being organised internationally to display technological developments during the life cycle of the project.

## Get involved!

To become a member of CALLAS, participate and take advantage of member services you are invited to join the Community free of charge with an on-line membership registration form at the project's communication website: **[www.callas-newmedia.eu](http://www.callas-newmedia.eu)**

You can join the Community as an organisation or an individual, providing a short description of your professional activities which will be included on your own profile on the website that can include a link, logo and contact.







## BEHIND THE SCENES

### The CALLAS Consortium

- Engineering Ingegneria Informatica S.p.A. - Italy (Co-ordinator)
- VTT Technical Research Centre of Finland
- British Broadcasting Corporation - United Kingdom
- Metaware S.p.A. - Italy
- 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 / National Technical University of Athens - Greece
- Faculty of Engineering, Mons (FPMs) - 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

### Project governance

The CALLAS consortium consists of nineteen partners, whose skills come from diverse multicultural fields in a well-concerted and balanced way such as: research and academic excellence, large industrial players and innovation led SMEs. This balance is essential to successfully implement CALLAS' ambitious goals. Moreover, the CALLAS Consortium represents nine European Community countries plus one international partner, creating a valuable mixture of cultural perspectives as well as a critical mass to maximise project impact.

### The project management structure

The Integrated Project Coordinator (IPC) is Massimo Bertoncini, working for Engineering Ingegneria Informatica S.p.A.

In addition to the CALLAS IPC, CALLAS' management structure is also composed of Workpackage Directors, who will supervise the seventeen Workpackages, in order to ensure coherence and consistency between them. The seventeen Workpackages have been clustered in functional components called sub-projects, which reflect the specific structure of the Integrated Project.