Emotional Multimodal Interfaces for Digital Media: the CALLAS challenge

Massimo Bertoncini[1], Marc Cavazza[2]

[1] Engineering Ingegneria Informatica, R&D Lab, Via San Martino della Battaglia 56, 00185 Roma, Italy, massimo.bertoncini@eng.it
[2] University of Teesside, School of Computing, Middlesbrough, TS1 3BA, United Kingdom, m.o.cavazza@tees.ac.uk

Abstract. Emotional multimodal interfaces aim at achieving the highest level of naturalness in human-computer interaction. One of the main challenges for CALLAS European R&D project 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. 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. Such key aspects are currently investigated within the CALLAS project in the specific area of Art and Entertainment applications.

Keywords: Affective Interfaces, Multimodal Interaction, Emotion detection.

1 Introduction

People naturally communicate combining gestures, movements, speech and 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. 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).

2   The CALLAS Project Objectives

The CALLAS is addressing the following high-level objectives:

1)   To advance the state-of-the-art in Multimodal Affective Interfaces by:
   i)   developing new **emotional models** that are able to take into account a comprehensive user experience in Digital Arts and Entertainment applications and
   ii)   ii) new **modality-processing techniques** to capture and elicit these new emotional categories

   2) To research, develop, and integrate advanced software components, tailored to the processing of individual modalities supporting the semantic recognition of emotions, making them available through a "living" repository, called the CALLAS "shelf".

   3) To establish 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 will eventually not require a deep understanding of the theories of Multimodality.

   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**
   - Interactive Installations for **Public Spaces**
   - Next-Generation **Interactive Television**

   CALLAS also aims to ensure the sustainability and the replicability of its technology results. This will be addressed mainly by supporting Technology Transfer, in particular towards SMEs operating in the new media sector, whether these SME are involved in Digital Arts and Entertainment or are innovative technology spin-offs.

3   The Emerging Vision for Digital Media

In recent years, Digital Media Entertainment has developed largely in terms of both technical sophistication and richness of digital content. On the other hand the recent progresses in emerging technologies, such as ubiquitous computing, augmented and virtual reality, human-computer Interaction, and context and location-awareness, are

paving the way for a profound paradigm shift towards embracing users' natural behaviour as the centre of human-computer interaction.

The term New Media or Digital Media is used to refer to all means of digital communication that evolved from traditional media (text, video, sound/music) sometimes through a "remediation" process [1]. Most New Media are actually *interactive*, and rely on digital content for which user interaction plays a central role. One long-term endeavour is for them to support interaction with sophisticated cultural content: the resulting media would constitute future Digital Arts and Entertainment systems. Concrete instances of such new media include Digital Theatre, Mixed Reality Art, Interactive TV, as well as Mobile and Ubiquitous systems supporting emergent interactive storytelling, which are the application areas targeted by CALLAS.

Interaction with cultural content determines a specific type of user experience that can be described in terms of a wide range of users' emotions and feelings. There have been significant developments in affective interfaces in the past years, yet their emphasis has been on the role of emotions in communication. To take its full dimension, research in emotional interfaces has been extending to the context of interactive media.

Within such a framework a core set of different streams activities have been carrying out aimed at:

1) To advance the understanding of emotional interaction in the context of Arts and Entertainment and develop corresponding models of Multimodal processing. This issue is taking into account the semantic aspect of emotion and will aim at handling extended not-Ekmanian emotional categories to better describe the user experience in Arts and Entertainment systems, and is improving the performance of basic emotional components for existing modalities at input level (speech recognition, gesture recognition, body motion tracking, face detection, sound capture, haptic tracking, natural language processing) when an extended set of semantic categories is going to be used. This has been made possible thanks to the encapsulation of Individual components for the processing of single modalities in software ready-to-use components, whose collection constitutes the CALLAS *shelf*. These are meant to evolve throughout the project with the progression of research into the processing of individual modalities;

2) To design and implement emotional components for new modalities, such as emotional Natural Language Processing and to extend the emotional synthesis by including acoustic, visual and linguistic features, while designing and applying suitable software engineering methodologies allowing a rapid development of multimodal affective applications through the adoption of a suitable Software Architecture approach (called the CALLAS *Framework*) enabling different modes of integration, as required by various applications, offering pre-assembled, re-usable, and semantic fusion components;

3) To develop prototypes (CALLAS *showcases*) in the domains of Digital Art and Entertainment, which will complement other research activities throughout their life cycle, by providing context and user data, artistic requirements, as well as evaluation settings, demonstrating the validity of the CALLAS approach for the advance in the state of the art of multimodal affective interfaces and raising awareness about the CALLAS technologies through the setting up of an open source user-community

whose members will be given free access to the CALLAS technologies including both application providers and final users;

4) To promote Technology Transfer, in particular towards SMEs operating in the new media sector, whether these SMEs are involved in Digital Arts and Entertainment or are innovative technology spin-offs, reducing substantially the cost for developing affective multimodal interfaces and creating an effective sustainability model where new revenue streams, based on the delivery of new services based on the CALLAS technologies and showcases, are emerging.

References

1. Bolter and Grusin, Remediation: Understanding New Media, MIT Press 1999