Universidade de Aveiro Departamento de Electrónica, Telecomunicações e Informática,
2018

Tiago de Figueiredo Henriques

A systematic approach for the integration of emotional context in interactive systems

Uma abordagem sistemática para a integração de contexto emocional em sistemas interativos

Tiago de Figueiredo Henriques

A systematic approach for the integration of emotional context in interactive systems

Uma abordagem sistemática para a integração de contexto emocional em sistemas interativos

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia de Computadores e Telemática, realizada sob a orientação científica do Doutor Samuel de Sousa Silva, Investigador do Instituto de Engenharia e Eletrónica e Informática de Aveiro e da Doutora Susana Manuela Martinho dos Santos Baía Brás, Investigadora do Instituto de Engenharia e Eletrónica e Informática de Aveiro.

o júri / the jury

presidente / president	Prof. Doutora Maria Beatriz Alves de Sousa Santos Professora Associada com Agregação da Universidade de Aveiro
vogais / examiners committee	Doutora Ana Correia de Barros Investigadora no Fraunhofer AICOS
	Doutor Samuel de Sousa Silva Investigador do Instituto de Engenharia e Eletrónica e Informática de Aveiro (orientador)

agradecimentos / acknowledgements

Em primeiro lugar, quero agradecer aos meus orientadores, pelas excelentes condições de trabalho fornecidas (laboratórios IEETA), material disponibilizado (computador e dispositivo móvel) e, especialmente pela oportunidade de realizar esta dissertação.

Ao Doutor Samuel Silva, um especial obrigado pelo apoio e interesse constante, pelas diversas sugestões de detalhes técnicos de implementação, de reescrita do documento da dissertação e também pela ajuda fundamental na escrita do artigo científico. À Doutora Susana Brás, a principal impulsionadora pelo meu interesse nesta dissertação e na área de "affective computing", um obrigado pelo apoio, disponibilidade e ajuda na escrita do artigo científico.

De salientar ainda, a Prof. Sandra Soares, do departamento de Educação, pelos conhecimentos transmitidos na área da psicologia e pelo contributo dado no artigo científico, o Prof. Doutor António Teixeira, pelas sugestões de melhoria no artigo científico, e o Prof. José Maria Fernandes, pela disponibilização de material de trabalho.

Ao Nuno Almeida, um sincero obrigado, pela paciência e pelos conhecimentos transmitidos no bom uso do módulo da framework multimodal.

Aos meus companheiros de laboratório de cima, um obrigado pelo bom ambiente de trabalho, pela companhia nos cafés e pelas conversas. Aos meus companheiros do laboratório de baixo, um obrigado pelos momentos de maior descontração e pela banda sonora "discutível" que por vezes tocava ali ... Um especial agradecimento para o meu caríssimo amigo João Paulo Melo, pelo apoio, pelas risadas, pela companhia nas diversas idas de descompressão ao ginásio, pela companhia durante os almoços e pelas noitadas (não foram muitas) no laboratório do IEETA.

A todos os participantes envolvidos no meu caso de estudo, um agradecimento pelo feedback e pelo tempo dispendido a realizar as tarefas que me possibilitaram adquirir os dados para as minhas conclusões.

Por fim, quero dedicar este trabalho às pessoas mais importantes da minha vida. Os meus sinceros agradecimentos, aos meus pais, ao meu irmão e aos meus avós por nunca deixarem de acreditar em mim. Nunca vos vou conseguir agradecer por tudo o que fizeram e fazem por mim, por isso, espero que um dia tenham orgulho em mim, tal como eu tenho em vocês.

palavras-chave

Computação afetiva, interação multimodal, desenvolvimento centrado no utilizador, ASD, modalidade afetiva, sistemas que consideram emoções, Affective Spotify, MoodDiary, modalidades.

resumo

Em sistemas interativos, ter conhecimento do estado emocional do utilizador não é apenas importante para perceber e melhorar a experiência global do utilizador, mas também é de extrema relevância em cenários onde tal informação pode impulsionar a nossa capacidade para ajudar os utilizadores a gerir e a expressar as suas emoções (por exemplo, ansiedade), com um grande impacto nas suas atividades do dia-a-dia e como estes utilizadores interagem com outros. Embora exista um elevado potencial em aplicações que tenham em consideração as emoções, inúmeros desafios impedem a sua disponibilidade mais ampla, por vezes resultante da baixa natureza translacional da pesquisa feita na área de computação afetiva, outras pela falta de métodos que permitam uma fácil integração de emoções nas aplicações. Embora já múltiplas ferramentas tenham sido propostas para a extração de emoção a partir de um conjunto de dados, como texto, áudio e vídeo, a sua integração ainda requer um esforço de desenvolvimento considerável que necessita de ser repetido para cada aplicação implementada. Tendo em conta estes desafios, apresentamos uma visão conceptual para a consideração de emoções no âmbito de sistemas de interação multimodais, propondo uma abordagem desacoplada que também pode fomentar uma articulação com a pesquisa na área da computação afetiva. Com esta visão em mente, desenvolvemos uma modalidade afetiva genérica, alinhada com as recomendações do W3C para arquiteturas de interação multimodal, que possibilita o desenvolvimento de aplicações que têm em consideração as emoções, mantendo os sistemas interativos desenvolvidos (e os programadores) completamente dissociados dos métodos computacionais afetivos considerados. De forma a complementar o trabalho realizado, e de modo a ilustrar o potencial da modalidade afetiva proposta na apresentação de contexto emocional em cenários interativos, foram instanciadas duas aplicações demonstrativas. A primeira, permite uma interação multimodal com o Spotify, considerando o contexto emocional do utilizador para adaptar a música que é reproduzida. O segundo, Mood Diary, serve como uma prova de conceito de como uma aplicação que considera as emoções pode ajudar os utilizadores na compreensão e expressão de emoções, uma característica potencialmente relevante para aqueles que têm necessidades especiais, como é o caso, dos perturbações do espectro do autismo.

keywords

Affective computing, multimodal interaction, user-centred design, ASD, affective modality, emotionally-aware systems, Affective Spotify, MoodDiary, modalities.

abstract

In interactive systems, knowing the user's emotional state is not only important to understand and improve overall user experience, but also of the utmost relevance in scenarios where such information might foster our ability to help users manage and express their emotions (e.g., anxiety), with a strong impact on their daily life and on how they interact with others. Nevertheless, although there is a clear potential for emotionally-aware applications, several challenges preclude their wider availability, sometimes resulting from the low translational nature of the research in affective computing, and from a lack of straightforward methods for easy integration of emotion in applications. While several toolkits have been proposed for emotion extraction from a variety of data, such as text, audio and video, their integration still requires a considerable development effort that needs to be repeated for every deployed application. In light of these challenges, we present a conceptual vision for considering emotion in the scope of multimodal interactive systems, proposing a decoupled approach that can also foster an articulation with research in affective computing. Following this vision, we developed an affective generic modality, aligned with the W3C recommendations for multimodal interaction architectures, which enables the development of emotionally-aware applications keeping the developed interactive systems (and the developers) completely agnostic to the affective computing methods considered. To support the work carried out, and to illustrate the potential of the proposed affective modality in providing emotional context for interactive scenarios, two demonstrator applications were instantiated. The first, enables multimodal interaction with Spotify, considering the user's emotional context to adapt how music is played. The second, Mood Diary, serves as a proof-of-concept to how an emotionally-aware application can support users in understanding and expressing emotion, a potentially relevant feature for those suffering from conditions such as autism spectrum disorders.

Contents

List of Figures

List of Tables

*

List of Acronyms

IM Interaction Manager

API Application Programming Interface

OS Operating System

TTS Text-to-Speech

HTTP Hypertext Transfer Protocol

ASD Autism Spectrum Disorder

EDA Electrodermal Activity

ECG Electrocardiography

HR Heart Rate

EMG Electromyography

SDK Software Development Kit

AI Artificial Intelligence

W3C World Wide Web Consortium

UCD User-centred Design

CDC Centers for Disease Control and Prevention

PDD-NOS Pervasive Development Disorder-Not Otherwise Specified

AAC Alternative Communication

PECS Picture Exchange Communication System

SCXML State Chart XML

IDE Integrated Development Environment

IEETA Institute of Electronics and Informatics Engineering of Aveiro

REST Representational State Transfer

JSON JavaScript Object Notation

BaaS Backend as a Service

PSSUQ Post-study System Usability Questionnaire

Chapter 1

Introduction

The use of technology has been steadily increasing over the past years and it has moved away from merely making our lives more convenient to potentially helping and improving people's lives, in a wide range of domains. Today, we live surrounded by many devices that play a key role in many tasks, whether in industry, education, medicine, or even in simple tasks or moments of our day-to-day routine. In this context, it is extremely important to propose interactive systems that provide a more natural interaction and, to different extents, try to adapt to the user considering his abilities and motivations, a feature with high relevance for all, but particularly relevant when the target audience has special needs.

A system can adapt to particular contexts in a variable number of ways, whether by design, or explicit choice of the user, but a certain level of independence in sensing when and how to adapt, based on the current context, potentially allows a more natural experience and a faster response to tackle change. In this regard, we may expand our discussion and embrace the concept of implicit interaction [1]. One simple and very well known example of implicit interaction, i.e., inputs provided to a system that were not explicitly given by the user, is location. Obtained implicitly and presented in most mobile devices, it can contribute to the definition of the current context, enabling adaptation, for instance, of how applications respond (e.g., disabling sounds when entering the company's meeting room) or provide information (e.g., showing the closest coffee shops) to the user.

In line with these considerations, more of these implicit interactions, reflecting aspects of the users' experience and context, can and should be considered, with the emotional state of the user assuming particular importance, considering the critical role it plays in modulating human behaviour [2].

1.1 Motivation

Emotions rule our daily life experiences and much of our motivated behaviour. For example, the decisions we make and the activities and hobbies we choose are widely dependent on the emotions we experience (e.g., happiness, sadness, fear and anxiety). Importantly, considering the individual's subjective emotional experience (e.g., likes and dislikes), which reflect an explicit dimension, emotions also encompass implicit responses, such as those arising from the Autonomic Nervous System activity (e.g., high heart rate level or electrodermal activity). The evaluation of such implicit measures of emotion can be performed continuously, with minimally invasive equipment, yielding a great opportunity to monitor emotional states as an

input to the development of more flexible and adaptable systems [3].

In this regard, it is important to take emotions into account when a high adaptability to the user is required or when we want to deal with the behavioural aspects of users, in different situations, resulting in a greater insight on the user's context and experience, with a clear relevance across several domains (e.g., learning environments [4], comfort assessment in human building interaction [5]). Additionally, the integration of emotions in interactive systems can play a key role in aiding users understand [6], manage and communicate their emotional state, which then can result in significant improvements in their quality of life. An example in which implicit measures of emotions may provide an added value is the case of individuals who lack the ability to express emotions, such as in Autism Spectrum Disorder (ASD) [7].

1.2 Challenges

Despite the evolution of affective computing [8] and its potential importance, the consideration of the user's emotional context, in interactive systems, is still far from its full potential, and that is the main reason why only a few systems appeal to the user emotional state to foster increased adaptability [9]. The improvement of this panorama requires tackling several challenges, at different levels.

From the perspective of interaction design and development, it is important that deploying emotionally-aware applications does not entail mastering all the technology involved in detecting emotions, which can work as a barrier for its wide adoption. Instead, developers should have access to reusable components that hide this complexity while enabling access to the full range of options available. Additionally, regarding research in affective computing methods, the field would profit from a translational approach that could more rapidly make the transition between the lab's machine learning methods and their deployment in the scope of interactive ecosystems.

1.3 Objectives

Considering the challenges previously identified, our overall goal is to contribute to improve the current state-of-the-art in considering the user's affective context in interactive systems. In this regard, our work specifically aims to:

1. Conceptualize an approach for the systematic integration of emotional context in interactive systems that does not entail mastering all the technology, by the developer;

2. Propose a first implementation of the main modules required to support the conceptualized approach encompassing support for different kinds of devices and operative systems;

3. Create a proof-of-concept application that demonstrates the integration and potential of the proposed approach.

The work to be accomplished should be supported by continuous validation of the proposed features following an iterative design and development approach.

1.4 Publications and Presentations

The work developed in this dissertation has already been partially published in the following article:

- Tiago Henriques, Samuel Silva, Susana Brás, Sandra C. Soares, Nuno Almeida, António Teixeira, "**Emotionally-Aware Multimodal Interfaces: Preliminary Work on a Generic Affective Modality**", Proc. 8^{th} International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, Thessaloniki, Greece, June, 2018 [online]

This publication presents a conceptual vision to tackle the challenges identified and a first instantiation of an affective generic modality along with a prototype application, enabling multimodal interaction with Spotify, which illustrates how the modality can provide emotional context in interactive scenarios.

Additionally, the work has also been publicly presented at the students@DETI event and the corresponding poster is provided in Appendix 5.2.

1.5 Overview

The remainder of this dissertation consists of four chapters, organized as follows:

- **Chapter 2** focuses on background and related work, introducing the work on the field of Affective Computing, Multimodal Interaction and provides a brief overview on the aspects regarding user-centred approaches to application design and development to contextualize the methods followed later on. Regarding the affective computing field, we cover some emotional toolkits and services that could be used to extract emotional context from users' data. This chapter also covers autism spectrum disorders to provide the reader with a closer look into the characteristics and needs of a user group that can potentially profit from an adaptation of applications based on the affective status.

- **Chapter 3** starts by presenting our conceptual vision regarding how emotion can be integrated in a multimodal interactive scenario and progresses by describing the overall architecture and main blocks for a core element in such vision: the affective modality. Then, it addresses the technical aspects that need to be considered in order to develop a generic affective modality, the main implementation steps, core blocks details and outlines the backend services and methods that support the modality.

- **Chapter 4** presents two applications proposed to support the developed work and illustrate its potential, showing how the developed affective modality enables these applications to integrate user's emotions during its use.

- Finally, **Chapter 5**, the conclusion, discusses the outcomes of our work, presents some insights about future work to evolve it, and identifies new paths of development made possible by our proposals.

Chapter 2

Background and Related Work

Considering the goals established for this work, two areas are deemed relevant for its progression: ***affective computing***, and ***multimodal interaction***. In what follows, we provide a brief overview of the key aspects for the current state-of-the-art and related work on the field of affective computing and summarize relevant information regarding multimodal interaction, particularly regarding how to support its development.

Considering the nature of our aims, and the associated time constraints, we do not foresee a strong emphasis on end-user applications beyond proofs-of-concept. Nevertheless, we aim to illustrate how our proposals can be considered in the wide scope of interactive systems design. Therefore, we provide a brief introduction regarding a user-centred design and development approach, considering users and their context when developing applications.

The chapter ends with a characterization of autism spectrum disorders (ASD), describing the characteristics of an audience that can take advantage of an adaptation of applications based on the affective context.

2.1 Affective Computing

The field of affective computing studies and proposes systems able to recognize, interpret, process and simulate emotions [10]. This research field has evolved substantially in recent years in providing methods and mechanisms able to account for the emotional state of the users via several inputs, such as physiological signals (e.g., *electrodermal activity (EDA), heart rate (HR))*, speech data, text and facial expression analysis, amongst others (Figure 2.1). The integration of several inputs allows for more accurate estimates and interpretation of the user's emotional state, which will then serve as basis for adapted responses by applications (thus resulting in more adaptable and interactive systems), made possible by the use of sophisticated machine learning methods. Given the implications of emotions in determining our behaviour, the computing ubiquity and the wide dissemination of technology in modern times, Affective Computing attracts attention in many areas, ranging from the industry of video games, automobiles, marketing, to mental health applications [11].

The implicit manifestations of emotions can be captured in several ways using, for example, video camcorders, able to record emotional facial expressions, microphones to record vocal inflection changes, skin-surface to sense muscle tension, heart-rate variability, skin conductivity and blood-glucose levels [12]. Facial expressions, in particular, represent a fast and easy way to access emotional responses. In fact, inferring the emotional state by means of

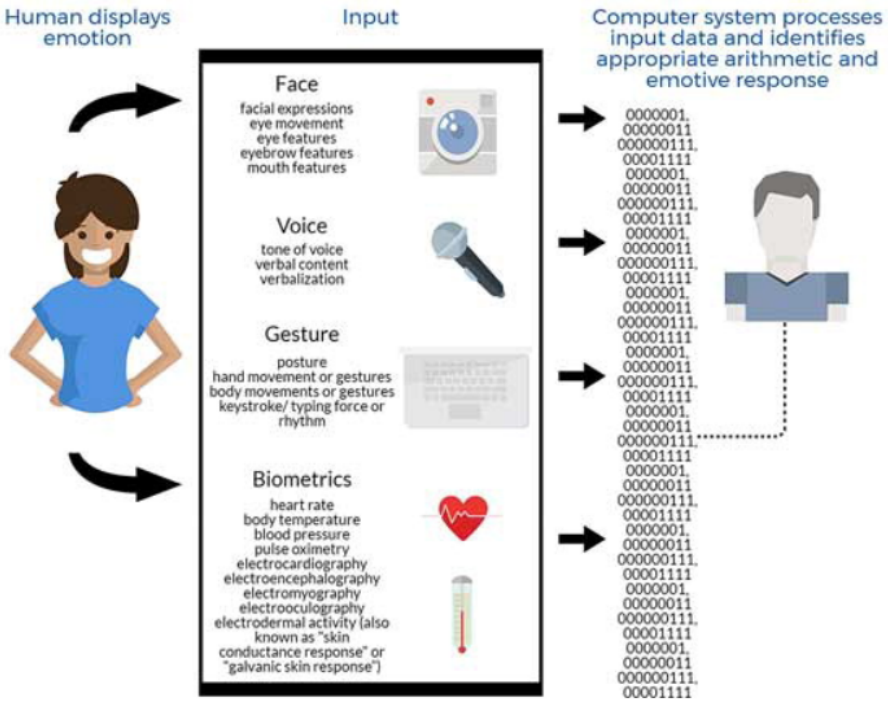

Figure 2.1: Basics of Affective Computing process; Humans display body signs (face, voice, gesture, biometrics), that are analysed by a computer system that identifies the appropriate emotional response. (source: `goo.gl/9NDfN2`).

---

the facial expression, possible while using a webcam or an image file, is rather common, with the available data showing high accuracy rates in the recognition of emotional responses from these features [13].

2.1.1 Emotional/Behavioural Profile and Importance

The innate response to stimulus and situations (e.g., running away when perceiving threat or danger) produces a response by our body: as a physiological response or as a behavioural response. The characterization of a person reaction will allow to define a person, his/her feelings, and preferences. Therefore, it is crucial to identify emotions in order to adapt contents to the person and situation. To accomplish such goal, the psycho-physiological profile of each person should be built by the processing of physiological signals (*e.g.,* ECG, EMG) and behaviour responses (*e.g.,* facial expression, posture). Figure 2.2 represents an example of a physiological response, fight-or-flight, that usually occurs in moments of high stress or when in the presence of something that is frightening, physically or mentally. This response triggers two different endings: stay and fight with the threat or run away to safety.

These responses can be used to measure attention, arousal, relaxation and emotion. This dissertation focuses on this last parameter ***(the emotion)***, and its importance in the development of interactive applications, specially those designed for people with special needs, particularly those with difficulties in expressing and understanding emotion. In these particular cases, carers (e.g., teachers or educators) can profit from this knowledge to act or better understand their behaviour; or, in alternative, interactive systems could adapt to avoid, for

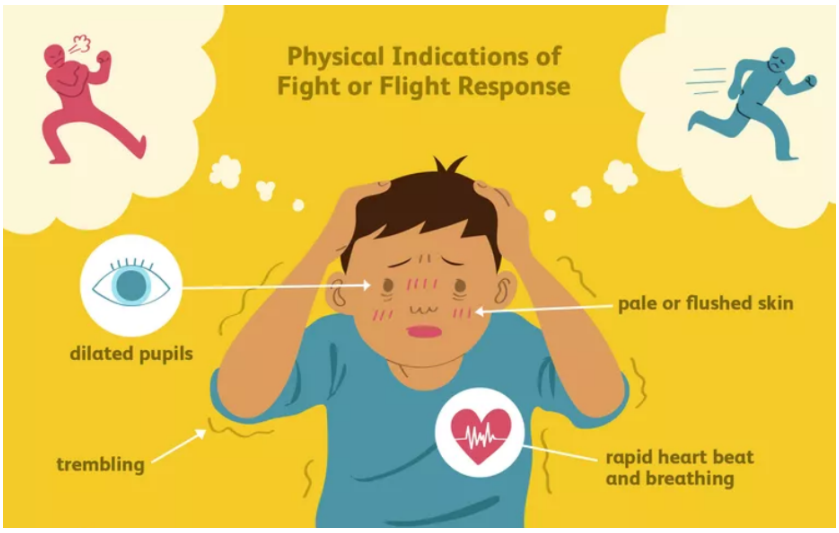

Figure 2.2: The fight-or-flight response was originally proposed by an american physiologist named Walter Cannon in 1920. (source: `goo.gl/KU7oRd`).

instance, anxiety attacks.

2.1.2 Emotional Toolkits and Services

The importance and interest on the affective computing field, in a wide range of contexts, has motivated the proposal of several toolkits and services that can be used to extract emotional context from diverse data types such as image (or video), plain-text, and speech audio data (refer to table 2.1 and 2.2 for a list of notable examples.).

Regarding the analysis of spontaneous facial expressions, examples of notable resources are: *Affectiva*[1] and *Sightcorp*[2], working with any optical sensor, camera device or standard webcam; *Microsoft Face API*[3] and *Kairos*[4], additionally enabling face detection, face recognition, face verification and face identification; *Google Vision API*[5] and *Amazon Rekognition*[6] providing an additional set of resources to identify people, objects, text and activities, as well as detect any inappropriate content. The detection and recognition of emotions from any form of text often distinguishes two components: sentiment and emotion analysis, consistent with dimensional and categorical models of emotion, respectively [14]. More specifically, sentiment analysis aims to detect valence (positive, neutral or negative) through text analysis, while emotion analysis directs its focus to the extraction of specific emotions from text, such as happiness, sadness, anger, or surprise. Notable services to extract emotional and language tones from text are IBM Watson's *Tone Analyzer*[7] and the *Text Analytics API* by Microsoft[8].

[1] goo.gl/8Apkfm
[2] goo.gl/xKZgB5
[3] goo.gl/ZaFYgv
[4] goo.gl/DT1hrf
[5] goo.gl/YSq1J2
[6] goo.gl/gRr1KH
[7] goo.gl/jz6Gon
[8] goo.gl/3n8HKH

Name	Main features	Data type	Resource
Microsoft Face API	- Face detection, analysis and emotion extraction; - Methods for face verification, grouping, identification and storage;	Image	API, SDK
Google Vision API	- Face detection, analysis and emotion extraction; - Image content analysis; - Detect inappropriate content; - Detects labels, logos, web entities and text;	Image	API
Affectiva	- Face detection, analysis and emotion extraction; - Emotion recognition API for speech in beta version;	Image, video	SDK
Sightcorp	- Face detection, analysis and emotion extraction; - Clothing colours extraction;	Image, video	API, SDK
Kairos	- Face detection, analysis and emotion extraction; - Methods to face verification and authentication;	Image, video	API, SDK
Amazon Rekognition	- Face detection, analysis and emotion extraction; - Image content analysis; - Detect inappropriate content; - Celebrity recognition; - Detects objects, scenes, activities and text;	Image, video	API

Table 2.1: Summary of the most notable features of the described services/toolkits (part 1).

Name	Main features	Data type	Resource
IBM Watson's Tone Analyzer	- Text detection, analysis and emotion extraction; - Document and sentence level;	Text	API
Microsoft Text Analytics API	- Text detection, analysis and emotion extraction; - Only document level; - Language identification; - Key phrases/words extraction;	Text	API
DeepAffects	- Text detection, analysis and emotion extraction; - Audio analysis and emotion extraction; - Audio de-noising API;	Audio, text	API
Beyond Verbal	- Audio analysis and emotion extraction; - Temper, valence, arousal and emotion group detection; - Progress and summary results;	Audio	API
Empath	- Audio analysis and emotion extraction; - Only summary results;	Audio	API
Vokaturi	- Audio analysis and emotion extraction; - Only summary results;	Audio	API, SDK

Table 2.2: Summary of the most notable features of the described services/toolkits (part 2).

Both API's are deep linguistic analysis tools that can be used to analyse words' relations, phrases and structure to extract information with its integrated sentiment scoring functionality. Our tone of voice and the manner in which we utter words and phrases may also provide important cues to infer an emotional state, but voice-driven emotions detection is still a field with much to explore. *Beyond Verbal's*[9], *Empath*[10] and the *Vokaturi*[11] enable to extract a person's set of emotions and character traits based on raw vocal pronunciations (e.g., from an audio file). One aspect to be taken into consideration is that the processed audio stream must have a significant duration to provide a good accuracy on emotion recognition. *DeepAffects*[12] is a speech and text emotion API, and additionally, enables speaker separation features. On figure 2.3, two screenshots were taken considering Google Vision and Microsoft Text Analytics API web interfaces, where emotional content was extracted from an image file and a piece of text, respectively.

2.2 Multimodal Interaction

Multimodal interaction [15, 16] provides multiple modes of communication between users and interactive systems to foster a more natural and versatile interaction when compared with single-modality interactive systems. Multimodal systems can offer a flexible, efficient and potentially more accessible environment, allowing users to interact using modalities such as speech, gestures, touch, and gaze, and to receive information in a diverse set of alternatives including, but not limited to, speech synthesis, smart graphics or haptic feedback. Adding modalities to an interactive system is not motivated by a need to replace the previous modalities, but rather to complement the system with new types of interaction, potentially widening the audience to a more diverse set of contexts and tasks [17].

2.2.1 Multimodal Adaptive Applications

Research and development in multimodal interactive systems has been increasing, at a fast pace, boosted by a widespread availability of low cost support technologies (e.g., smartphones, interactive gloves, eye trackers), with several commercial products already exploring some of its potential, such as Ford Sync[13], a system integrated in vehicles, controlled by voice, that enables users to make hands-free calls or control the music, for example. Other examples of such technologies include Personal Digital Assistants such as Microsoft Cortana, Siri and Google Now. Figure 2.4 presents the mentioned personal assistants' interfaces.

Additionally, multimodal technologies like Dance Dance Revolution, Wii and Kinect (figure 2.5) have the potential to train and improve different skills on different areas (e.g. body motor development, communication, interaction, learning abilities) for individuals with autism spectrum disorders (ASD), for example [18].

This growth in multimodal interactive systems, it is also visible in the cinematographic and in the interactive video games fields (figure 2.6). At the early 00's, several movies appeared where the use of interactive multimodal systems was clear. Minority Report[14] is one of

[9] goo.gl/2i3gZT
[10] goo.gl/xxQBrR
[11] goo.gl/JMNThL
[12] goo.gl/JzD49v
[13] goo.gl/nAVGSW
[14] goo.gl/Q1MmVr

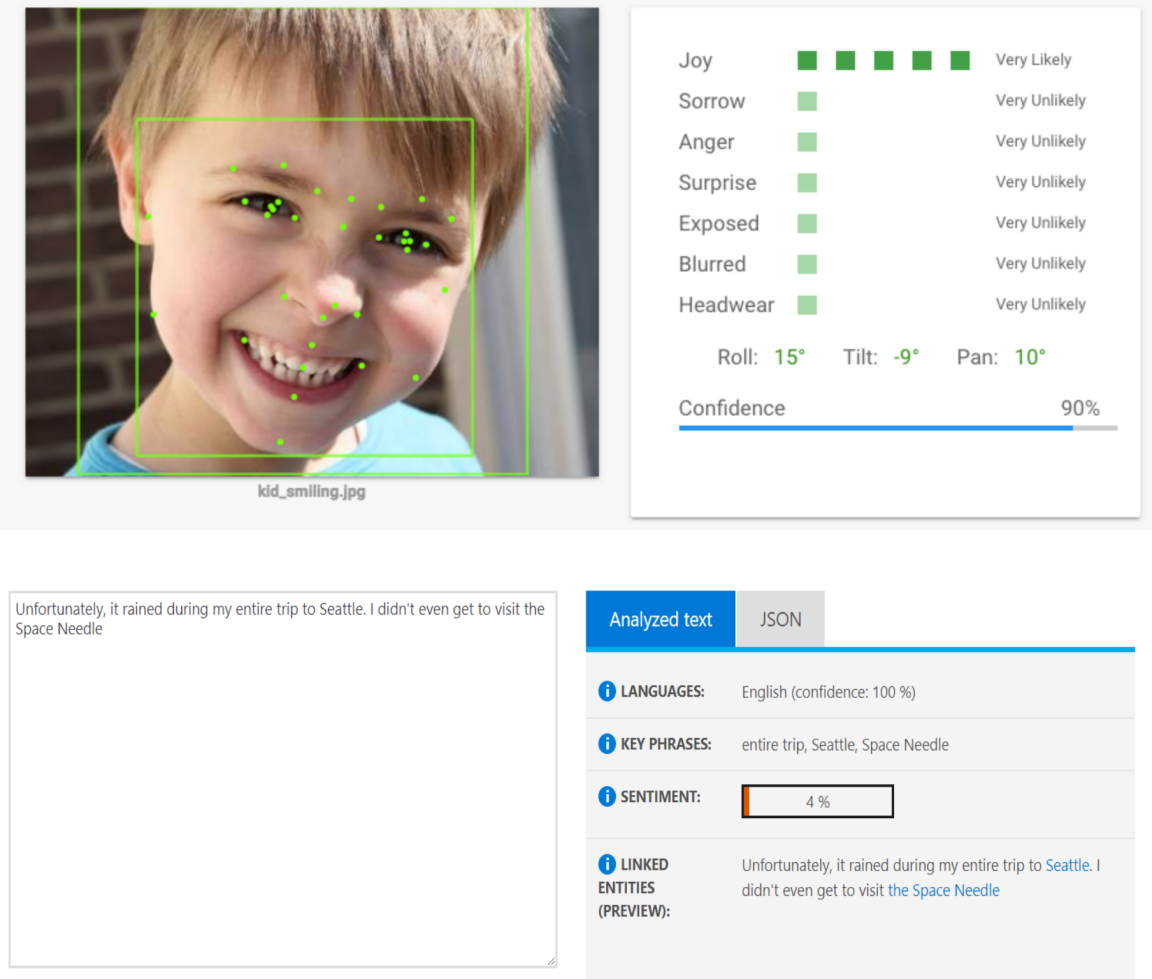

Figure 2.3: On the top-image, an example of an image file being analysed by Google Vision API is presented; A very likely level of joy is being extracted using this service. (source: `goo.gl/e75p8t`); On the bottom-image, an example of a piece of text that is being analysed by Microsoft Text Analytics API is also presented; A sentiment analysis of 4% (from 0 to 100) is extracted, indicating a detection of a negative emotion (e.g., sadness). (source: `goo.gl/Ygbgcn`).

---

the most successful movies that explored this feature, where John Anderton played by Tom Cruise uses gesture-controlled gloves to operate a digital interface. More recently, in the field of video games, the Playstation Move technology appeared, making use of the gestures modality, applied in several videogames, for example the EyePet[15], where a virtual pet interacts with people and objects in the real world.

In recent years of multimodal interaction evolution, a lot of applications that feature multiple interaction capabilities reached the market, however, the support for these multiple modalities encompasses a great development exercise and entails a meticulous expertise on

[15] goo.gl/7aNxDe

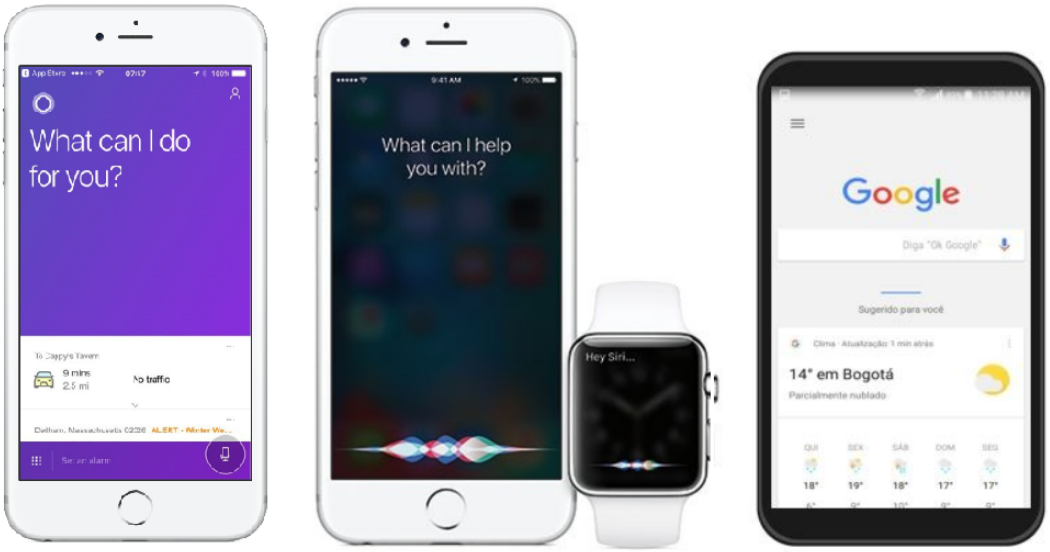

Figure 2.4: Examples of existing personal assistants. From left to right: Microsoft Cortana, Apple Siri and Google Now.

Figure 2.5: Examples of existing multimodal technologies. From left to right: Dance Dance Revolution, Wii and Kinect devices.

---

mastering the different technologies that accommodate each modality. In this regards, several architectures have been proposed to tackle this effort.

2.2.2 Multimodal Architectures

According to the literature [16], distinct architectures have been presented to support multimodal interaction, that can be splitted into categories [19]. Notable examples of architectural multimodal categories include agents based architectures (e.g., QuickSet (Cohen et al., 1997), HephaisTK (Dumas, Lalanne, & Ingold, 2009)); components based architectures (e.g., OpenInterface (Serrano et al., 2008), ICARE (Bouchet & Nigay, 2004)); layer based architectures (e.g., MUDRA (Hoste, Dumas, & Signer, 2011)); server based architectures (e.g., SmartWeb Handheld (Sonntag et al., 2007)) and other distributed architectures (e.g., I-TOUCH (Pocheville, Kheddar, & Yokoi, 2004)). Considering that a lot of different multimodal frameworks with different characteristics are accessible, the W3C [20] proposed a

Figure 2.6: Examples of multimodal systems in movies and video games. From left to right, Minority Report and Playstation EyePet pictures. (source: `goo.gl/PSPb4w` and `goo.gl/yCQNsG`, respectively).

---

standard Multimodal Architecture, that is characterized by its decoupled nature and a set of standards [21] for the specification and communication of its modules (figure 2.7). Overall, the architecture has three notable components: **the modalities**, in charge of handling interaction inputs and outputs to the system; **the interaction manager**, responsible for managing the events coming from the modalities and communicating with the application logic; and **the data model**, storing information about the current state of the application. All these components communicate through an asynchronous protocol based on life-cycle events. The fusion engine presented, in figure 2.7, inside the interaction manager, is the component responsible for providing truly multimodal interaction by merging sequential or redundant events from different modalities into new events. For instance, merging the speech input "Zoom here", with the gesture of pointing to a location on screen.

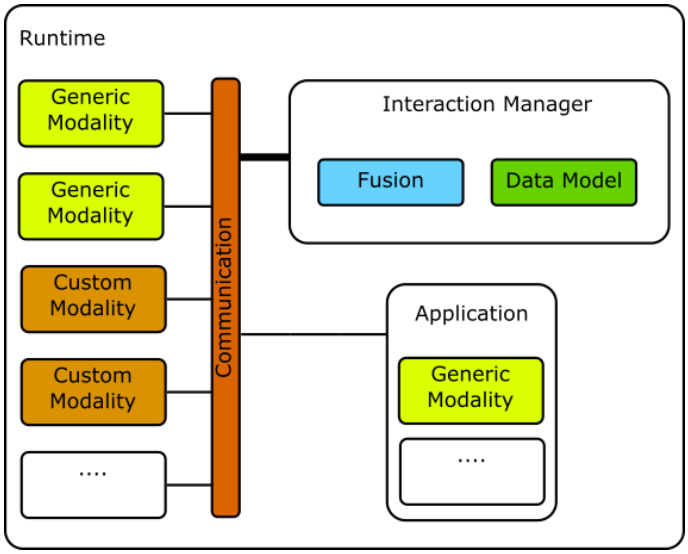

Figure 2.7: Main components of an architecture supporting multimodal interaction, abiding to the W3C recommendations.

Taking into consideration the W3C recommendations [20], a framework has been proposed

to support the development of multimodal interactive applications [22] and supporting a set of interaction modalities, with a particular emphasis on a generic speech modality enabling multilingual speech interaction [23]. The generic modalities are available, in the framework, for off-the-shelf usage, by developers, providing an easy way to add interaction options to applications without mastering the required technologies. These modalities encapsulate most of the complexity through the use of automated methods for their configuration (e.g., Almeida et al. [23]) and rely on (remote) support services.

2.2.3 User-context and interactive Systems

The increase on the research and work related to the field of pervasive computing as well as the rising number of devices supporting it [24], makes relevant that computing systems become context-aware [25]. In this context, applications need to adapt to a set of conditions including for example, the device type, the connection state, the user environment, and so on. Additionally, these applications also need to take into consideration the user context and environment and try to adapt to that particular situation without an explicit request from the user.

While we are most accustomed to consider user interaction as explicit inputs (e.g., whenever a user presses a button to perform an action), we may also consider implicit interaction (i.e. inputs to the system not coming explicitly from the user) [1]. Current research in context-awareness presents a strong work-oriented focus on location [26], considering the mobile application field, enabling adaptation of how applications respond (e.g., disabling sounds when entering the company's meeting room) or provide information (e.g., showing the closest coffee shops) to the user.

Recently, in the summer of 2016, Pokémon Go[16] (figure 2.8), an augmented reality mobile game that connects in-game objects (Pokémons) with real world objects, revolutionized the mobile phone gaming market and according to statistics[17], this location-aware game has entered the top-10 Android games in America and took the first place in AppStore[18] within 24 hours after its release.

In a similar way, more of these implicit interactions, reflecting aspects of the users' experiences, environment and context, can and should be considered, with the emotional state of the user playing a major role in modulating human behaviour [2]. In fact, the emotional profile of each person should be adapted to that person and situation. Our mood is influenced by social factors [27], which prints a new challenge in the development of systems and contents. The only way to deal with this and to enable the person to be engaged with the computational systems is to adapt these systems to the user's needs.

2.3 User-centred design

Despite the design and development of applications not being the focus of our work, one of the goals we aim to achieve is the development of demonstrator applications that support the developed work. In this context, it is important to have a background on user-centred

[16] goo.gl/CiiMn8
[17] goo.gl/7G5Z1o
[18] goo.gl/iXPXC6

Figure 2.8: Examples of screenshots of the location-aware game, Pokémon Go, that concerns to the user location to provide related content. (source: `goo.gl/RE6rd7`).

---

design that empowers the creation of personas, scenarios and requirements that need to be considered in the scope of these applications.

According to Human-centred design processes for interactive systems, ISO 9241-210 [19] (2010), "Human-centred design is an approach to interactive system development that focuses specifically on making systems usable. It is a multi-disciplinary activity". In this context, user-centred design process *(UCD)*, also called human-centred design process, is an iterative design approach that considers users, their needs and interests in each stage of the design cycle [28, 29].

Four distinct phases characterize each iteration of the *UCD* cycle [30, 31] (figure 2.9). The first phase involves an understanding, by designers, of the context in which users could use a product. Then, it is required to define the users' requirements (second phase). The third phase focuses on designing and developing solutions. The final phase is related to an evaluation process, where an analysis is made to the users' context and to the previously established requirements, in order to verify how well a design approach is operating. A continuous cycle between these four phases is performed until the evaluation outcomes are suitable.

Considering the user-centred design scope, it is still relevant to have some background in personas, scenarios and goals, in order to better understand users, their context and motivations.

[19] goo.gl/9ndDPD

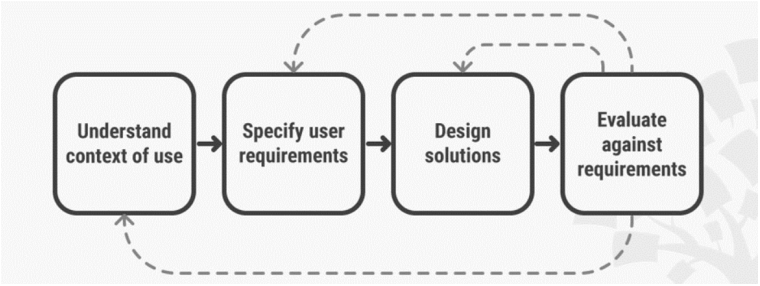

Figure 2.9: The four phases involved in each iteration of the *UCD* process. (source: `goo.gl/PSPb4w`).

2.3.1 Personas, scenarios and goals

The demonstrator applications we propose to develop will require to have in mind the users, their context and the scenario in which they will be integrated. In this regards, we provide a brief introduction on personas, goals and scenarios.

Personas [29, 32] are fictitious characters, created with the intention of representing types of users that might use a service, product or application. Usually, Personas are defined upon research and observations of experiences, behaviours, needs and expectations of real people. Personas are considered a tool that makes the planning and design task less painful and are intended to help researchers and developers.

Generally, the definition of a Persona is associated with a scenario that intends to accomplish certain(s) goal(s). A scenario is a situation that defines the details of the persona story, detailing when, where and how a sequence of events takes place. The goal is associated with the motivation that triggers actions by the persona. The scenario reaches an end, when the goal or goals are achieved [29].

2.4 Autism Spectrum Disorders

The work to be developed aims to be able to be applied to multiple scenarios and contexts, however, we define a scenario (demonstrator application) with a public with special needs, children with ASD, as a possible target audience that benefits from a adaptability of applications that takes into account their emotional states. In this context, it is important to have a background on autism spectrum disorders and their abilities, considering several domains.

Difficulties in social interaction and communication are some of the most verifiable diagnostics in children with autism spectrum disorders [33, 34]. These children have extreme difficulties in making eye contact, participating in social experiences and most importantly, they fail in recognizing people's emotions from their facial expression [35, 36, 34]. In this context, this target audience could profit from a high level of adaptability, regarding applications. In this regards, and considering the evolution in the affective computing field and the importance of multimodal interaction in the context of the user, it becomes indisputable that the technology specially developed to tackle this audience's needs is still far from its potential [33].

Autism Spectrum Disorder *(ASD)* is a group of neurological and developmental disorders

characterized by a lack of skills concerning communication and behaviour as well as unusually narrow, repetitive interests (American Psychiatric Association 1994), affecting how persons interact and communicate with others. This condition is estimated to be about in 1 in 68 children according CDC's Autism and Developmental Disabilities Monitoring [37]. Typically, symptoms are diagnosed in the first two years of life and they last throughout a person's lifetime. These disorders can be split in three different groups[20]: *Autism, Asperger's Syndrome and Pervasive Development Disorder-Not Otherwise Specified (PDD-NOS).*

Autism is the most common and critical type of autism spectrum disorder diagnosis, characterized by significant language delays, social and communication challenges along with motor disabilities. People with Asperger syndrome have average or above-average intelligence and might experience some atypical behaviours and interests. *PDD-NOS* covers situations that do not match the other two types of conditions and *PDD-NOS* individuals are often socially oriented.

Studies [18] have shown that almost all children with this condition considers it a very demanding task to manage their emotions, which is vital, for example, for their integration in classrooms or other social groups. This capability to accurately perceive, recognize and interpret facial expressions is particularly important when considering social interactions [38]. Therefore, it is extremely important, to possess the ability to recognize and interpret emotions through facial expressions [38], for example. Studies [39] also show that reinforcements in learning emotions recognition techniques on different contexts (e.g., classroom or home environments), can result in improvements of this same perception on daily activities for this targeted public.

2.4.1 Applications for Children with ASD

In this context of ASD disorders, the usage of technology is extremely important in order to provide methods that allow children with this condition to better communicate and better express their feelings. In this regards, several applications specially developed for this targeted public reached the market.

Augmentative and Alternative Communication Applications

There is a significant number of applications on the market that concerns to augmentative and alternative communication (AAC). These systems are designed to replace standard means of communication and are extremely important for persons with ASD conditions [40].

Picture Exchange Communication System *(PECS)* is the most common AAC technique used [41] and helps children with autism to communicate using picture cards with different meanings. However, when considering a large number of cards, these systems fail to provide portability [41] and this flashcard therapy involves an accurate memorization of facial emotions expressed in the cards [42], ending up having the opposite result expected on developing social interaction skills.

Taking into consideration the challenges above, there has been a lot of research and development on building AAC applications that can run on mobile devices such as smartphones and tablets [41]. Examples of remarkable AAC mobile applications (figure 2.10) on the market

[20] goo.gl/MB4WhU

that aim to help children with ASD to communicate and express themselves include *Proloquo2Go*[21], a symbol-based alternative communication application developed by AssistiveWare that runs on iOS devices, uses symbol packs called SymbolStixs intended to help people with difficulties in communication. It has been successfully used by people with autism, cerebral palsy and down syndrome; *Avaz*[22] is a PECS mobile application developed specially for children with communication challenges. Avaz aims to make speech therapy more effective, improves the child's language skills and increases the child's desire to communicate and interact with others. Similarly to *Proloquo2GO*, it enables users to create messages using picture symbols and additionally, it provides high-quality voice synthesis and a third-party provider sharing mechanism that enables users to share their messages; *Predictable*[23] is a mobile self-learning word prediction that helps people with difficulties in communication to easily build sentences. Predictable generated content such as phrases will automatically be stored in the cloud, enabling a secure and easy access to personal data.

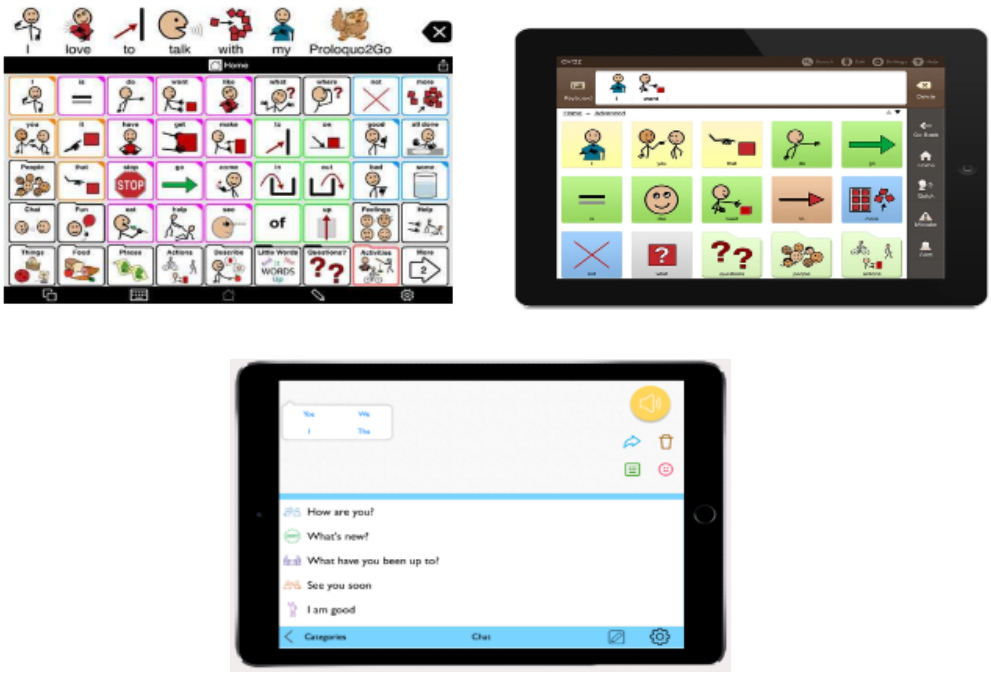

Figure 2.10: Examples of existing AAC applications. On the top-left, Proloquo2Go (source: `goo.gl/2ep9Dv`); on the top-right, Avaz (source: `goo.gl/qHth8p`); on the bottom-center, Predictable (source: `goo.gl/1zqgn1`)

Emotions and Social Behaviour applications

Children with ASD present not only difficulties in communication and interaction with other people, but according to [35, 36, 6], this target group also shows serious limitations

[21] goo.gl/N2SvtC
[22] goo.gl/Smc2mG
[23] goo.gl/CQ6RDp

regarding emotions recognition through facial expression analysis. In this context, there are also several mobile applications (figure 2.11) intended to bring the emotional context closer to children with this condition.

CaptureMyEmotions [6, 43] is probably the mobile application, developed for this audience, that takes more into account the user's emotional context. It uses a wireless physiological sensor (*Affectiva's* sensor) and facial expression recognition to infer emotional states, combining user's photos, videos or audio content. As this application lets children take picture and record audio or video content, it increases the chance of getting a genuine emotion, differentiating itself from the other applications available on the market that can return inconclusive results due to the memorization of the analysed data. AutismXpress[24] is an Android and iOS application designed to encourage people with autism to recognize and express their emotions through its appealing interface. It uses emojis for emotions representation; Let's Learn Emotions[25] offers autistic children multiple ways of targeting emotions and it contains five interactive modes: two matching games, a guessing emotions game, a memory game, and a debate game related to the origin of each emotion. iSequences[26], developed by Fundacion Planeta Imaginario[27], is also a mobile application designed for children with autism that enables them to play the role of 6 characters and experience 100 different day-to-day situations (e.g., brushing their teeth, washing their hands, going to the beach). The application aims to match the daily activities performed and the emotion inferred from those activities. *The Let's Face It! (LFI!)* Emotion Skills Battery [38], a complement to the *Let's Face It!* [44], also aims to explore the facial expression recognition skills of participants with autism. This mobile application contains three games: Name Game, Matchmaker Expression and Parts-Whole Expression. The Name Game is intended to assess the child's skill to label basic facial emotions. On the other hand, the Matchmaker Expression game expects the child to match a specific emotion from three visual targets while the Parts-Wholes Expression game tests the child's abilities to identify happy and angry expressions.

Table 2.3 presents the most notable characteristics of the mentioned applications, specially design for the ASD public, intended to better help this public to communicate and to recognize/detect emotions.

2.5 Discussion

The state-of-art previously described intends to give an overview of the current technologies in our field of research and its purpose usage, and mainly to present the major gaps we deem relevant to pursue.

With the wide range of devices (smartphones, tablets, laptops), tasks and purposes, it becomes extremely important to consider generic mechanisms that consider the users' emotional context, since the option of developing for each particular device is a possibility, but involves an exponential development effort by developers.

As exposed in the affective computing related work, a considerable number of toolkits that enables developers to detect emotional content is already provided, however, each mechanism

[24] goo.gl/Uu1cLk
[25] goo.gl/EE8e1A
[26] goo.gl/REbk3T
[27] goo.gl/Hk1rui

Application	Category	Subcategory	Operating System(s)	Device(s)	Available Language(s)
Proloquo2GO	Communication (AAC)	Phrases and Boards	Apple	Phone Tablet	English French Spanish
Avaz	Communication (AAC)	Phrases and Boards Remote Communication	Apple Android	Phone	English
Predictable	Communication (AAC)	Text	Apple Android	Phone Tablet	Danish Dutch English French Portuguese Others
CaptureMyEmotion	Emotions and Social behaviour	Emotions	Android	Phone Tablet	English
AutismXpress	Emotions and Social behaviour	Emotions Puzzle	Apple Android	Phone Tablet	English
Let's Learn Emotions	Emotions and Social behaviour	Emotions	Apple	Phone Tablet	English French Portuguese
iSequences	Emotions and Social behaviour	Daily Habits Social Stories and Skills	Apple Android	Tablet	Catalan English French German Spanish

Table 2.3: Summary of the most remarkable applications intended to help the ASD public to better communicate and to recognize/detect emotions, according to `https://goo.gl/pYp6v3`.

Figure 2.11: Examples of existing emotions and social behaviour applications. On the top-left, CaptureMyEmotions (source: [6]); on the top-right, Let's Learn Emotions (source: `goo.gl/WeissN`); on the bottom-left, AutismXpress (source: `goo.gl/kMwpWN`); on the bottom-right, iSequences (source: `goo.gl/WbAk51`).

---

used for this purpose involves, to the developer, an intensive study and mastering of the technology for the appropriate use of this tool. When considering several of these instruments, the complexity of the solution increases exponentially, which is infeasible for developers to manage.

The generic modalities, considering multimodal interaction related work, are available, in the proposed framework, for off-the-shelf usage, by developers, providing an easy way to add interaction options to applications without mastering the required technologies. Such an approach, we argue, could serve the integration of affective computing with multimodal interactive systems through the consideration of a generic affective modality, encapsulating the complexity of the supporting computational methods, and providing a simpler entry point to the design and development of emotionally-aware applications. In light of this evidence, the next chapter proposes a novel systematic solution to support emotionally-aware applications that addresses these issues.

Chapter 3

Generic affective modality to Support Emotionally-aware Applications

Considering the importance of affective computing in the context of interactive systems and the current state of the art, this section illustrates our overall vision of the scenarios that need to be considered and how these can be conceptually accommodated in the scope of a multimodal interactive architecture. Additionally, we propose a generic affective modality to provide a decoupled solution to develop emotionally-aware applications. In what follows, we describe the main aspects of the adopted conceptual vision, architecture and implementation.

3.1 Conceptual Vision

Nowadays, workspaces such as offices, laboratories, classrooms or even leisure spaces such as parks and museums make use of a variety of interactive technologies and applications to make a more pleasant experience for the user. Therefore, for the context of our work, we consider an environment where users can interact with multiple devices (e.g., personal computer, smartphone, tablet, interactive board) and applications for different purposes. Considering these scenarios, some of these devices (figure 3.1) and applications could have mechanisms or sensors used to capture data that can be used to derive emotional context.

Figure 3.1: Examples of devices that contain sensors (camera, audio recorder, textual input or even heart-rate monitor) to capture data that could be used to infer a particular emotional state.

Given the sensitivity of sharing user data that can be used to extract emotions (such as photos, videos, or even psycho-physiological measurements), users can choose to accept to

share just their emotional state with other applications and devices. Considering this logic of thought, figure 3.2 presents our overall vision of the main components envisaged to support emotionally-aware multimodal interaction. By taking in consideration an affective modality, a particular application, for example, can be endowed with access to the emotional status of the user. This modality makes use of different support services and toolkits to, for instance, given a particular image of the user's facial expression, captured by the smartphone camera, determine the current emotional state.

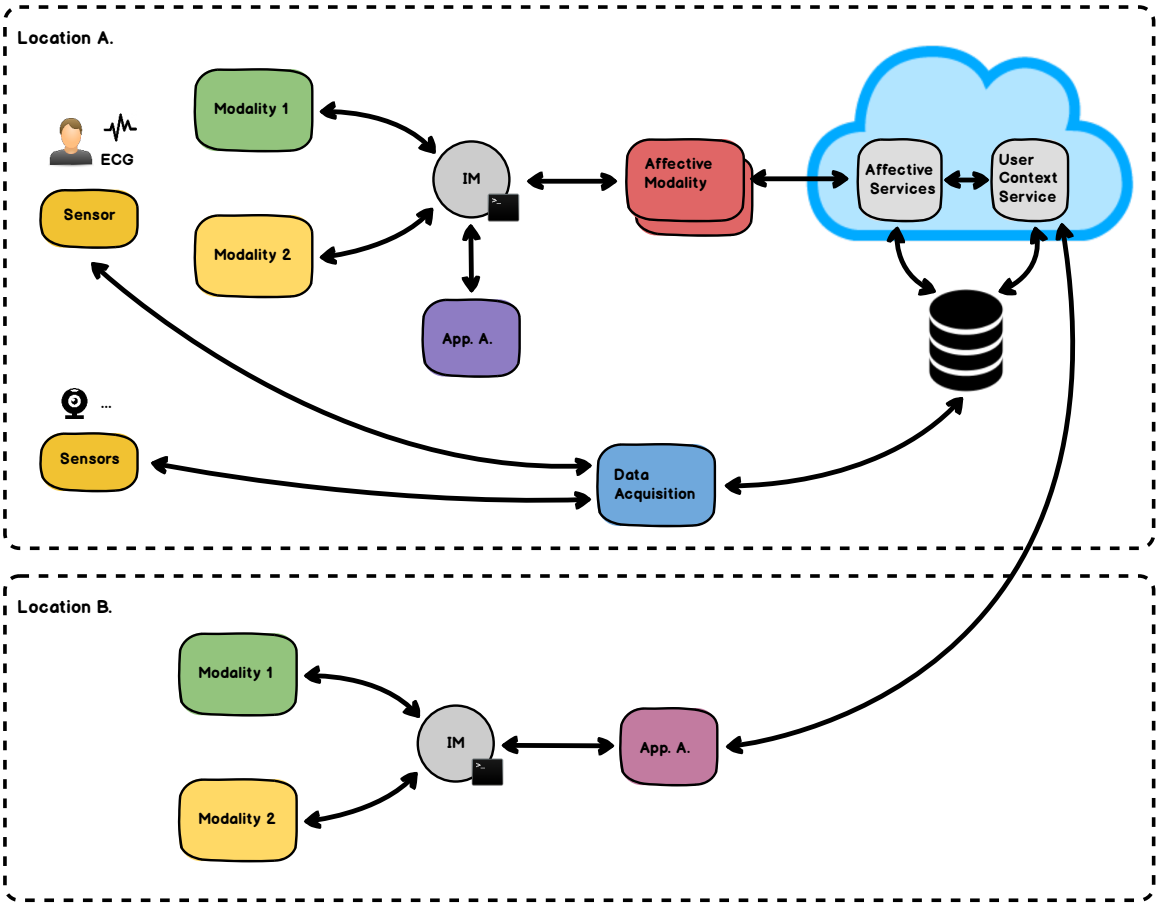

Figure 3.2: Overall conceptual vision for emotionally-aware multimodal interfaces.

Additionally, the affective modality might be more complex and rely on an additional infrastructure, represented on figure 3.2 for location A. For instance, different kinds of sensors, independent from the application, might be available, in the environment (e.g., a video camera, a microphone array) or in a wearable (e.g., *ECG*) and store data in a building server or backend service. When any application needs the current emotional status, it provides the user identification to the affective modality and the support services will enquire if data is available for that user and if a compatible method exists to process it, returning that user's emotional state. The reason for this approach is twofold: first, application developers do not need to address data acquisition, although a sensor managed by the application is possible; and second, the application does not have access to the collected data, ensuring privacy,

since it may contain sensitive information about that user, such as video footage or psycho-physiological data from an ECG signal. For instance, a third-party application, installed at home, and connected to the affective modality will only get the emotional status and will be completely blind to the data from which it was extracted.

Finally, at a different location (figure 3.2, location B.), an application without access to an affective modality may still be able to adapt to particular emotional states by enquiring the user context service. This service may be updated with the emotional status, obtained from the affective services, at fixed intervals or according to what the user authorizes. For instance, an application in the fridge, based on the user context, may suggest a mood lifting beverage when the user comes around, based on an overall negative mood, during that morning. A simple storyboard that intends to represent this situation can be found on figure 3.3.

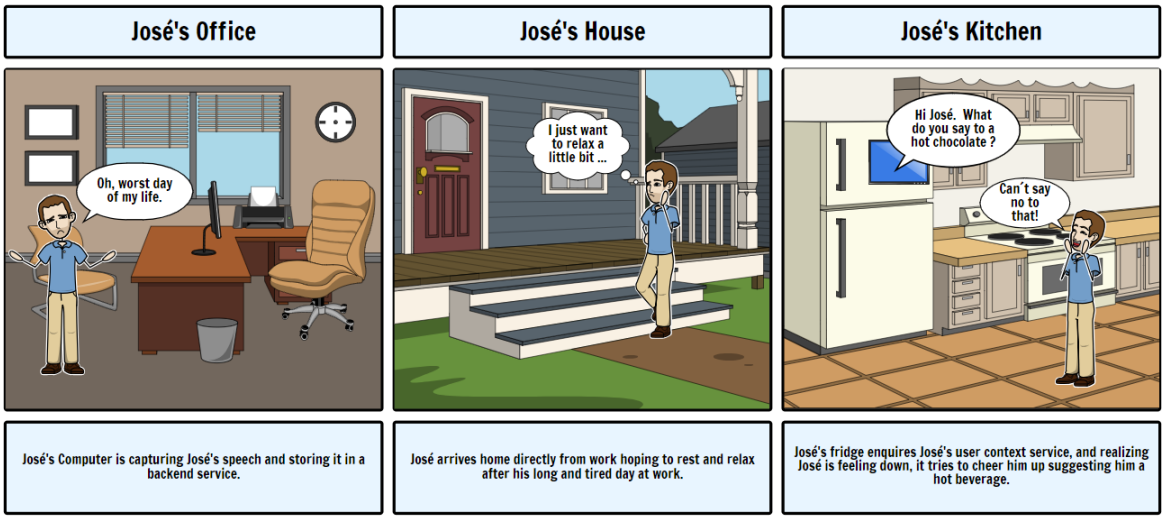

Figure 3.3: A figurative scenario, designed on `goo.gl/rSRTrT`, that considers an user named *José* and an application, that runs on *José's* fridge, accesses *José's* user context service and suggest him a hot beverage based on his recent emotional state.

A notable difference between the two scenarios (locations A. and B. of figure 3.2) of accessing the emotional status, i.e., an affective modality vs emotional status obtained from the user context, concerns how the application uses it. In the first, emotion is an input, such as pressing a key or issuing a voice command, to which the application can (or may be expected to) react. In the second, the application uses emotions to customize itself and, such as location, in a mobile phone, it may not always be available.

3.2 Modality Architecture

Figure 3.4 presents the main modules of the proposed and developed affective modality architecture. This modality communicates with a module of a multimodal framework [20, 19], **the interaction manager** (*IM* on figure 3.4), developed on our research institute, *IEETA*.

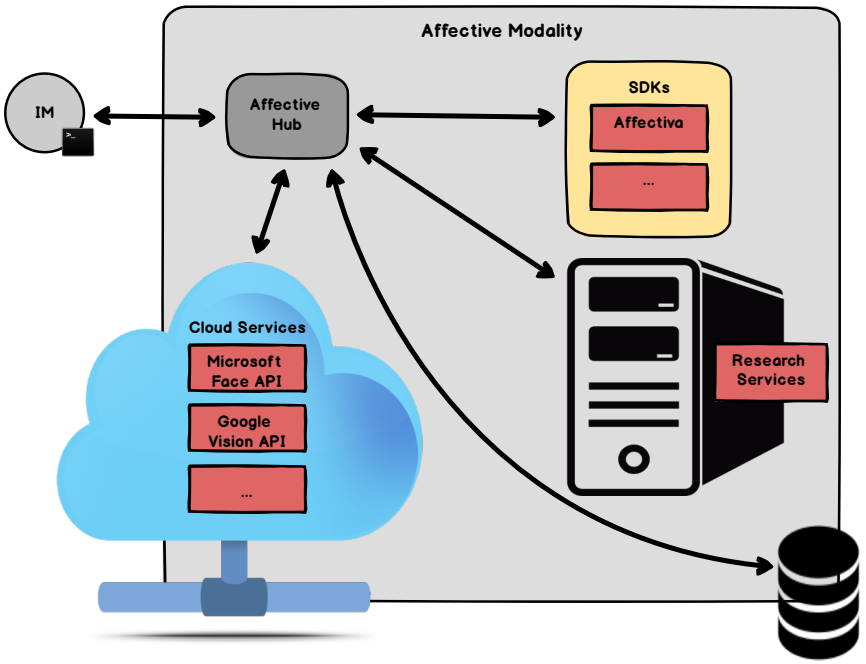

Figure 3.4: Depiction of the overall components for an affective modality. An affective hub service manages the events exchanged with the interaction manager and articulates the different methods available locally, in the cloud, or as part of a research effort.

Interaction Manager

The multimodal framework [23, 19] module was designed to simplify the integration of multimodal capabilities into new applications. In this context, the interaction manager is responsible for handling user's interaction within applications, connecting all other modules. The communication between modalities (or applications) uses the HTTP protocol, and it considers HTTP GET and POST requests for the communication. The events received from the modalities or applications are parsed and processed by this manager using a state chart machine (defined with *SCXML*[1]), following W3C recommendations.

Affective Modality

The affective modality is responsible for sending, receiving and processing events, from and to the interaction manager *(IM)* and it contains a module responsible for the management of affective responses, **the Affective Hub**. The whole process inherent to the decision of the predominant emotion output, i.e, the calculation of the dominant emotion is made by the modality itself, after receiving an affective response by the hub. The modality still contains a backend service, where users' data such as images and audio or text files are stored.

[1]https://goo.gl/jFMDQF

Affective Hub

This service provides responses with an affective basis and it is able of handling multiple client requests. Depending on the type of request specified by the client, it can receive data to process (e.g., a video frame captured from a webcam) or check for relevant data stored in the database (backend system) regarding a particular user. Then, depending on the type of data available, the affective hub chooses the best suitable method to process it and extract emotional context. The chosen method will depend if any was specifically requested, by the client, or which method, from those available, is able to deal with the data. The available methods can be grouped in three different categories: those that are **local to the infrastructure**, as can be the case, for instance, of *Affectiva*; those that are **cloud services** (e.g., *Microsoft Face API*); and those that stem from **ongoing research** (e.g., at our Research Institute, *IEETA*).

3.3 Affective Modality development

This section comprises all the aspects concerned with the overall implementation process of the affective modality. In this context, we begin by covering aspects related to the development of the modality, the affective hub block and the backend service that supports its features.

3.3.1 Modality Implementation

We developed the Affective Modality using Visual Studio[2] Enterprise 2015 IDE and the C# programming language. In what concerns development matter, this modality connects to the Interaction Manager, and contains an Affective Hub (figure 3.4) that manages all the events and articulates the different options available for obtaining the emotional status of the user. From the development standpoint, and to ensure a more uniform outcome from the affective modality, whatever the emotion recognition method used, we considered that any processing related to the decision of what emotion should be considered, from the data provided by each service or toolkit, should be performed by the modality, with the Interaction Manager and all connected applications receiving only the output of the dominant emotion *(e.g., sadness or happiness)*. In this regards, the modality receives, from the hub, a response with affective context, analyses it, and returns to the IM the output of the predominant emotion. On figure 3.5, an example of a response with affective context, that will be analysed by the modality, is provided.

Affective Hub module

The developed system can be used by an audience that does not master technological knowledge. Therefore, it is extremely important to apply technologies that do not impose overload on the user. In this regard, users will never need to make updates to software versions or SDK's (e.g., updating the *Affectiva SDK version*), and it is the developer's responsibility to maintain the REST-based service and its dependencies.

In this context, the affective hub block is a REST-based service, developed using the C# programming language and placed in a virtual machine (running *Windows Server 2012*) of the research institute *(IEETA)*. Whenever this module needs to extract emotional content

[2] goo.gl/t2bhbq

```
[
    {
        "faceRectangle": {
            "top": 84,
            "left": 239,
            "width": 492,
            "height": 492
        },
        "faceAttributes": {
            "gender": "male",
            "age": 53,
            "glasses": "NoGlasses",
            "emotion": {
                "anger": 0.094,
                "contempt": 0.049,
                "disgust": 0,
                "fear": 0,
                "happiness": 0,
                "neutral": 0.853,
                "sadness": 0.004,
                "surprise": 0
            }
        }
    }
]
```


Figure 3.5: Example of an affective response provided by the affective hub service. In this particular case, the neutral emotion output (value *0.853*, from 0 to 1, corresponds to the predominant emotion of the analysed data) is being sent by the modality to the IM.

---

from any sort of data, it uses the appropriate service or toolkit, receives the response in *JSON* format and sends that response to the modality that implements all the rationale in analysing the response.

Modality Services/Toolkits

The affective hub makes use of 7 services or toolkits for emotional states extraction: 3 for image or video content, 2 for audio and 2 for plain-text content. *Affectiva, Microsoft Face API and Google Vision API* analyse image and video data. *DeepAffects and Beyond Verbal* are the services in charge of processing and analysing audio content. Concerning textual content, the *Microsoft Text Analytics API and the Tone Analyzer by IBM Watson* take care of extracting an emotional state from any sort of text input.

We chose to consider at least 2 methods (3 for image or video content) for the processing and extraction of emotional content for each type of data because we considered that the hub must have redundant mechanisms in the process of emotion analysis and extraction (e.g., if the *Google Vision API* is unable to extract an emotion, the modality uses other service or toolkit that deals with this data type, for example the *Microsoft Face API or Affectiva*).

Modality backend service

The modality requires a backend service for data storage, a user management facility, and a Backend as a Service (BaaS) was the solution found for this project.

The backend of modality was implemented using Firebase[3] (figure 3.6), a cloud-based *NoSQL* database that uses JavaScript Object Notation *(JSON)* to store and structure information. Additionally, it supports real-time data syncing and provides features such as cloud storage, push notifications or social networking integration and it enables an easy integration into applications thanks to its own API's.

The integration of this backend service enables the modality to store multiple data types (e.g., image, video, audio, text files), profiting from a firebase bucket[4]. The affective hub requires to the modality backend service when inserting new data or when it needs to check if exists any particular data concerning that user that could be used to infer an emotional state.

Our project explore three features of Firebase: *authentication, real-time-database and cloud storage*. As Firebase does not offer support for a C# Software Development Kit (SDK), we use two open source C# Firebase libraries. The firebase database library[5] is a simple wrapper built on top of Firebase Real-time Database REST API and supports, among other features, to listen to backend insertions and updates (real-time notifications). The firebase authentication library[6] enables the modality to properly authenticate users (email and password) by generating an authentication token to be used with REST queries against Firebase Database endpoints.

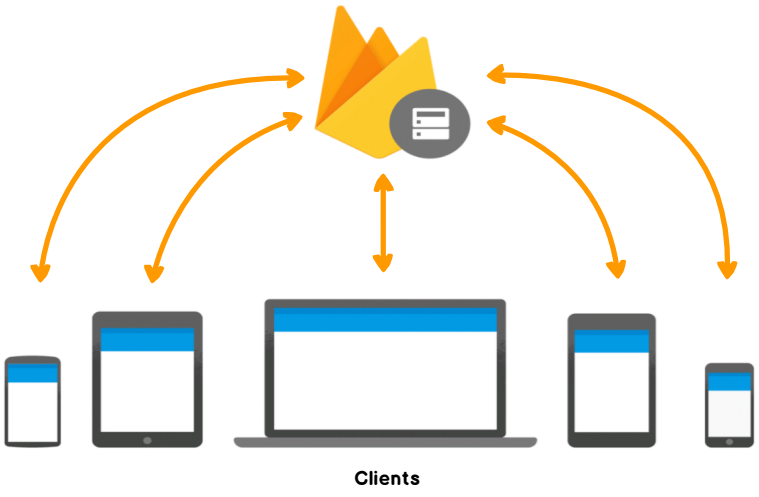

Figure 3.6: Firebase data syncing on multiple clients and devices, enabling applications to listen for new insertions or updates on the backend service. In the scope of our project, only the developed modality is linked to firebase, that manages all the information and notifications to connected clients.

[3]https://goo.gl/DhssNA
[4]goo.gl/8dGPYc
[5]goo.gl/JnMD1a
[6]goo.gl/9FkR3T

3.3.2 Emotions Output Language

In multimodal interaction, an action can be triggered by inputs from different modalities. For instance, to scroll left, a user can say "go left" or make a gesture to the left (e.g., swipe left). While the input is different, the outcome is the same action, which means that both inputs have the same meaning, i.e, semantics. In a similar way and applying a similar rationale, we consider a uniform output, regardless of the toolkit, that is being used by the modality. In this regards, as a first simple approach, the same six basic emotions *(happiness, sadness, fear, disgust, surprise and anger)* defined in the Facial Action Coding System *(FACS)* (Ekman, 1978) [45, 46] are applied (figure 3.7). In addition, we chose to consider one more emotion (*neutral*), which, at this point of our work, represents a detection of a emotion that does not fit into any of the 6 previously mentioned groups. Therefore, the modality uses the most suitable service or toolkit to analyse that data type, and returns the emotion output of the predominant emotion. In the event that the service or services inquired are unable to label an emotion semantic (e.g., considering an image, it may not contain any facial expression or if it does, it may not perceptible enough to infer an emotion from it), the modality sends, to the IM, the emotion output *'NO_EMOTION_DETECTED'*.

Figure 3.7: The six basic emotions according to Ekman and Friesen presented on facial expressions pictures. (source: `goo.gl/pGFb3N`).

3.3.3 Modality modes of operation

The modality works with two modes of operation: **explicit** and **implicit** requests. The modality makes use of the explicit mode of operation when an application that is using the modality makes a request to obtain a particular emotion from a data type (e.g., a user takes a photo and wants to tag that photo with an emotion).

On the other hand, the implicit mode of operation, allows an application to adapt itself to the emotional context of a user without making any kind of explicit request to the modality

(e.g., an application that lacks mechanisms of data collection can use the modality to adapt to the user's affective state, if the modality contains any data related to that same user, or if it receives data from that same user through any other connected application).

Explicit mode of operation

The explicit mode of operation does not require the user identification to make requests of emotional content to the modality. This mode of operation is proposed to answer to sporadic explicit requests from applications. The explicit mode of operation can be divided into three mechanisms: when an application sends, to the IM:

- **A valid url** with data that is intended to be analysed;
- **A filepath** that represents the path of the data intended to be analysed;
- **A binary data sequence** (that represents a particular data) to be analysed;

The flow of information between the application that desires to make explicit requests, Interaction Manager and Affective Modality is represented on figure 3.8.

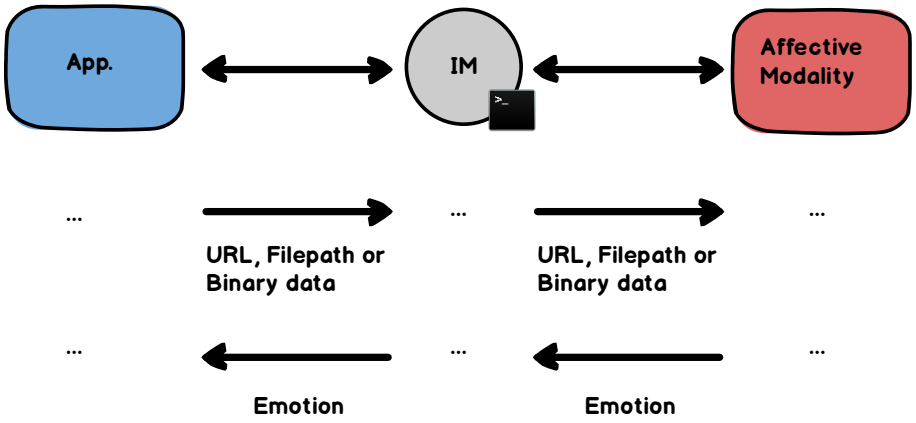

Figure 3.8: Explicit request flow of information representation between the application, IM and affective modality. The application sends, to the IM, data to be analysed, that will be directed to the Affective Modality. After the emotion extraction by the modality takes place, an emotion output reaches the application.

Implicit mode of operation

The implicit mode of operation needs to take into consideration the user in order to operate. In this regards, the application that intends to connect to the IM, needs to provide the user identification (his credentials) (figure 3.9). The modality will then, keep the user's credentials *(email, password, username)* in a *JSON* file (figure 3.10 provides an example of the modality configuration file), when a connection is established. This file also contains the priority services to be used by the modality.

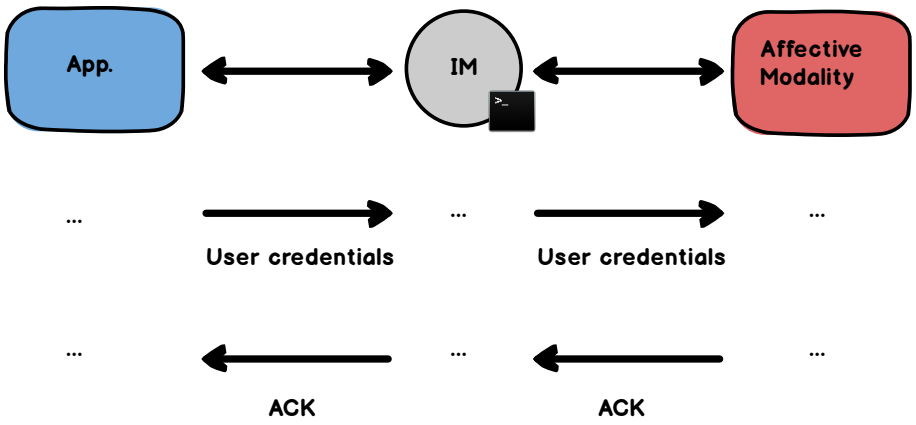

Figure 3.9: Implicit mode of operation: Flow of information between the application, IM and affective modality. The application sends, to the IM, the user credentials, which will subsequently reach the modality. The modality responds with an acknowledge message, that reaches the application.

```
{
    "firstName": "Roger",
    "lastName": "Waters",
    "username": "rogerwaters",
    "email": "roger.waters@pinkfloyd.com",
    "password": "wishuwerehere",
    "videoService": "Microsoft Face API",
    "audioService": "Deep Affects",
    "textService": "Tone Analyzer"
}
```

Figure 3.10: The JSON file features the user first name, last name, username, email and password (for firebase authentication capabilities) and priorities services/toolkits to be used. First name, last name and the priority services are optional values, which means they don't need to be specified by the application. So, in case the priority services are not specified by the user, the modality makes use of the default services.

The implicit mode of operation follows the listening strategy, that is described bellow:

Modality listening strategy

The modality listening strategy complies with the following rationale (figure 3.11): when a new user connects with the IM, the modality checks, for that particular user (based on the username), if there is data stored on the backend service that could be used to infer emotional content. The system analyses the most recent data from that user, requires to the most suitable service/toolkit to treat that data type, and returns the emotion output to the IM, enabling any application that is currently communicating with this framework module to adapt to the current emotional state of this user. If the modality could not derive a particular emotional state (e.g., the captured image does not contain a facial expression or the captured audio stream does not have a significant duration that enables its analysis), it requires to

the alternative services or toolkits to attempt to extract an emotion. In case the modality makes use of all mechanisms for extracting emotional content for that data type and, still, no emotion has been derived, the system returns *'NO_EMOTION_DETECTED'* to the IM.

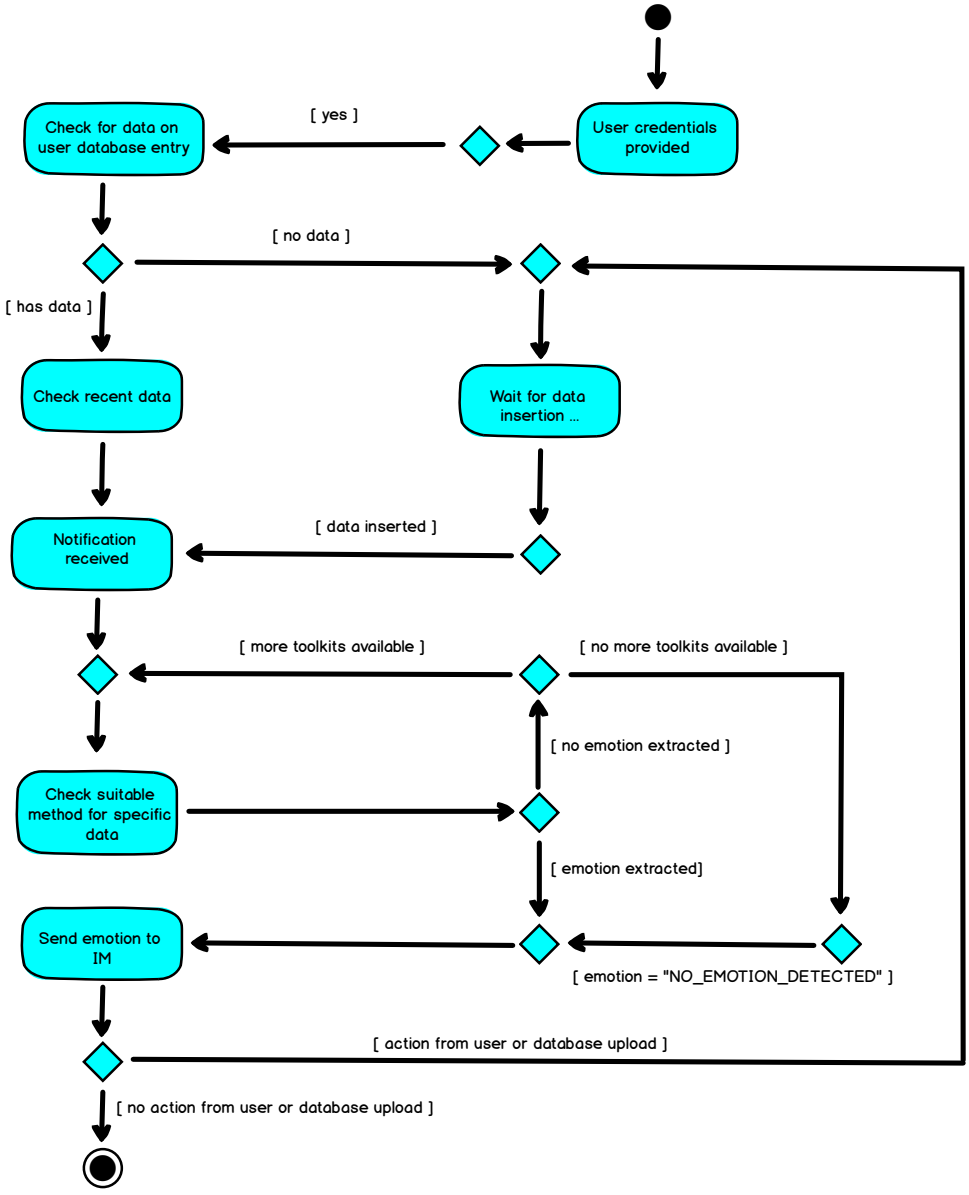

Figure 3.11: Modality listening strategy applied, concerning to the modality implicit mode of operation.

After that, the modality keeps listening to any changes that may occur in the database, i.e, if any data is inserted (figure 3.12), the modality triggers a real-time notification, using the modality backend service capabilities, and immediately makes use of the most viable service or toolkit to extract an emotional state, following the same rationale previously described.

This synchronization mechanism is very useful when considering a scenario where a data

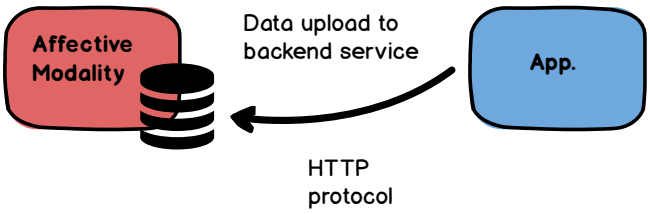

Figure 3.12: Data that is uploaded to the modality backend service triggers a notification for the modality listening strategy.

---

acquisition application (e.g., a video tracking application on an infrastructure entry point) runs on location *A.*, and a device where this modality is integrated runs on location *B.* Any application that uses the modality can adapt to that user emotional state even if it does not contain any mechanism of data acquisition.

3.4 Discussion and Conclusions

This chapter presents a **generic solution** for supporting the **integration of emotional context in interactive systems** encompassing an overall conceptual proposal and the implementation of a core element of that vision, the affective modality. The proposed affective modality is fully integrated with the multimodal framework developed at *IEETA*, following the W3C recommendations for multimodal architectures, and communicates with the interaction manager, which deals with the events coming from the modalities or applications.

By placing most of the complexity of integrating emotion extraction methods in remote services or toolkits, our solution hides the complexity in the process of extracting emotions, and the developer only needs to integrate the developed modality and deal with the semantic contents of the events that reach the application. With this solution, we potentially reduce the complexity and time required by developers to consider emotions in their own applications.

This decoupled solution was designed and developed in a scalable manner, meaning that new methods, mechanisms or services can be easily added to enable extracting emotional states from data types not considered, at this stage, and without requiring changes to the interactive systems already integrating the affective modality.

Chapter 4

Demonstrator Applications

Two demonstrator applications were developed to support the proposed work on the affective modality that intends to illustrate its potential, showing how the proposed modality enables these applications to integrate user's emotions during its use.

The first application (***Affective Spotify***) was not formally submitted to any evaluation assessment, and it was developed with the main purpose of providing the grounds to support the modality development while we sought for the most suitable services and toolkits. The second (***MoodDiary***), on the other hand, was developed at a more advanced stage of the modality development, and we tried to take full advantages of the modality features while placing both the application and the modality under evaluation.

4.1 Affective Spotify

The purpose of the first developed application was to create a multimodal interactive test bed for the development of the proposed affective modality. In this regards, a first desktop application prototype, referred as ***Affective Spotify***, was developed enabling multimodal interaction with Spotify[1]. This application supports voice and gestures interaction to perform actions such as play or pause a song or playlist, change song and volume, and add a song to a playlist. The user's emotional state is considered through the proposed affective modality and the application recommends what songs to hear based on the user's mood.

This application enables, among other features, to access to the user's playlists as well as each playlist tracks (figure 4.5). Users can choose the playlist they want to listen whether using voice or gestures (e.g., by clicking on the playlist they want) commands.

4.1.1 Requirements

At this development stage, our purpose was to create a multimodal interactive test bed for the development of the proposed modality. In this context, we didn't specify user requirements but instead, we did specify system requirements. The following requirements for the overall system were inferred:

- The overall system must adopt a multimodal architecture;
- The system must support more than one modality;

[1] goo.gl/4CEBpE

- The emotional context must be relevant to the depicted scenario;
- The modality must resort to local and cloud methods for emotional states extraction.

4.1.2 System Overview

Figure 4.1 presents the main modules integrating the overall architecture of the proposed prototype application. There are three independent modules corresponding to each of the modalities (**speech**, **gestures** and **affective** interaction). These modalities communicate with a fusion engine (profiting from work developed on a multimodal framework [23]), responsible for the merging of events from different input modalities, configured with SCXML [47, 48]. Merged events are sent to the IM, which communicates with the Spotify application issuing the requests to the Spotify API.

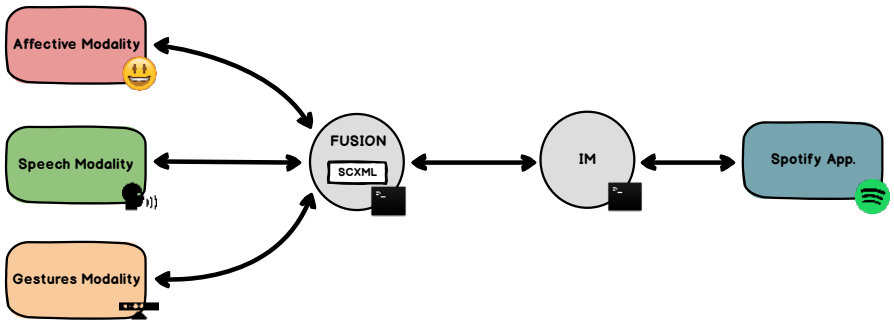

Figure 4.1: Main component blocks of the prototype application, Affective Spotify.

4.1.3 System Implementation

The system (figure 4.2) was developed in a laptop (HP Envy, 15.4-inch, 2.40GHz Intel Core i7-4700MQ, 8GB RAM, running Windows 10). For gestures detection, a Kinect from first version *(v1)* was used and additionally, we take advantage of an Asus Cerberus Headset with a microphone to improve the quality of voice commands to the speech modality.

The application was developed taking advantage of the Visual Studio[2] Enterprise 2015 IDE using the C# language. The multimodal framework module, IM, was responsible for the communication between all modalities and the Spotify application. At this stage of development, the events that came from the affective modality are not being merged with other events from the other modalities. The fusion engine merges events from the speech and gestures modalities (e.g., "play this one" (voice input) followed by a click on the desired playlist).

The gestures input modality is supported by a Kinect camera and the speech modality receives input from a microphone connected to the headset and synthesized speech messages are transmitted to the system using Microsoft's Speech Synthesis text-to-speech (TTS) and the European Portuguese language pack. Finally, the affective modality can receive a video stream, from the laptop's webcam, or a single image, also obtained from the webcam (*Spotify App.* on figure 4.1).

[2]goo.gl/2TJcwb

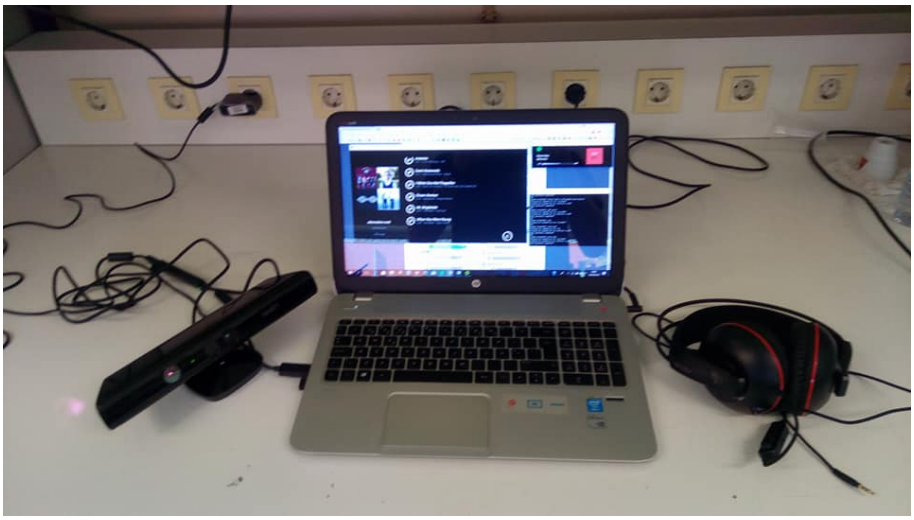

Figure 4.2: Affective Spotify system deployment. From left to right, a Kinect, a laptop and an Asus Headset.

At this development stage, all independent modules (the three **modalities**, the **fusion engine** and the **IM**) ran locally on the laptop. In this regards, the laptop ran an instance (runtime) of the IM and an instance (runtime) of the fusion engine.

During the development of this application, several libraries and SDK's were used. Table 4.1 provides a short description of each of them.

Library	**Description**	**Application usage**
Json.NET	A .NET library to process JSON data.	Used to iterate over JSON data.
SpotifyAPI-NET	A C# .NET library for the Spotify-Client and the Spotify Web API communication.	Used to communicate with the spotify API.
Microsoft Speech Recognition	A SDK that provides functionalities to acquire and monitor speech input and to create speech recognition grammars.	Used to convert audio stream into text transcription and to enable developers to create grammars, concerning the speech modality.
Microsoft Speech Synthesis	A SDK that contains classes that enable to initialize and configure a speech synthesis engine.	Used to create text-to-speech outputs, concerning the speech modality.
Microsoft Kinect	A SDK that enables developers to create applications that support gesture and voice recognition, using Kinect sensor technology on computers running Windows OS's.	Used to detect and recognize gestures inputs, concerning the gestures modality.

Table 4.1: Notable libraries and SDK's used in the Affective Spotify application development.

Affective Modality Implementation Stage

Our purpose with this prototype was to create a multimodal interactive test bed for the development of the proposed modality. Therefore, at this development stage, the affective modality was still at an early stage of development.

For this first implementation of the modality, we just considered methods for image or video processing. In this context, we included support for the *Affectiva* and *Microsoft Face API*

services, that analyse the video stream or image captured by the laptop's webcam. Particular consideration was given to the fact that these toolkits, although similar in what they can extract from facial expressions, vary in the method of operation: Affectiva runs **locally on the infrastructure** and Microsoft Face API is a **cloud service**. *Affectiva's SDK*, *Affdex*, returns facial expression output in the form of metrics: the level of recognition for seven emotions*(anger, contempt, happiness, fear, sadness, surprise and disgust)*; twenty facial expression metrics *(e.g., attention, lip press)*; thirteen emojis expressions *(e.g., laughing, relaxed, rage)*; and four appearance metrics *(e.g., age, ethnicity, gender and glasses)*. Similarly to Affectiva, Microsoft's Face API face detection also extracts several face related characteristics such as gender, age, head pose, facial hair, smile intensity, glasses and enables the detection of the same seven emotions metrics as *Affectiva's SDK*. Both methods score each detected emotion with a percentage of confidence ranging from 0 (no expression) to 100 (fully present expression). Intending to reduce false positive emotion detections, we chose to only consider scores above a threshold of *95* that result in an event, sent to the IM, reporting the output of the recognized emotion (e.g., joyful, sad) at that precise moment.

Figure 4.3 illustrates the facial detected features using the Microsoft Face API. In this particular case, a high level of happiness is detected, for the current song.

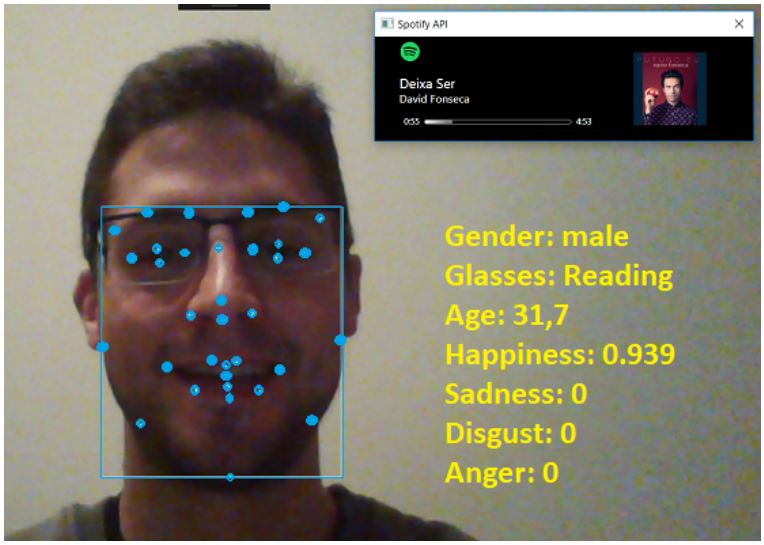

Figure 4.3: Sample data returned by Microsoft Face API while being used to detect the user's emotion during song playing with Spotify. In this particular case, a high level of happiness is detected.

4.1.4 Illustrative Scenarios

A formal evaluation was not carried out, however, we collected feedback from users regarding aspects that could be improved and how the application could adapt itself to the emotional context of users. Figure 4.4 represents a real interaction between a master degree student with ***Affective Spotify***.

The application enables listening to several songs and playlists with Spotify (figure 4.5), and provides multimodal interaction via speech and gestures modalities, thus allowing an increased level of accessibility, potentially reaching a wider audience. The application detects,

Figure 4.4: A student while interacting with Affective Spotify, making use of the three available modalities: speech, gestures and affective interaction.

---

through facial recognition, the user's emotions, in real-time. The only requirement is that the user, while interacting with the application, has his or her face framed within the laptop's camera.

The application, besides reacting to negative emotion events, detected while playing a song, also considers the dominant emotion for two distinct contexts: one that considers the **current track** (single context) and another that considers the **previous four songs** (session context).

These two buffers are populated along time with the emotions detected above the decision threshold, either in the context of just one song, or several songs. When deciding the overall mood for a particular song or session, these buffers are analysed and a majority rule applied.

The following scenarios depict how Affective Spotify considers emotion, through the proposed affective modality, to take action.

Single track context

In this scenario (illustrated in figure 4.6) only the current song is taken into account and three different emotional contexts are considered.

Negative emotion detected: while playing a music that the user dislikes (e.g., a negative emotion is being detected by the modality), the application will ask immediately if the user desires to skip that song and play the next one instead.

Positive emotion single track overall: While playing a music that the user likes (e.g., the single track buffer is positive at the end of the song), the application will ask the user, at

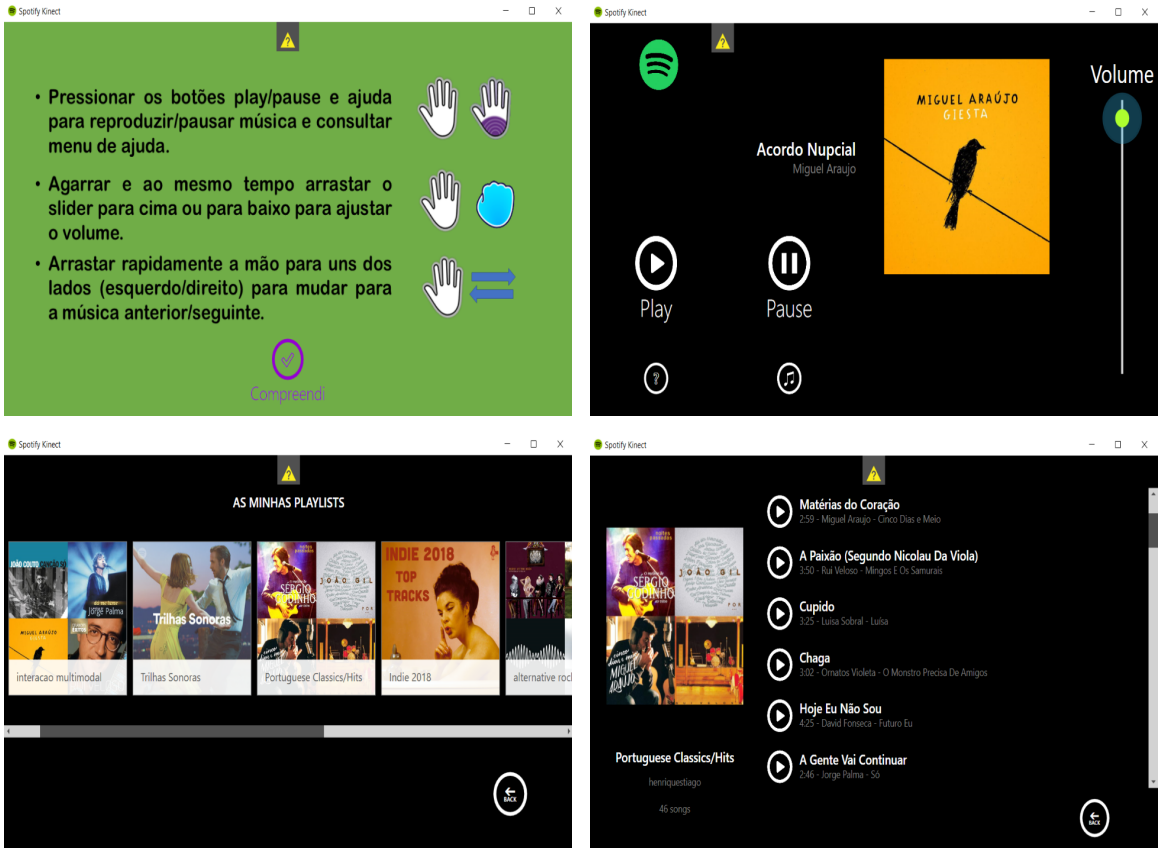

Figure 4.5: On top-left: visual representation of the initial menu that explains the body gestures recognized by the gestures modality; on top-right: visual representation of the application main page; on bottom-left: visual representation of the playlists' page; on bottom-right: visual representation of a playlist's list of songs page.

---

the end of the song, if he/she desires to add that specific track to the user favourites musics list.

Negative emotion single track overall: Although negative emotions were detected, by the application, the user chose not to switch to the next song. At the end of the song, as the overall emotion (single track buffer) is still negative, the application will ask the user if another music genre or category should be played.

Session context

In this scenario, illustrated in figure 4.7, multiple songs (four songs) are taken into account in order to establish a session profile. This scenario can be defined as an aggregation of multiple instances of the previous scenario, which means that a session profile is defined after joining individual track profiles.

Overall positive emotion session: The overall emotion of the session profile (session buffer) is positive, which could mean that the user is enjoying the songs played. The application asks the user if he/she desires to listen to that week's featured playlists, since the user

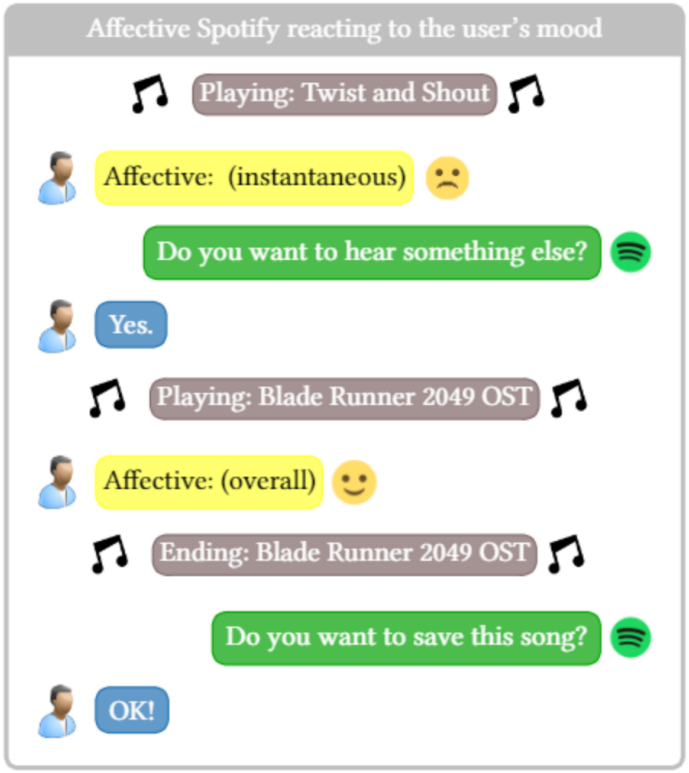

Figure 4.6: Reacting to the user's mood during a song: if a negative mood is detected, Affective Spotify proposes to play something else; if the user likes the music, the application suggests adding it to the favourite's list.

---

is potentially feeling good and could be available to listen to other genres and music categories.

Overall negative emotion session: The overall emotion of the session profile (session buffer) is negative, without the user accepting the initial proposal to change to another song, which could mean that the user is not feeling well or may be going through some hard times. The application will try to cheer the user up by proposing to play more joyful songs in order to boost the user's mood.

4.2 MoodDiary

The purpose of the second developed demonstrative application, ***MoodDiary***, was to illustrate the ease of integration of the developed modality across multiple platforms and OS's. This application represents a part of the envisaged scenario depicted on figure 3.2, for location A, where an application captures data that is stored on the modality backend system, and another application adapts its content to the emotional states being extracted from the captured data. This emphasizes that a new application, beyond not having to tackle the complexity of different affective methods, may even not be required to feed the affective modality with any data. The data may be obtained by external acquisition systems and this

Figure 4.7: Adaptation to the session overall mood: if the overall mood for the last four songs has been negative (part 1), Affective Spotify proposes hearing something potentially more uplifting; if the session is positive (part 2), overall, the application suggests hearing some novelties.

---

approach ensures an extra degree of privacy, since new applications do not need to access to the data, only to the emotions it unveils.

We also intended to address how a user with difficulties regarding emotions' understanding might profit from an adaptability of applications considering the affective context. In this regard, this application has been developed considering a scenario where a child with autism interacts with a mobile device, for instance, in a classroom context. First evaluations have been considered, to assess the overall concept, by performing a heuristic and a usability evaluation.

4.2.1 Methods

Based on our state-of-the-art, overall challenges and goals to achieve, we adopted a Persona [49] for a child diagnosed with *ASD* and another for a special education teacher. In this regards, we established a usage scenario and defined a set of requirements for the application, based on it.

Personas

For this work, and given the time constraints on creating new Personas, we chose to adopt the work proposed by [50], properly evaluated by a panel of experts and already considered in the design of a first application prototype for children with ASD [51], in the creation of

a Persona of a child with autism, *Nuno Rocha*. This proposed method for the creation of Personas was based on the methodology described by [49].

Table 4.2 lists a simplified description of *Nuno's* Persona, extracting some details of the original Persona, not directly important for the current context, and changing some aspects of Nuno's needs and characteristics and most importantly, his motivations, specially inferred from [35, 36, 34].

Table 4.2: Persona for Nuno Rocha, a child diagnosed with ASD.

Nuno Rocha born in 2009, in Aveiro district, Portugal, lives with his father, mother and a 13 year old sister. At the age of 3, he was diagnosed with Autism Spectrum Disorder (level 2 in the scale of severity), with associated cognitive deficits. He currently attends the 4th grade in a Basic School, were he benefits from a specific individual curriculum, including Special Education support, using a structured learning model (TEACCH), and Speech Therapy sessions. At home, he prefers to watch TV and play computer games. Although he uses the computer, he is not able to research information on any search engine, nor does he use the social networks for communication. The elected mean of communication is speech. He struggles to perform basic social interactions, he has difficulties to communicate and express himself and he is mostly capable of using short and simple sentences (subject + verb + object). He appears to understand simple oral material, having difficulties on the comprehension of longer sentences that lack visual support or that are out of the context. As far as it concerns reading, he recognizes all the letters from the alphabet, but he seems to struggle on the reading process, mostly syllabic, associated to a loss of purpose and hesitations. He writes with orthographic correction but he needs support on the structuring of small texts and in answering questions. He makes requests in his areas of interest, and when questioned he has difficulties in answering, sharing daily experiences, and beginning and keeping a conversation. He shows difficulties in keeping eye contact, respecting interaction shifts and adjusting to the context and to the interlocutor. In some situations, he verbalizes incoherent phrases and out of context (delayed echolalia). Nuno appears not to be able to recognize any type of emotions whatsoever.

Motivation: *Nuno* would like to express his emotions and understand the emotions others are expressing.

Considering the motivations and roles of additional stakeholders (among family and educators), we also chose to use the Persona of a special education teacher, named *Isabel Oliveira*, who can use the proposed application, in order to understand what can motivate the child and what can help to calm it down. A description of this Persona is presented in Table 4.3, omitting some details of the original Persona, and complementing some aspects of Isabel's needs, characteristics and most importantly, her motivations based on input collected from professionals with daily contact with children with ASD. Other potential stakeholders are also considered on our usage scenario (mother and father), however, at this stage they were regarded as Serving Personas [49] and, for the sake of brevity, we chose not to explicitly deal with their Personas, at this time.

Table 4.3: Secondary Persona for Isabel Oliveira, a Special Education teacher.

Isabel Oliveira was born in France, in April, 1972, and currently lives in Aveiro. She is married and has two daughters and a son. She has a BSc in Language, Literature and Cultures, with a major in Portuguese and French, and post graduation in Special Education. She has 19 years of teaching experience, 7 of those in Special Education, having a very good level of knowledge regarding her field of work. She constantly works to be up to date with recent knowledge and practices. From her point of view, information and communication technologies can be an asset during the learning process of kids with special educational needs and, during her work with them, she often uses computers and tablets with educational software. Her main interests are literature, cinema, cooking, and writing, but, during her free time, her family is the main priority. Despite her experience, Isabel feels she can not detect and recognize when Nuno is feeling down or when he is feeling happy. She hopes that with the advancement of technology, intelligent mechanisms will emerge that will help her infer an emotional state for Nuno.

Motivation: *Isabel* would like to be able to assess and judge the emotional states that *Nuno* is feeling, so she could intervene and help him when she considers it appropriate. She also desires that there was some technology capable of storing information that could help to calm down Nuno.

Scenario

Our project work aims to explore the possibilities of a child using a mobile device as a tool for school interaction and to develop his skills on emotions recognition and understanding. In this regards, it is considered an application scenario, in the context of the classroom, where we explore how the child can interact with the application and how a special education teacher can take a part in its interaction. The usage scenario adopted follows:

- Scene 1: **Add a new diary entry**

Nuno just finished his school activities and wants to report and share this moment. When he uses the smartphone, the main menu is composed of two options: “My Diary” and “My Happiest Moments”. *Nuno* touches the “My Diary” button, and the mobile device displays all current diary entries. He then, clicks the button to add a new diary entry, placed on the top-right corner of that view.

- Scene 2: **Take a picture**

Next, the application displays the add diary menu, where *Nuno* should provide a small description of the diary entry. The application also provides a button that should be clicked by the child to take a picture. When pressed, the smartphone displays the current view obtained from the device’s camera, and after *Nuno* frames his face with the device’s camera, he presses the button to capture the photo, which will be stored on the device, and it will be tagged with the corresponding emotion.

- Scene 3: **Comment the entry**

Additionaly, *Nuno* is presented with one final feature, that enables him to comment the diary entry. He then, presses the “Press to comment” button, and a new view appears where he is

able to write what he is feeling. At the end, he presses the "Add" button and the comment will be added to that diary entry alongside with the tagged emotion extracted from it. *Nuno* can then, choose to create the diary entry or cancel it, by clicking the corresponding button.

- Scene 4: **Check his happiest moments**
Nuno just finished his activity in the speech therapy session, and he is potentially feeling down. *Isabel*, using the smartphone, goes to the section "My Happiest Moments", where all diary entries with an overall "happy" emotion associated are presented. *Isabel* then, shows to *Nuno* some of his "happy" entries to try to cheer him up by showing him moments that were, in previous situations, associated with happiness.

Requirements

Considering the envisaged scenario, an application for children with some level of autism needs to be developed for some mobile platform(s), properly adapted to the context of school. The proposed application can be used either by the child, by the special education teacher or by the child's family members.

We inferred the following requirements for the mobile application, considering the overall context, Personas and scenario:

- Taking or selecting photos;
- Saving the photos;
- Commenting photos or diary entries;
- Emotional tagging on photos, comments and diary entries;
- Viewing photos, comments and corresponding emotional tagging;
- Deleting comments or diary entries;
- Access to data (photos or comments) that can be used to calm down a specific individual of this targeted public;

4.2.2 System Overview

Figure 4.8 presents the main modules (mobile application, IM, data acquisition application and affective modality) that integrate the overall architecture of the proposed system. The mobile application uses the HTTP protocol, issuing GET or POST requests, to a central cloud-based IM, while the data acquisition application uploads data to the affective modality backend system.

4.2.3 System Implementation

The current system deployment consisted on a laptop (HP Envy, 15.4-inch, 2.40GHz Intel Core i7-4700MQ, 8GB RAM, running Windows 10), an iPod (Touch 6th generation running iOS version 9.2) and a virtual machine (running windows server 2012). Figure 4.9, contains a laptop, where a video data acquisition application collects visual contents concerning the user facial expression. The user interacts with the mobile application running on the iPod and the

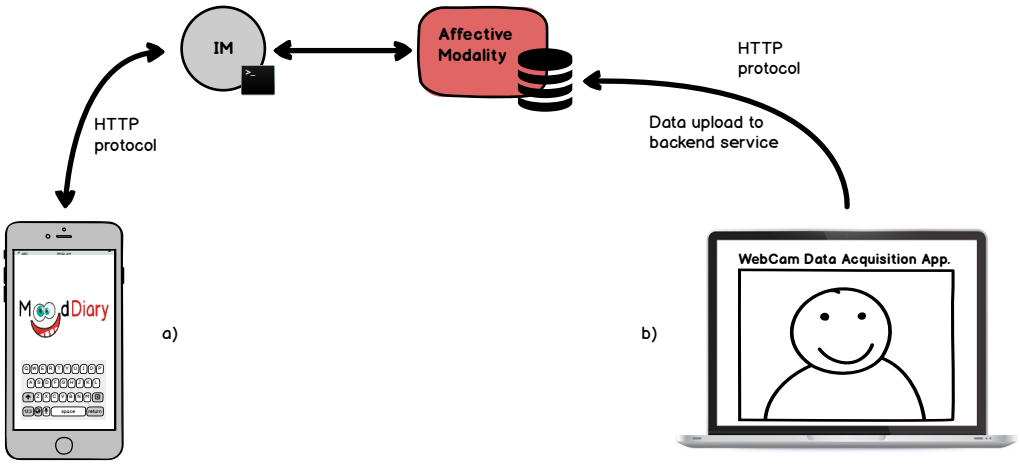

Figure 4.8: Current system deployment. The mobile application (figure 4.8 a)) uses the HTTP protocol to communicate to a cloud-based IM and a remote video data acquisition application (figure 4.8 b)) uploads data to the modality backend service.

video data application is constantly storing data in the affective modality (that runs on the virtual machine) backend system. Considering communication aspects, the application that runs on the mobile device communicates with the IM which communicates with the modality. Here, a different approach from ***Affective Spotify*** has been used. A cloud-based IM runs remotely and the mobile application connects to that IM.

During the development of this system, several libraries were used. Table 4.4 provides a short description of each of them.

Library	**Description**	**Application usage**
SwiftyJSON	SwiftyJSON library makes it easy to deal with JSON content in Swift language.	Used to iterate over JSON data.
Alamofire	Alamofire is an HTTP networking library written in Swift.	Used to communicate with the interaction manager through the HTTP protocol.
SWXMLHash	SWXMLHash is a simple way to parse XML in Swift.	Used to process XML data when receiving responses from the interaction manager.
PasswordTextField	A custom textfield with a switchable icon that shows or hides the password. This library is written in Swift.	Used to show or hide the password field when identifying the user.

Table 4.4: Notable libraries used in the MoodDiary iOS application development.

Affective Modality Implementation Stage

Unlike ***AffectiveSpotify***, the overall ***MoodDiary*** system was developed after the development of the affective modality, therefore, at this development stage, this system already uses the most recent version of the modality (refer to section 3.3), being able to profit from

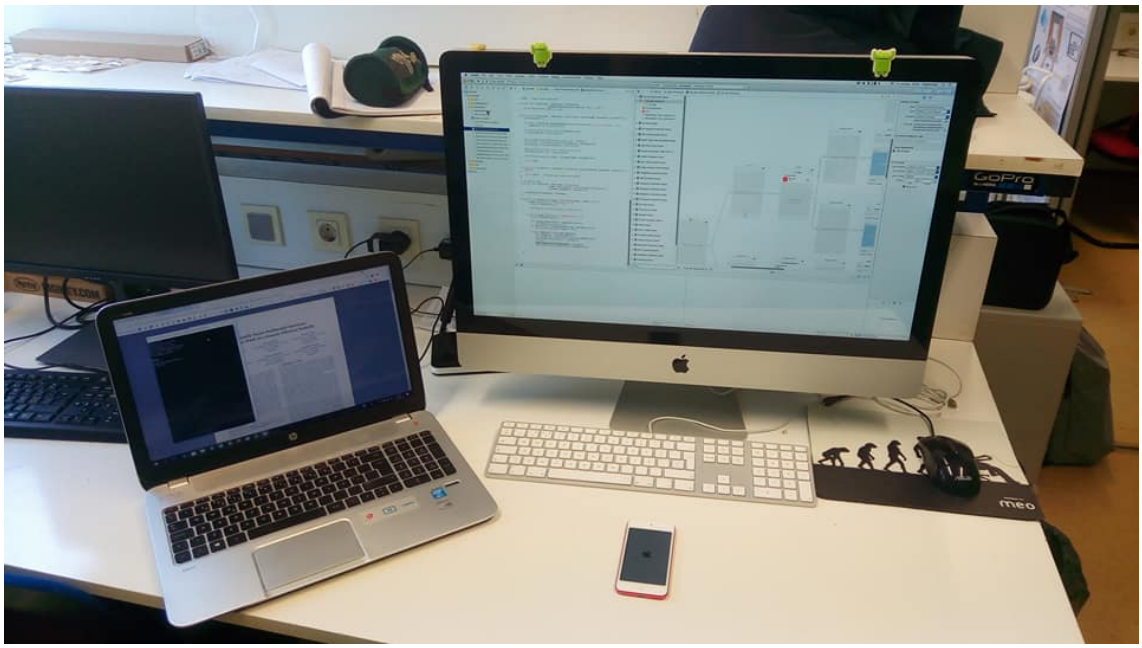

Figure 4.9: Case study experimental setup: a laptop where the video data acquisition application runs and an iPod where the mobile application runs. In the figure, it is also possible to note the iMac where the mobile application was developed.

---

all of its features. The system considers the modality methods for image, video or text processing. The chosen services or toolkits considered by the modality will depend if any was specifically requested, by the user, or which method, from those available, is able to deal with the data. In the event that no services or toolkits are specified by the user, the modality will make use of the standard services or toolkits.

Video Data Acquisition Application

This application was developed in a laptop (HP Envy, 15.4-inch, 2.40GHz Intel Core i7-4700MQ, 8GB RAM, running Windows 10) taking advantage of the Visual Studio[3] Enterprise 2015 IDE using the C# language. To collect user facial expression data, this application requires at the laptop's webcam to capture images every 5 seconds. For each collected image, the application uses the modality to store the captured images frames in that user database entry. This allows other applications (such as the application that runs on the mobile device) to adapt to that user's emotional context, using as input this application acquired data. Figure 4.10 represents a sample of data (e.g., facial expression images) that is being captured by this application.

MoodDiary Mobile Application

The application that runs on the mobile device (iPod Touch 6th Generation running iOS version 9.2), referred as ***MoodDiary*** (figure 4.11), was developed on an iMac (27-inch),

[3]https://goo.gl/2TJcwb

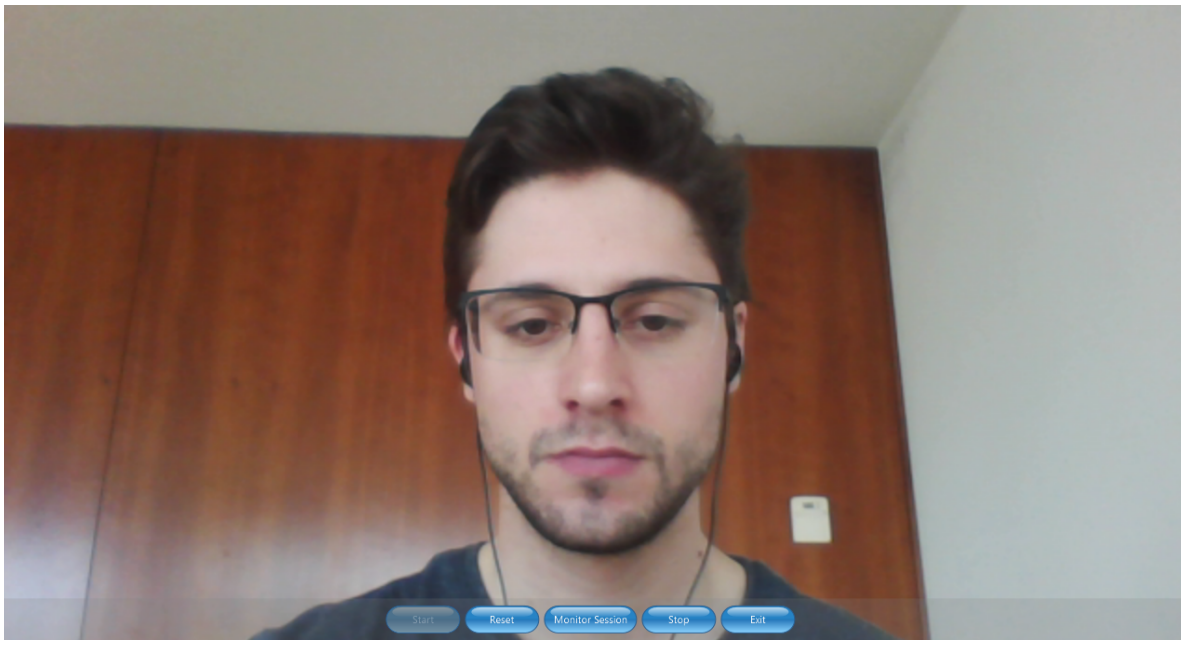

Figure 4.10: Example of a data sample that is being captured by the video data application that runs on the laptop. A screenshot is captured every 5 seconds and stored into the modality backend service.

3.2GHz Intel Core i3, 8GB 133MHz DDR3, ATI Radeon HD 5750, running El Capitan OS (version 10.11.5) using Xcode[4] Software version 8.2.1 capabilities and Swift[5] (version 3.1) programming language.

Figure 4.11: MoodDiary logo - Mobile application.

This application features the main capabilities of a virtual diary (such as creating specific entries for a particular day), and additionally, it enables to associate diary entries with images and comments. The application is divided into two sections: "**MyDiary**" and "**My Happiest Moments**". The first section allows users to view, edit or delete diary entries as well as create new entries. Additionally, users can add photos and comments to diary entries. The second section allows users, on the other hand, to view the diary entries associated with the "happiness" emotion. So, when creating new entries in the "**MyDiary**" section, if the diary entry is associated with "happiness", it will be added to the "My Happiest Moments" section.

[4] goo.gl/BXUfqp
[5] goo.gl/vDwFah

This application uses the proposed modality to associate emotions with diary entries, photos and user comments. Concerning the proposed modality implementation details, this application considers the two modes of operation by the modality:

- **Explicit emotional requests** (figure 4.8, section a)) enables this application to associate emotions with content provided sporadically by the user. In this context, it analyses a photo (inserted or selected by the user) and associates an emotion with it. Following the same principle just applied, it analyses comments (made by the user) and associates an emotion with it;
- **Implicit emotional requests** (figure 4.8, section b)) are always being considered by the application and are directly related to the user mood. In this context, this application uses the data captured by the video data acquisition application (remote data), to associate emotional states to diary entries, reflecting the user mood when visualizing contents. Additionally, the MoodDiary application still adapts the screen content to the user mood (e.g., it changes to the "**My Happiest Moments**" section if a high number of occurrences of the "sadness" emotion is detected).

Figure 4.12 presents samples of the application views, presenting the application interface and main features.

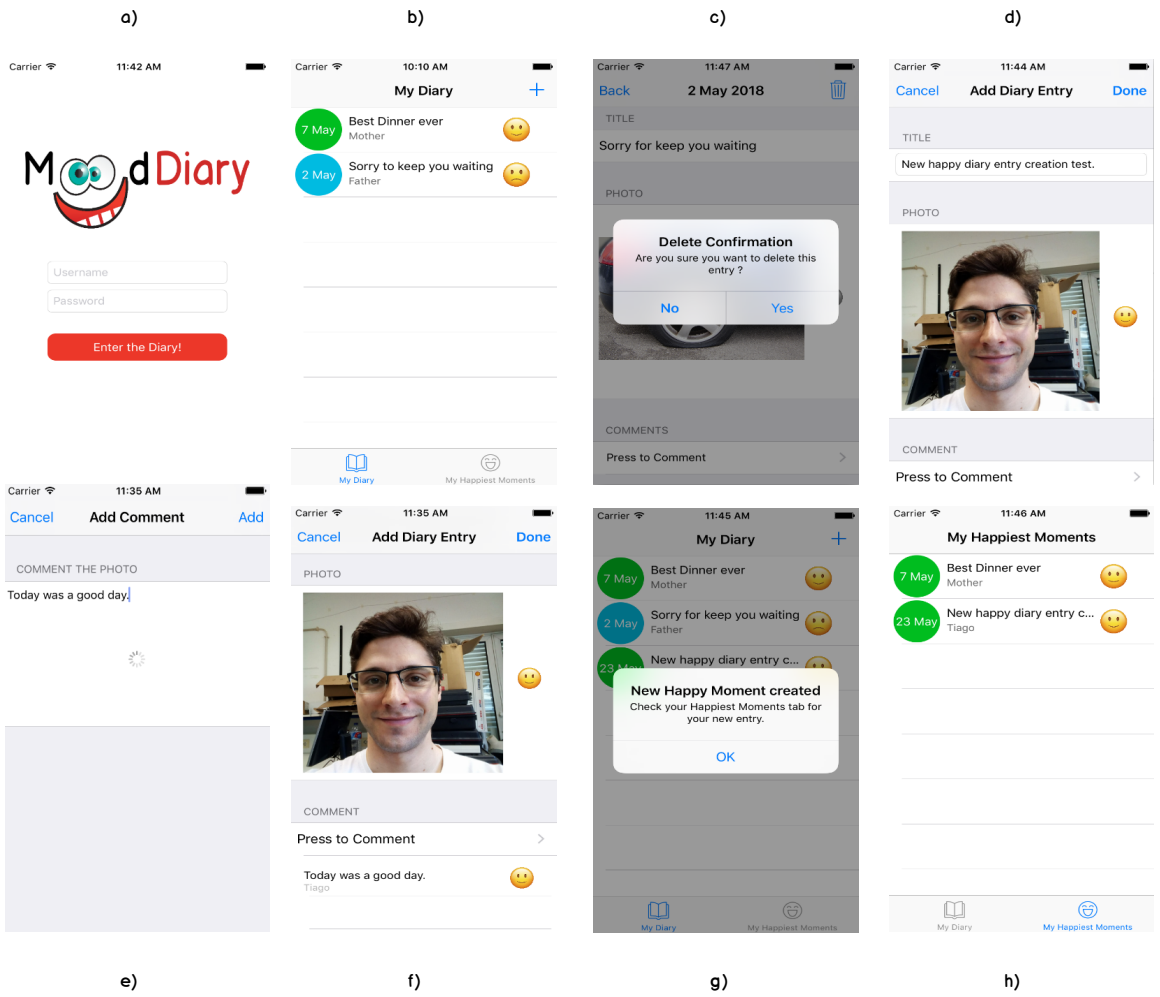

Figure 4.12: Example of some views of the mobile application: A - Initial application login screen; B - My Diary section view controller; C - Alert confirmation when deleting diary entries; D - Add diary entry view with emotion tagging on an image; E - Add comment view with an activity status spinner (waiting for the affective response by the modality); F - Add diary entry view with emotion tagging on image and comment; G - Alert on the creation of a new happy moment (it occurs when a diary entry is associated with the "happiness" emotion); H - "My Happiest Moments" view controller.

Emotions to emojis matching

For the representation of emotions, emojis were chosen to make the application interface more appealing (replacing text for images). Figure 4.13 shows the association between emotions and emoji used in the application. In the first version of the application, when no emotion could be detected by the modality, no emoji association was being made, but after some discussion, we came to the conclusion that a suggestive emoji should be used to represent this particular situation.

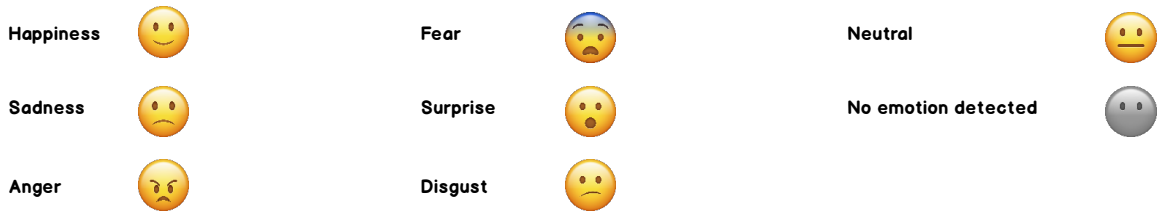

Figure 4.13: Visual representation of emotions through emojis representation.

4.2.4 Evaluation

The overall system and the modality were then placed under assessment in order to improve the application layout, making it more appealing and user-friendly. The video data acquisition application intends to represent a remote acquisition sensor, therefore, it does not integrate our evaluation process. A case study was carried out in our research institute *(IETA)*, aiming to test, in a first stage, the modality and the mobile application (MoodDiary), before inserting them in the envisaged scenario.

The modality and the developed application were placed under 2 evaluations: a **heuristic** and an **usability** evaluation.

Firstly, a heuristic evaluation was carried out with the main goal to try to perceive some application visual interface elements that make less sense, such as the colours used or icons position or size, for example, and at a later stage, it was then performed the usability evaluation, with some of the suggestions of the heuristic evaluation already considered and solved.

Heuristic Evaluation

Heuristic evaluation [52] is a usability mechanism used to find usability problems in an interface. This evaluation method involves a small number of evaluators that review the interface and compare it against a set of usability principles *(the heuristics)*. With this approach, it is intended to identify any issues associated with the design and development of user interfaces. Among several sets of heuristics, we chose to follow the *10* usability heuristics for interaction design defined by Jakob Nielsen[6] in *1994*.

In this first phase, in order to evaluate the set of heuristics considered, we recruited four evaluators, all Master students in a background on applying heuristic evaluation obtained from a previous course in Human-Computer Interaction. The evaluators were asked to interact with the mobile application, without any type of tasks or scenarios to complete, aiming to find usability problems that we were not considering.

[6] goo.gl/m33Yhy

Due to the difficulty of obtaining the best usability results through the evaluation of only a single user, we decided to conduct the heuristic evaluation [53] using *4* evaluators, increasing the chances of finding the greatest number of usability issues. Each evaluator was asked to detect any violation of the provided heuristic (Table 4.14 depicts the problems detected by each evaluator) and, whenever possible, proposed solutions. Among the *4* evaluators, we found *13* usability problems, and for the sake of brevity, in table 4.5 we summarise the proposed solutions for those problems.

Problem	**Description**	**Heuristic**	**Severity**
A	The circular date icon should change its colour regarding the calculated emotion of the diary entry	Aesthetic and minimalist design	1
B	Change the Diary Entry View Controller header from Description to Title.	Match between system and the real world	1
C	The insertion of new comments should be triggered by a click on a suggestive button icon instead of the current approach.	Flexibility and efficiency of use	1
D	The mechanism to delete diary entries and comments is not very clear and only iOS consumers are familiarized with it.	Flexibility and efficiency of use	2
E	When no emotion is detected by the modality, no visual feedback is given in the application. An emoji should be chosen to represent the specified situation.	Recognition rather than recall	2
F	There's no message or alert mechanism when users chose to cancel a new diary entry.	Error Prevention	3
G	Instead of text, the application should present standard icons that represent the same particular action or field.	Match between system and real world	1
H	The system does not provide other languages options to users.	Match between system and real world	2
I	Icons that represent actions (such as add or return actions) should be the standard used by Apple.	Consistency and standards	1
J	In some parts of the application, the best words were not chosen to describe actions or present information.	Consistency and standards	1
L	The application does not provide a mechanism that informs the user that the modality is detecting emotions.	Visibility of system status	2
M	The error messages are not clear enough and don't inform the user the required information that describes the error.	Help users recognize, diagnose, and recover from errors	2
N	The application does not present the user confirmation, alerts, for example, when users' decide to delete diary entries or comments, which can result in the deletion of undesired data.	Error Prevention	3

Table 4.5: Usability Problems found by the 4 evaluators, along with its heuristic classification.

From the analysis of table 4.6, the problems ***D***, ***E*** and ***N*** were those that were considered by the evaluators more times (with *3* occurrences in *4* possible ones). According to [54], the severity level varies from value *0* to *4*. In this context, for each problem, a severity level was defined as the worst value on that rating scale, among all severity levels considered by the evaluators, where the value "0" corresponds to "I dont agree that this is a usability problem at all" and the value "4" corresponds to "Usability catastrophe: imperative to fix this before product can be released". The problem N was the first to be corrected because it was considered with severity level of *3*. The remaining problems (D and E), were also corrected, although they were considered only severity level of *2*.

Evaluator	A	B	C	D	E	F	G	H	I	J	L	M	N
1	X	X	X	X	X	X							X
2				X	X		X	X			X		
3			X			X			X	X	X		X
4				X	X							X	X

Table 4.6: Detected usability problems by each one of the 4 evaluators.

Usability Evaluation

Usability testing [55] aims to evaluate a product, service or application by giving participants a set of tasks that need to be completed. With this approach [56], we intended to collect data so it can be analysed to determine the participant's accomplishment and satisfaction while interacting with the product, service or application.

For the usability evaluation, we recruited 10 volunteers (all of them Computer and Telematics Engineering students) to be part of our evaluation process (average age of 23 years old). Each volunteer had to perform a set of tasks and, in the end, they were asked to fill a Post-study System Usability Questionnaire (PSSUQ) [57], which is divided into 19 questions that need to be answered using a 7-point Likert-like scale where the value "1"corresponds to "I totally agree"and the value "7"corresponds to "I totally disagree"(questionnaire can be found here[7]). At the end of all tasks, participants' data were collected, by an observer, and grouped so it could be analysed later. A questionnaire related to PSSUQ is divided into 4 groups [57]:

- **General**: Questions 1 to 19;
- **System usability**: Questions 1 to 8;
- **System information quality**: Questions 9 to 15;
- **Interface quality**: Questions 16 to 18;

The tasks to be performed during the usability evaluation were designed in order to create a natural and sequential flow that allowed the volunteers to perceive the features of the application, and its adaptability concerning the users' emotional state in the depicted scenario. To specify the tasks, we adopted a scenario that considers a previous use of the application by the child's parents and, thus, it is populated with some contents they created. The task list description is presented on table 4.8, that additionally, points the possible steps to be performed by participants for each task. The PSSUQ questionnaire results, related to questions 1 to 19, is presented on figures 4.16 and 4.17.

In table 4.7, it is possible to see, the results achieved for each component of the PSSUQ questionnaire. The results were very positive, which revealed that all features that would

[7]https://goo.gl/forms/VSVrm0UbCZKNYQNj1

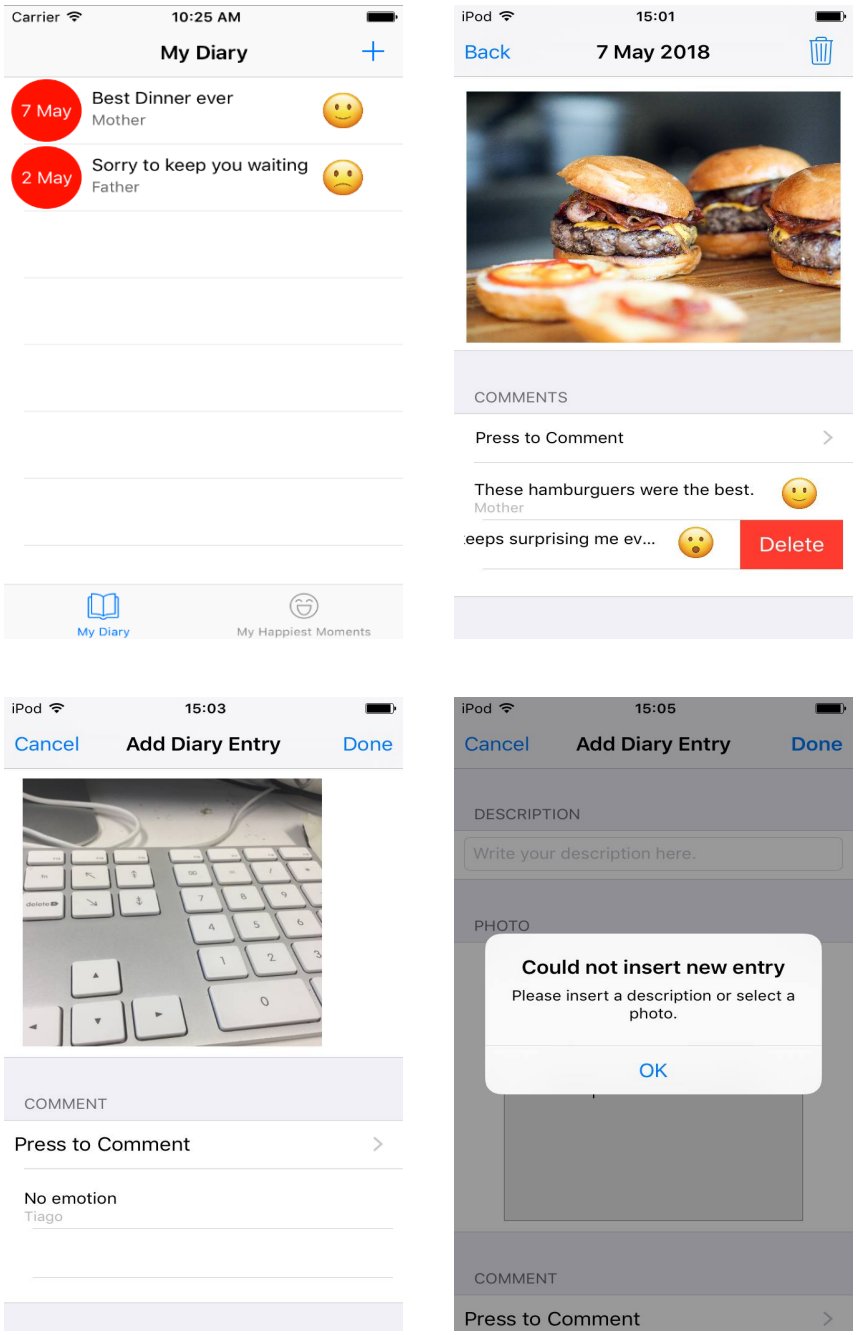

Figure 4.14: Visual representation of the problems referred by participants on the heuristic evaluation. On top-left: representation of Problem A; on top-right: representation of Problem D; on bottom-left: representation of Problem E; on bottom-right: representation of Problem M.

---

be expected in an application of this type are integrated, providing a good experience of interaction to the participants and a increased level of adaptability by the application to the participants' emotional state. Concerning the mean results, the group that obtained the worst

results refers to the interface quality (**1,400**), which demonstrates the special care that must be taken into account during the design of interactive applications. On the other hand, the system usability was the group with the best results (**1,325**), which to some extent proves the usefulness in considering the evaluation of heuristics previously performed. Concerning the median results, all groups have the same results (**1,000**). On figure 4.15, it is presented a volunteer while performing the usability evaluation, while he interacts with the mobile application ***MoodDiary***.

Group	**PSSUQ mean**	**PSSUQ median**
General	1.353	1.000
System usability	1.325	1.000
System information quality	1.357	1.000
Interface quality	1.400	1.000

Table 4.7: Results of the PSSUQ questionnaire made.

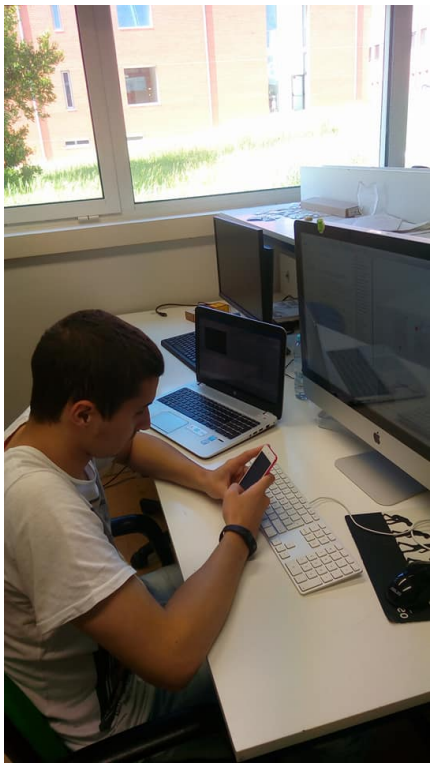

Figure 4.15: A volunteer while performing the usability test, while interacting with the mobile application (MoodDiary).

Task	Description	Possible Steps
A	Login into the application	Insert username and password Press "Entry the Diary" button
B	Identify the "MyDiary" section	Press the "MyDiary" button
C	The diary entry created by the "mother" should be edited	Select the "mother" diary entry Delete the "mother" comment Delete the "father" comment Create a new comment
D	Create a new diary entry	Press the Add button Insert title Take a picture Insert a comment Press done
E	Remove the diary entry created by the "father"	Slide right the "father" diary entry to delete Select the "father" diary entry Press the delete button
F	Frame the face with the the laptop's webcam (where the video data acquisition application runs) and express a sad face for a while	Frame the face with the laptop's webcam
G	Leave the application	Press the menu device button

Table 4.8: Tasks description and possible steps to be performed by participants during the usability evaluation.

4.3 Discussion and Conclusions

This chapter presented two demonstrator applications developed to support the development of the proposed affective modality and illustrate its potential.

Affective Spotify System

This demonstrator application allowed the first integration of the affective modality in a multimodal system and enabled to evolve the modality and assess the different technical aspects of its development. In this first prototype application, all modules ran locally (speech and gestures modalities, IM, fusion engine and affective modality).

The features considering a sensing database, containing data from sensing technology available in the environment (e.g., *EDA, ECG*) and the integration with research-stage methods (e.g., from our groups' research [58], providing advanced methods to process additional types of data, were not addressed at this development stage.

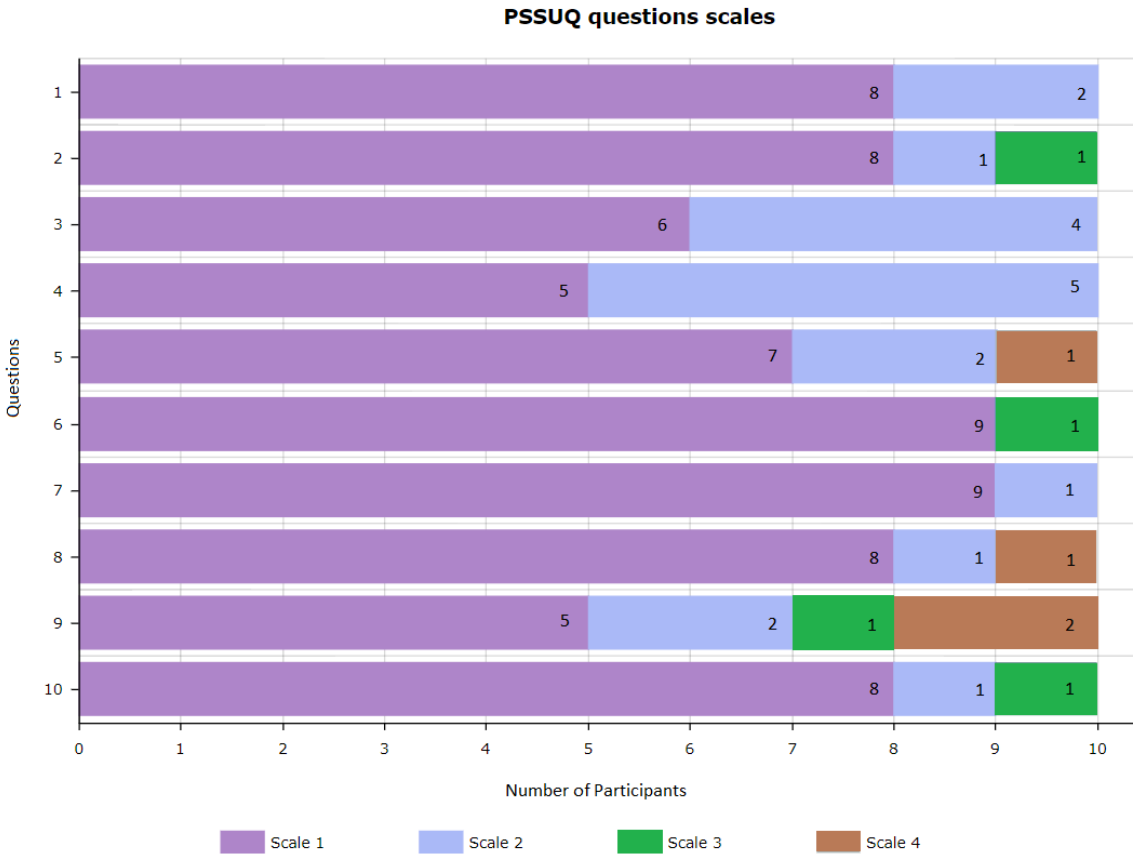

Figure 4.16: PSSUQ questionnaire results from question 1 to 10, referring to the 10 participants.

MoodDiary System

On the other hand, the second demonstrator system enabled to prove the ability to integrate the proposed modality into new applications and OS's (mobile application). Here, a different approach from Affective Spotify has been used, and a cloud-based IM runs remotely and the mobile application connects to that IM. Additionally, it made possible to illustrate part of the conceptual vision depicted on figure 3.2, location A, where data acquired remotely is being used to infer emotional states on another application.

For demonstrative purposes, and considering that we have not mechanisms to identify people using any sort of data (e.g., through facial expression, audio content), we assume that the user interacting with both applications (mobile and desktop applications) is the same.

The simple evaluation carried out provided good feedback on the overall application. Naturally, the main purpose was not to assess if emotion was properly being detected (as this goes beyond the scope of our work) or considered, at this early stage, but to have a first validation of the application concept. The obtained feedback should be used, in the future, along with a stronger work on the user characterization and scenarios, to evolve the concept in closer relation with the end-users and stakeholders.

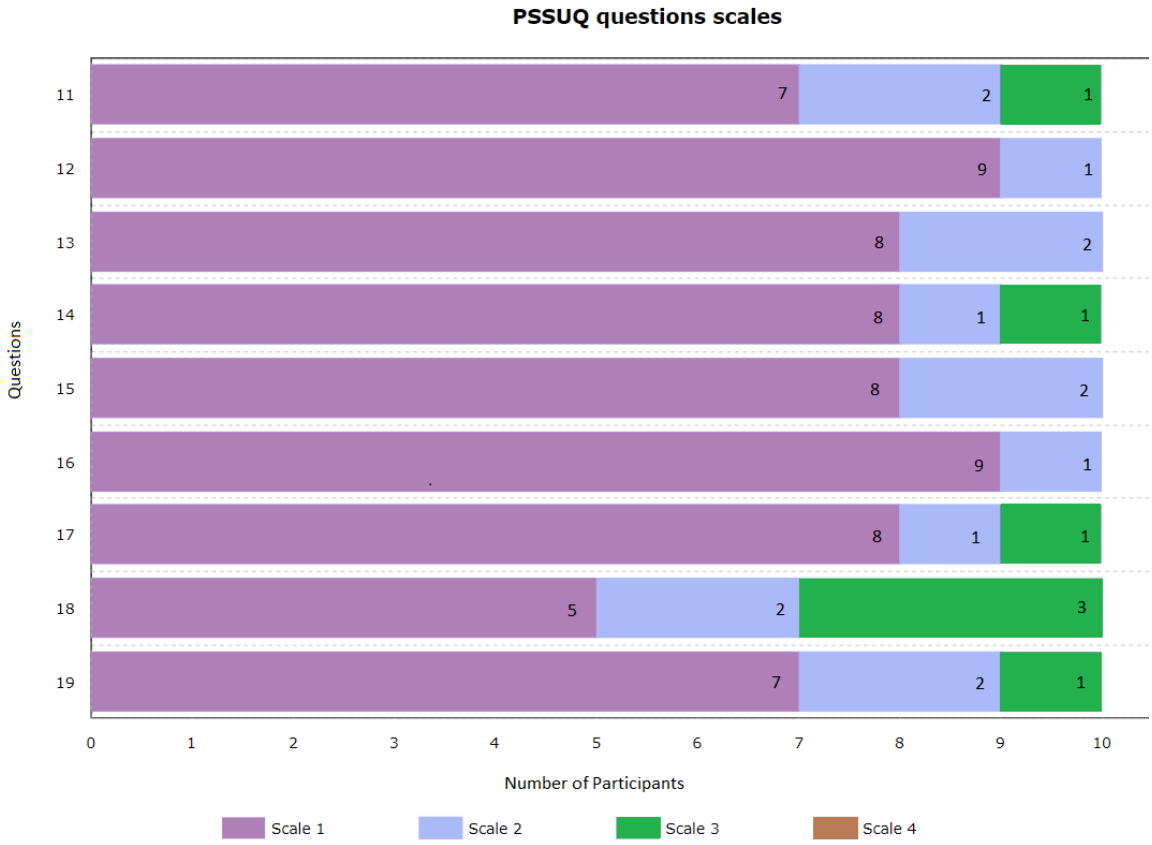

Figure 4.17: PSSUQ questionnaire results from question 11 to 19, referring to the 10 participants.

Chapter 5

Conclusions

In this final chapter, we perform an overall summary of the work carried out and analyse its outcomes in light of the initial goals. Some ideas regarding future work are also addressed, in order to improve or expand the developed work.

5.1 Overall Analysis

In light of the work carried out, presented in this document, we consider that the goals established for this project have been successfully reached. We propose a conceptual vision encompassing how emotion can be brought into interactive contexts and how the efforts in researching and developing affective computing methods can be brought closer to the application level.

Aligned with our vision, we present an affective modality, a generic decoupled solution to provide emotional context to interactive applications. By placing most of the complexity of integrating emotion extraction methods in remote services or toolkits, the developer only needs to integrate the developed modality and deal with the semantic contents of the events that reach the application. With this solution, we reduce the complexity and time required by developers to consider emotions in their own applications. The system was designed and developed in a scalable manner, meaning that new methods, mechanisms or services can be easily added to enable extracting emotional states from data types not considered, at this stage, and without requiring changes to the systems already integrating the affective modality.

The developed proof-of-concept applications (*Affective Spotify*, in a preliminary stage, and *MoodDiary* in a final stage) prove the successful integration of the affective modality in different applications with different OS's. There are many scenarios and environments where the modality can be applied, and the developer is only required to define how to deal with the emotional context, whether it be by reacting to the implicit input or by annotating a particular task, content or action with its emotional impact.

Our option of using a real multimodal interactive application, Affective Spotify, as a test bed for the development of the affective modality proved to be an important decision since it made us aware, sooner, of potential "real-world" difficulties. Additionally, it offered a more realistic scenario for integration with the multimodal architecture, which then translated into an easier transition into the novel application, MoodDiary.

The application MoodDiary is our effort to provide additional evidence of the affective modality features and demonstrate how the vision initially proposed can be instantiated, high-

lighting, for instance, the possibility of data being acquired independently of the application (and device) to extract the user's emotional status.

MoodDiary was not intended as an effort to fully address the issues regarding emotion in children with autism spectrum disorders. To that effect, the consideration of the previously validated Personas and scenarios, although a good starting point, should be further expanded, for instance, by more explicitly including the potential users and stakeholders in defining the motivations, scenarios and requirements. Nevertheless, we consider that MoodDiary can already support conceptual validation, for instance regarding how to consider emotion in the envisaged scenarios, an adequate starting point for the discussion with different end-users. The short evaluation cycle carried out drew its motivation from this purpose. Based on the collected data, and from the questionnaire results obtained, we can conclude that the work carried out provides a promising ground for further development in this field.

Naturally, the time constraints associated with this work demanded a pragmatic approach that, in some situations, left room for future improvement, and its innovative nature has also created new routes to follow, as discussed in the next section.

5.2 Future Work

The main requirements for our work were met. However, more functionalities can be added to the modality in order to make it more robust and enable it to handle more data types and sensor measurements.

In the Affective Spotify demonstrator application, the affective modality is not being used on the fusion engine. By merging events from such modality with speech or gestures input, several conclusions can be accomplished on whether our mood is directly related with specific actions (e.g., users often listen to a playlist when they are sad).

In the MoodDiary proof-of-concept, and for the sake of simplicity, we are considering that the user who is interacting with the application is the same who is being filmed by the webcam. Since this data is being acquired independently, and stored in a database, in a real world scenario, assuming such a match is limiting. This evidences the need for a mechanism that could ensure proper match between the data considered for extracting emotions and the user. In the case of video (and the same rationale might be considered for audio or even ECG), a service such as Microsoft's Face API could provide face verification and identification features used to tag the video data obtained independently (e.g., a video camera in a building lobby), storing the user identification for later consideration when retrieving data for the affective modality.

The MoodDiary proof-of-concept application could also be improved in some aspects such as enabling users to share diary contents with friends or family members, however, considering that the development of applications was not the scope of our work, such improvement was not fulfilled.

Adding to the contributions made, by the presented work, to the field of multimodal interaction considering the affective state, it also opens a fast track to improve the translational nature of the research in affective computing methods. This can be performed by deploying services encapsulating novel methods that, in turn, can be seamlessly integrated and updated in the affective modality. One example includes the possibility of the modality to integrate emotional states derived from physiological signals (e.g., *ECG, EMG, EDA*), profiting from work developed on our research institute [58]. With this strategy, emotions inferred from data

mining mechanisms on such signals can be added to the modality backend system, making them available to the applications connected with the modality, thus providing a faster way into more realistic testing scenarios.

Bibliografia

[1] Albrecht Schmidt. Implicit human computer interaction through context. *Personal Technologies*, 4(2):191–199, June 2000.

[2] Peck D. Hincks E. Jacob S. Robert J. K. Solovey T., Afergan E. Designing Implicit Interfaces for Physiological Computing. *ACM Transactions on Computer-Human Interaction*, pages 1–27, 2015.

[3] Lisa Feldman Barrett. *How emotions are made: The secret life of the brain.* Houghton Mifflin Harcourt, 2017.

[4] Chai M Tyng, Hafeez U Amin, Mohamad NM Saad, and Aamir S Malik. The influences of emotion on learning and memory. *Frontiers in psychology*, 8:1454, 2017.

[5] Hamed S. Alavi, Himanshu Verma, Michael Papinutto, and Denis Lalanne. Comfort: A coordinate of user experience in interactive built environments. In *IFIP Conference on Human-Computer Interaction*, pages 247–257. Springer, 2017.

[6] Valerie Gay, Peter Leijdekkers, Johann Agcanas, Frederick Wong, and Qiang Wu. CaptureMyEmotion: helping autistic children understand their emotions using facial expression recognition and mobile technologies. *Studies in Health Technology and Informatics*, 189:71–76, 2013.

[7] Madalina Sucala, Pim Cuijpers, Frederick Muench, Roxana Cardo, Radu Soflau, Anca Dobrean, Patriciu Achimas-Cadariu, and Daniel David. Anxiety: There is an app for that. *Depression and Anxiety*, 34:518–525, 2017.

[8] Arvid Kappas. Smile when you read this, whether you like it or not: Conceptual challenges to affect detection. *IEEE Transactions on Affective Computing*, 21:38–41, 2010.

[9] Alexandru Popescu, Joost Broekens, and Maarten Van Someren. GAMYGDALA: an emotion engine for games. *IEEE Transactions on Affective Computing*, 5:32–44, 2014.

[10] Rosalind W Picard. Emotion research by the people, for the people. *Emotion Review*, 2(3):250–254, 2010.

[11] Tara J Brigham. Merging technology and emotions: Introduction to affective computing. *Medical reference services quarterly*, 36(4):399–407, 2017.

[12] Rana El Kaliouby, Rosalind Picard, and Simon Baron-Cohen. Affective computing and autism. *Annals of the New York Academy of Sciences*, 1093:228–248, 2006.

[13] Byoung Chul Ko. A brief review of facial emotion recognition based on visual information. *Sensors*, 18(2):401, 2018.

[14] Rafael A Calvo and Sunghwan Mac Kim. Emotions in text: dimensional and categorical models. *Computational Intelligence*, 29(3):527–543, 2013.

[15] Nielsole Ole Bernsen and Laila Dybkjaer. *Designing Interactive Speech Systems: From First Ideas to User Testing*. Springer-Verlag New York, Inc., December 1997.

[16] Bruno Dumas, Denis Lalanne, and Sharon Oviatt. Multimodal interfaces: A survey of principles, models and frameworks. In *Human machine interaction*, pages 3–26. Springer, 2009.

[17] Sharon Oviatt, Rachel Coulston, and Rebecca Lunsford. When do we interact multimodally?: Cognitive load and multimodal communication patterns. In *Proc. 6th Int. Conf. on Multimodal Interfaces*, pages 129–136, New York, NY, USA, 2004.

[18] Multitouch tabletop technology for people with autism spectrum disorder: A review of the literature. *Procedia Computer Science*, 14:198 – 207, 2012.

[19] Nuno Almeida, Samuel Silva, António Teixeira, and Diogo Vieira. Multi-device applications using the multimodal architecture. In *Multimodal Interaction with W3C Standards: Toward Natural User Interfaces to Everything*, pages 367–383. November 2017.

[20] Deborah A. Dahl. The W3C multimodal architecture and interfaces standard. *Journal on Multimodal User Interfaces*, 7(3):171–182, April 2013.

[21] Deborah A Dahl. *Multimodal Interaction with W3C Standards*. Springer, 2016.

[22] A. Teixeira, N. Almeida, C. Pereira, M. O. e Silva, D. Vieira, and S. Silva. Applications of the multimodal interaction architecture in ambient assisted living. In *Multimodal Interaction with W3C Standards*, pages 271–291. Springer, 2017.

[23] Nuno Almeida, Samuel Silva, and António Teixeira. Design and development of speech interaction: a methodology. In *International Conference on Human-Computer Interaction*, pages 370–381. Springer, 2014.

[24] Joel Birnbaum. Pervasive information systems. *Commun. ACM*, 40(2):40–41, February 1997.

[25] T. Chaari, D. Ejigu, F. Laforest, and V. M. Scuturici. Modeling and using context in adapting applications to pervasive environments. In *Proc. ACS/IEEE International Conference on Pervasive Services*, pages 111–120, June 2006.

[26] Davies N Friday A. Cheverst K, Blair G. The support of mobile awareness in collaborative groupware. *Personal Technologies*, 3:33–42, 1999.

[27] Janine D. Flory, Katri Rikknen, Karen A. Matthews, and Jane F. Owens. Self-focused attention and mood during everyday social interactions. *Personality and Social Psychology Bulletin*, 26(7):875–883, 2000.

[28] Stephanie Chamberlain, Helen Sharp, and Neil Maiden. Towards a framework for integrating agile development and user-centred design. In Pekka Abrahamsson, Michele Marchesi, and Giancarlo Succi, editors, *Extreme Programming and Agile Processes in Software Engineering*, pages 143–153, Berlin, Heidelberg, 2006.

[29] Interaction Design Foundation. Personas: A simple introduction. `goo.gl/HJJPmi`. Accessed: June, 2018.

[30] Jan Gulliksen, Ann Lantz, and Inger Boivie. User centered design in practice - problems and possibilities. April 1999.

[31] Rudy Den Buurman. User-centred design of smart products. *Ergonomics*, 40(10):1159–1169, 1997.

[32] UX Planet. User personas, scenarios, user stories and storyboards: Whats the difference? `goo.gl/YS9hKS`. Accessed: June, 2018.

[33] Adolphs R. Kennedy DP. Perception of emotions from facial expressions in high-functioning adults with autism. *Neuropsychologia*, pages 3313–3319, 2012.

[34] Sanna Kuusikko, Helena Haapsamo, Eira Jansson-Verkasalo, Tuula Hurtig, Marja-Leena Mattila, Hanna Ebeling, Katja Jussila, Sven Bölte, and Irma Moilanen. Emotion recognition in children and adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 39(6):938–945, June 2009.

[35] W. E. Jones and Ami Klin. Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature*, 504 7480:427–31, 2013.

[36] P. Washington, C. Voss, N. Haber, S. Tanaka, J. Daniels, C. Feinstein, T. Winograd, and D. Wall. A Wearable Social Interaction Aid for Children with Autism. pages 2348–2354, 2016.

[37] Braun KV et al. Christensen, DL. Baio J. Prevalence and characteristics of autism spectrum disorder among children aged 8 years autism and developmental disabilities monitoring network. *MMWR Surveill Summ 2016*, 65:123, 2012.

[38] James W. Tanaka, Julie M Wolf, Cheryl Klaiman, Kathleen König, Jeffrey Cockburn, Lauren E Herlihy, C Kent Brown, Sherin S. Stahl, Mikle South, James C. McPartland, Martha D. Kaiser, and Robert T Schultz. The perception and identification of facial emotions in individuals with autism spectrum disorders using the let's face it! emotion skills battery. *Journal of child psychology and psychiatry, and allied disciplines*, 53 12:1259–67, 2012.

[39] Rosalyn Adamowycz and Sorcha Parker. Interpreting social contexts and emotions and ASD. *Procedia - Social and Behavioral Sciences*, 93:1148 – 1153, 2013.

[40] David Beukelman and P. Mirenda. *Augmentative and Alternative Communication*. January 2012.

[41] Harini Sampath, Bipin Indurkhya, and Jayanthi Sivaswamy. A communication system on smart phones and tablets for non-verbal children with autism. In Klaus Miesenberger,

Arthur Karshmer, Petr Penaz, and Wolfgang Zagler, editors, *Computers Helping People with Special Needs*, pages 323–330, Berlin, Heidelberg, 2012.

[42] Dorothea Lerman, Christina Vorndran, Laura Addison, and Stephanie Kuhn. A rapid assessment of skills in young children with autism. *Journal of applied behavior analysis*, 37:11–26, February 2004.

[43] Peter Leijdekkers, Valérie Gay, and Frederick Wong. CaptureMyEmotion: A mobile app to improve emotion learning for autistic children using sensors. In *Proc. 26th IEEE Int. Symposium on Computer-Based Medical Systems*, pages 381–384, 2013.

[44] James Tanaka, Julie Wolf, Cheryl Klaiman, Kathleen Koenig, Jeffrey Cockburn, Lauren Herlihy, Carla Brown, Sherin Stahl, Martha Kaiser, and Robert Schultz. Using computerized games to teach face recognition skills to children with autism spectrum disorder: The Let's Face It! program. *Journal of child psychology and psychiatry, and allied disciplines*, 51:944–52, August 2010.

[45] Agata Kołakowska, Agnieszka Landowska, Mariusz Szwoch, Wioleta Szwoch, and Michał R Wróbel. Modeling emotions for affect-aware applications. In Stanisław Wrycza, editor, *Information systems Development and Applications*, pages 210 – 227. Faculty of Management, University of Gdansk, 2015.

[46] Christine Mohn, Heike Argstatter, and Friedrich-Wilhelm Wilker. Perception of six basic emotions in music. 39, October 2010.

[47] N. Almeida, A. Teixeira, S. Silva, and J. Freitas. Towards Integration of Fusion in a W3C-based Multimodal Interaction Framework: fusion of events. *Proc. Iberspeech, pp.*, 1:291–300, 2016.

[48] Jim Barnett. Introduction to SCXML. In *Multimodal Interaction with W3C Standards*, pages 81–107. Springer, 2017.

[49] Alan Cooper, Robert Reimann, and Dave Cronin. About face 3.0: The essentials of interaction design. johnwiley & sons. *Inc., Indianapolis, Indiana, USA*, 2007.

[50] Leal A. Silva S., Teixeira A. On the creation of a persona to support the development of technologies for children with autism spectrum disorder. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9739:213–223, 2016.

[51] Almeida N. Silva S. Teixeira A. Vieira D., Leal A. Tell your day: Developing multimodal interaction applications for children with ASD. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10277 LNCS:525–544, 2017.

[52] Usability.gov. Heuristic evaluations and expert reviews. `goo.gl/javT8B`. Accessed: June, 2018.

[53] Nielsen Norman Group. How to conduct a heuristic evaluation. `goo.gl/8dWD7e`. Accessed: June, 2018.

[54] Nielsen Norman Group. Severity ratings for usability problems. goo.gl/YS9hKS. Accessed: June, 2018.

[55] Usability.gov. Usability testing. goo.gl/47ePFi. Accessed: June, 2018.

[56] UX Mastery. How to conduct usability testing from start to finish. goo.gl/tg9YUp. Accessed: June, 2018.

[57] A. F. Rosa, A. I. Martins, V. Costa, A. Queirós, A. Silva, and N. P. Rocha. European portuguese validation of the post-study system usability questionnaire (pssuq). In *2015 10th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–5, June 2015.

[58] J. Ferreira, S. Brás, C. F. Silva, and S. C. Soares. An automatic classifier of emotions built from entropy of noise. *Psychophysiology*, 54(4):620–627, 2017.

Appendix

.2 Public Presentation

The following poster (figure 1) was created to present the dissertation work in the public event during 2018, students@deti, at Aveiro University.

students @deti 2018

A systematic approach for the integration of emotional context in interactive systems

Tiago de Figueiredo Henriques
Orientadores: Prof. Samuel Silva, Prof. Susana Brás

Dissertação em Engenharia de Computadores e Telemática.

Abstract

Knowing how a user is reacting to a system can improve our ability to adapt, whether to change a content, if the user is not liking it, or implementing measures to avoid negative moods or anxious states. Even though the area of affective computing is progressing at a fast pace, and there are already toolkits supporting the extraction of emotional status, developing emotionally-aware applications is still a challenge. In this work, we propose how to systematically integrate emotion into applications by proposing an affective modality as part of a multimodal interaction framework.

Keywords

Affective computing; Multimodal interaction, User-centred design; Interaction manager; Cloud services; Emotionally-aware; Modalities;

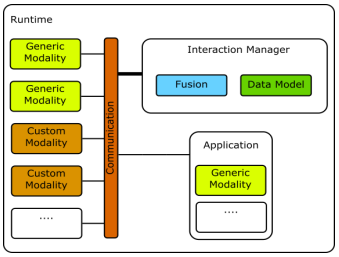

Fig 1 - Main components of an architecture supporting multimodal interaction, abiding to the W3C recommendations.

Affective Modality

The affective modality communicates with a module of a multimodal framework (interaction manager) and contains an affective hub (see Fig. 2):

- **Affective Modality -** sends, receives and processes events, from and to the Interaction manager (IM);
- **Affective Hub** – directs a request for the best suitable service and provides a response with an affective basis;

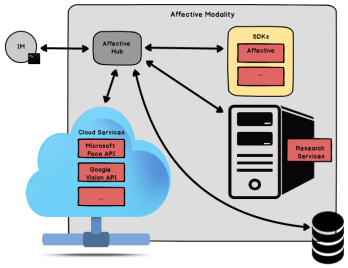

Fig 2- Depiction of the overall components for an affective modality.

Multimodal Architecture

The W3C multimodal architecture proposal has 3 notable components (see Fig. 1):

- **Modalities –** handles interaction inputs and outputs to the system;
- **Interaction Manager –** manages the events that comes from the modalities and communicates with the application logic;
- **Data Model** – stores information about the current state of the information;

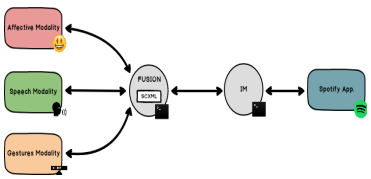

Fig 3 – Main component blocks of the overall architecture of the proposed Affective Spotify application.

Proof-of-concept Applications

Integration of emotional context in:

- **Affective Spotify**, enabling multimodal interaction (voice and gestures) with Spotify (see Fig. 3). The application detects, through facial recognition, the user's emotions, in real-time, and reacts to the current user mood;
- **MoodDiary**, - a virtual diary, developed considering special needs in emotion understanding and communication (see Fig. 4). The application goal is to consider the user emotional status to associate emotions with content, using the modality capabilities to annotate emotions present, for instance, in photos or text.

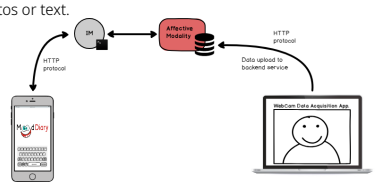

Fig 4 – Multi-device system deployment. The mobile application uses the HTTP protocol to communicate to a cloud-based IM and the desktop application uploads data to the modality backend system.

Affective Spotify

MoodDiary

universidade de aveiro
theoria poiesis praxis

deti departamento de electrónica, telecomunicações e informática

Figure 1: Poster of the generic affective modality and demonstrator applications on the student@deti public event.