Digital Emotions: the potential and issues of Affective Computing systems

Nerina Reitano

M.Sc. Interactive Digital Media

2018

Digital Emotions:

the potential and issues of Affective Computing systems

Nerina Reitano

A research paper submitted to the University of Dublin,

in partial fulfilment of the requirements for the degree of

Master of Science Interactive Digital Media

2018

Declaration

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at: http://www.tcd.ie/calendar

I have also completed the Online Tutorial on avoiding plagiarism 'Ready, Steady, Write', located at http://tcdie.libguides.com/plagiarism/ready-steady-write

I declare that the work described in this research Paper is, except where otherwise stated, entirely my own work and has not been submitted as an exercise for a degree at this or any other university.

Signed: ______________________

Nerina Reitano

11.05.2018

Permission to lend and/or copy

I agree that Trinity College Library may lend or copy this

Research Paper upon request.

Signed: ______________________

Nerina Reitano

11.05.2018

Acknowledgments

I would like to thank my supervisor Katja Zibrek for her kindness, help and guidance throughout the whole process.

Abstract

The last decade has witnessed a rapid evolution of the internet and an increased flow of information, which have encouraged the development of systems that try to distinguish between useful and non-useful information for users. If this growth has benefited the society, then by offering computers the opportunity to adapt their content to individuals' needs, systems based on information filtering have also facilitated the access to personal data, allowing computers and smartphones to *know* their users. The field of Affective Computing aims to use human emotions as a means to communicate preferences to machines, in order to lead to a more natural and efficient communication between humans and computers. The biggest challenge for Affective Computing is to understand how computers, which are supposed to be rational and precise, can understand, react and eventually express emotions. This research paper includes an analysis of the process of Affective Generation, namely the capacity for a device to understand the affective state of its user. It examines Affective Computing systems able to detect facial expressions, speech, body movements, and physiological information, with the purpose of evaluating which method has the greatest potential for improving the human-computer interaction. The limits of Affective Computing systems are dictated by technical issues, and especially by privacy threats, since emotional machines would be able to collect a much higher amount of information about their users than computers do at this stage. The purpose of this research paper is to identify and discuss the ethical issues concerning Affective Computing devices, and to assess if and how those issues can be overcome.

Table of Contents

1. Introduction

> In the last decade, the number of Information appearance (e.g., via radio, television, Internet) has grown tremendously. This development can be observed in almost every mode of iTV systems. The only strategy to control this flood of information is to filter it.
> Individuals vary in their interests and needs, thus information filtering strategies which seek to meet these individual profiles will differ. Each information filtering strategy is based on the personal knowledge and experience of the individual user (Wittig, 1999, p. 10).

The last ten years have witnessed a rapid evolution of the internet and an increased flow of information, especially due to the advent of smartphones and social media. This development has induced the need to filter data to offer individuals' the opportunity to deal with only useful information. It has become essential for systems to *get to know* their users, namely to obtain information about what the person likes and dislikes, which are his or her interests, and how users spend their free time. The systems keep track of users' activities to be able to create and update a profile. User data from almost every webpage is traced and collected by other subjects through various technologies, from cookies to more refined techniques that assign to each user a sort of *digital fingerprint.*

In the latest versions of iOS, Apple has implemented a section that tracks and memorises the user's movements, by showing a list of five geographical areas where the user has spent the majority of his time. By selecting one of the places, the user gets access to a map that displays the detailed positions with times. Facebook uses an algorithm called EdgeRank that aims to analyse the *feelings* that users have for each other. The algorithm examines how often and how intensively people interact with each other, with the aim of identifying who is important to whom. If a user visits a specific place every day between 8 pm and 8 am, then that place might probably be his or her home. If two users interact with each other every day, then they are probably friends. If a user never visits another user´s Facebook profile and never interacts with them, then she or he will probably not be particularly interested in seeing that user's pictures, which means that the information can be filtered out. Algorithms make decisions for users, and these decisions are based on assumptions, that turn out to be true in most cases. It is highly probable that computers will be able to analyse and interpret images, sentences or sounds published on social media, and therefore to collect more information about their users: who they are, what they like, how they spend their free time, to which advertising messages they might be most susceptible.

However, at this stage, machines are not able to independently learn the information needs of their users. Computers might be capable of filtering information according to a user's general interests, but they cannot adapt their content to his/her temporary needs, which makes the interaction between human and computers not particularly spontaneous and efficient. By combining computer science, psychology and cognitive science, Affective Computing aims to allow individuals to communicate their preferences with emotions, and therefore to express their needs in a natural and instant way. Ideally, machines will be able to identify their users' mood, and respond by giving them exactly what they need in a specific moment. A music player should suggest us songs according to our affective state, virtual tutors should be able to detect if we are interested or bored while doing a task, email services might tell us if the sender of an email was happy, or angry, and so on. Basically, human-computer interaction should simulate the human-human interaction, in which individuals are able to understand each other's emotions. Computer scientists have noticed that the same methods used by humans to express their affective state can be translated into something that computers can understand. This research paper analyses the processes of emotion detection through facial expressions, vocal behaviour, gestures and body movement, and physiological measurements, with the purpose of identifying which method has the highest potential to improve the interaction between humans and computers. However, if communicating affect to machines has the advantage of simplifying individuals' lives, it also entails a series of issues and concerns, due to fact that computers will have access to a much higher amount of personal information than they have at this stage. This research paper focuses on the ethical issues of Affective Computing devices, and attempts to predict what the positive and negative consequences due to the implementation of such systems might be.

2. Background

2.1 Introduction to Affective Computing

Affective computing is the field of creating machines that can recognize affective states and respond to humans using emotions, with the purpose of improving the *quality* of human-computer interaction (Picard, 1997). Machines, computers and other devices should ideally *get to know* their users, in order to better serve individual's needs and lead to a more natural and effective use of technology. The aim of affectively-aware systems is to use human-human interaction as a basis for improving human-computer communication, which is, at this stage, "affect-blind, affect-deaf, and generally speaking, affect-impaired" (Picard, 1997, p. 15). Affective Computing does not simply attempt to create algorithms that adapt computer responses to user preferences and interests. The ultimate goal is certainly to create devices that really *know* a person, but the primary aim is to enable computers to independently learn the information needs of their user. As stated by Picard (1997, p. 102), "It is not only more natural to communicate preferences with affect, but it is efficient; affect can be communicated through modulation of the voice, a smile, or nod of the head, without conscious effort or mental work". The core ideas of Affective Computing are based on a paper written by Picard (1997), who first suggested integrating the detection and interpretation of emotions into computer science. Picard's ideas were initially received with scepticism and distrust and received little support. This was probably due to a *presumed* absence of connections between emotions and computers. As stated by Picard in her book *Affective Computing* (1997), humans are used to consider emotions as "a disorganized response, largely visceral, resulting from the lack of an effective adjustment. Acting "emotionally" implies acting irrationally, with poor judgement" (p. 1). In fact, a large body of research suggests that emotions interfere with rational decision making by affecting people's thinking and motivation. Also, computers are rational, precise, and free of errors and misunderstandings. Emotions might be essential in social communications, romantic and friendly relationships, but the idea of implementing affect in the area of employment, decision making or marketing, could certainly lead to undesirable consequences. In her book, however, Picard demonstrates how these assumptions are unfounded, and at this stage, Affective Computing is one of the most important research topics in terms of human-machine communication. Its main objective is to improve this interaction, by empowering computers to perform calculations related to emotions, thereby enabling technical systems to respond as living beings. The attempt to create machines with human characteristics is mainly justified by studies that have shown that technical systems with human forms are more likely to be accepted as communication partners by humans than others (Picard, 1997). Affective Computing has the potential to transform several areas of society. It could offer concrete and relevant solutions to problems in many fields, both in business and in private sectors. Affect could also be implemented into the health/safety sector, by allowing vehicles to detect signs of stress or anger to prevent accidents. Emotional analysis can be integrated into training modules and improve the interaction between service employees and customers. Combining technology with emotions could especially bring considerable improvements to e-learning. In fact, an intelligent system might be able to detect a student's emotional state and respond appropriately. "Given affect recognition, the computer tutor might gauge if it is maintaining your interest during the lesson, before you quit out of frustration and it is too late to try something different" (Picard, 1997, p. 16). Tao and Tan (2005) have defined three areas of development in Affective Computing research: Affective Understanding, Affective Generation, and Application. Affective Understanding refers to the capacity for a device to understand the affective state of its user. Affective generation is the ability of a computer to react to a user's affective state, and affective application refers to the question about which fields and tasks can be improved through Affective Computing (Schwark, 2015).

2.2 Emotions

The first step to *understanding* an emotion, is certainly defining that emotion.

Theorists and psychologists propose varying interpretations of the concept of emotions. However, the dozens of emotion theories can fundamentally be divided into two categories, according to which emotions are either cognitive (namely related to the brain) or physical (related to the body) (Picard, 1997). William James (1884) first defined emotions as "the feeling" of the changes of the autonomic nervous system taking place on a visceral level after a stimulation. He argued that, while "bodily disturbances" are conventionally considered products or expressions of the so-called standard emotions (anger, fear, lust, etc.) these corporeal reverberations are actually the raw material of the emotion itself. With his theory, James highlighted the importance of physiological arousal and biological response to defining emotion as a mental process. William's theory was challenged by Cannon (1927) and later by Schachter and Singer (1962), who emphasized that the brain is strictly connected to affective states. The wealth of different and contrasting theories has confirmed that both the brain and the body play an important role in the generation of emotions (Picard, 1997).

The next, and most important, step for researchers, and for computer scientists in the field of Affective Computing is to assign names to emotions and divide them into categories.

Human affects can be classified into six typical emotions (Elfenbein & Ambady, 2003): happiness, sadness, surprise, fear, disgust, and anger. Other authors have been focused on the dimensions of emotion, such as negative or positive emotions. The three emotion dimensions can be categorised into "arousal" (calm/excited), "valence" (negative/positive) and "control" or "attention", which refers to the internal or external source of the emotion (Picard, 1997). Even though it is relevant for Affective Computing researchers to understand the theory of emotions, their main interest is to analyse *whether* and *how* it is actually possible to distinguish one emotion from another. As stated by Picard, "the distinction- are emotions physical first or cognitive first- is not as important to us as the question, 'How can emotions be generated in computers, recognised by computers, and expressed by computers?'" (Picard, 1997, p. 23).

As already mentioned, Affective Computing hold that an efficient human-computer interaction would ideally simulate the human-human interaction. For this purpose, researchers have analysed human behaviours and the methods individuals use to *express* their emotions and *understand* others emotional states. Human sciences such as psychology, anthropology and sociology have investigated nonverbal communication for several decades, demonstrating that signals such as facial expression, speech perception, gestures and body movements, as well as their positioning in an environment, are used by individuals to express their emotional states (Cristiani *et al.*, 2010). These cues have been grouped into five classes called codes, namely physical appearance, vocal behaviour, facial and eye behaviour, gestures and body movements, space and environment (Guerrero *et al*, 1999). Computer scientists in the field of Affective Computing have focused on vocal behaviour, facial expressions, gestures and body movement. Researchers have realised, that a further effective method to detect humans' emotion is to measure physiological data. No major efforts have been done to investigate the space and environment code, however, the emerging field of Virtual Reality (VR) investigates the way individuals use the space and interact with the environment. This could be the key element of how virtual systems can change the nature of verbal and nonverbal interaction.

3. Emotion detection methods

3.1 Facial expressions

"Facial expression is one of the most cogent, naturally preeminent means for human beings to communicate emotions, to clarify and stress what is said, to signal comprehension, disagreement, and intentions, in brief, to regulate interactions with the environment and other persons in the vicinity" (Pantic, 2014, p. 1). By just observing the changes in a person's face, we are usually able to understand his or her mood and, therefore, affective state. Because Affective Computing aims to replicate human-human interaction, the analysis of facial expressions should apparently be the most efficient way to detect emotions for computers as well. However, there are some issues related to facial expressions that can complicate the process of recognising affective states for machines. First of all, as stated by de Gelder *et al.* (1997), the way human beings interpret emotions might be rooted in "our experience of facial expressions built up over the course of communicative development" (Schwark, 2015, p. 762). Even though within the framework of Affective Computing, it is not that relevant to know whether facial expressions are *innate* or result of an experience, this uncertainty could complicate the process of emotion recognition. If the interpretation of human affect is based on experience and has a cultural basis, then the meaning of certain facial expressions might depend on the geographical area, and this makes it difficult for Affective Computing researchers to create a universal facial expression recognition system that can be applied to all individuals. For convenience, the existing technologies are obliged to use highly stereotyped personality types and emotional responses, which means that they do not replicate the human behaviour with 100% accuracy (Tao & Tan, 2005). Regardless of whether emotions are biological or a cultural/experiential in nature, research studies, including those conducted by Duchenne (1862) and deepened by Paul Ekman (1993), have shown that there is a close connection between emotion and facial expression. Both in human-human and in human-computer interaction, another issue regarding facial expressions is the fact that they can be confused and misinterpreted. This problem can easily be solved when facial expressions are combined with words and body movements, but in some cases, this is not possible. Labarre (1947) takes the example of colic babies who "smile" in their sleep, even though that facial expression aims to express pain rather than pleasure. People smile when they are nervous, they cry when they are happy and, sometimes, they laugh when they are angry, and people often fake their emotions. Fake facial expression might not be a big issue in AC at this stage because the need to fake emotions usually arises in social contexts. In the "face-to-face" communication between a user and a computer, there is no apparent need for an individual to perform an emotion. However, if Affective Computing succeeds in its intent, then affect recognition systems will be applied in everyday life, therefore, social contexts as well. In that case, the issue of fake emotions might have to be reconsidered.

3.1.1 Process: how to detect emotions through facial expressions

The aim of Affective Computing is to teach computers how to use images of users' facial expressions to recognise his or her emotional states. This means that facial expressions need to be translated into something that computers can understand (Calder, 2001).

The majority of face recognition software is based on a publication by Tuevo Kohonen (1989), who introduced the possibility of representing a face in a two-dimensional space, creating a unique face print in a format that most of the machines were able to understand. The Principal Component Analysis (PCA) algorithm is capable of isolating and representing the typical components of a set of data, and it is mainly used in patter recognition. PCA is used to define which are the features that characterize a face, maximising the differences between the faces. However, the side effect of this technique is that the algorithm also maximises the defects of the image, such as the low light conditions. To solve this problem, Peter N. Belhumeur *et al.* (1997) generated a new algorithm called LDA (Latent Dirichlet Allocation) that introduced the possibility of using a more accurate model of the face. More recent approaches use 3D-images for facial recognition. They determine an expression by analysing the movement and location of certain features (Gupta O. & Gupta P., 2015).

The Facial Action Coding (FAC) system developed in 1978 by Paul Ekman and his colleagues categorised the given expression into six basic emotions: disgust, sadness, anger, happiness, surprise and fear. This approach considers expressions to result from little deformations of the neutral face due to facial muscles called Action units (AUs) (Ekman & Friesen, 1971). In the 1990s, Facial Animation Parameters (FAP) were developed and incorporated in the MPEG-4 standard to develop human facial expression models. A face contains around 84 tracking points, but among them, only 11 are used to recognize emotions. These 11 tracking points include inner brow raise, nose wrinkle, and lip corner depression (Gupta O. & Gupta P., 2015). Viola and Jones (2004) introduced a new image representation method called *Integral Image*, through which features are computed very quickly, as well as a simple classifier that was build using the AdaBoost learning algorithm (Freund & Schapire, 1997) to select a small number of critical visual features from a very large set of potential features. This method also allows background regions of the images to be to be quickly discarded so most computation can be conducted on promising face-like regions. Despite the variety of face detection methodologies, the process of facial expression recognition can be analysed in three stages: *face detection*, *facial feature extraction* and *facial expression interpretation*. In Affective Computing, the final stage is replaced with the *Classification Process* (Pantic, 2014). *Face detection* is computer technology that involves the tracking of the face through specific biometric techniques that use (and combine) algorithms, to determine the position and the dimensions of a face within an image or a video. By using appropriate landmarks, it is then possible to extract some characteristics of the image (or video) for subsequently analysis *(facial feature extraction). Facial expression interpretation* is based on a comparison of the information extrapolated during the *face detection*, and *facial feature extraction* and is elaborated according to the algorithms that determine small changes in the expression. This stage determines the actual expression on the face and classifies it as one of the six basic emotions.

Once the computer has recognised an emotion, the next step is to analyse whether its answer is correct, but how is it possible to know whether the computer is right?

Picard (1997) presents a test that a computer should pass if it is able to accurately recognise human emotion through facial expression. In this test, a group of computers and a group of humans are asked to reveal which kind of emotion a person in a digital video is expressing. Given that 70% of the people recognise anger and 30% recognise hatred, if a single computer identified that person as expressing hatred, then it could not be penalised. The computers will succeed as long as 70% of them recognise anger and 30% recognise hatred. This means that the same "margin of error" must be granted both to humans and to machines because perfect performance cannot be guaranteed in either cases. As stated by Picard (1997, p. 51), "for a computer to imitate human recognition ability, we should be able to momentarily swap a computer for a human, and the computer should recognize the same emotions that the human would recognize". A practical example of a facial recognition system is FaceSense, the API of which is able to recognise an individual's affective state in real-time by analysing a video of his or her face. The API presents a framework that combines the processing of the face (e.g., a head nod or smile) with predictions of mental state models (e.g., interest and confusion) to interpret the affective state.

3.1.2 Implementation of facial expression recognition systems

The implementation of systems that detect individuals' emotions by analysing their facial expressions could improve several fields, including medicine, education and cognitive science. In computer science, such technology would ideally provide an easy, natural communication between machines and humans, even reaching the point at which it can replace the classic keyboard and mouse (Pantic, 2014).

Outside the field of Affective Computing, facial recognition systems are becoming increasingly popular because they offer a significant help in various fields, such as terrorism and crime fighting, and user authentication systems for improved security. Even though the main objective of these kinds of systems is to identify an individual for security purposes rather than to recognise his or her emotions, the existing facial recognition technologies present some issues (both technical and ethical) that could theoretically underpin the problems of emotion detection in Affective Computing. Apple's new FaceID facial recognition technology has been implemented into the iPhone X for unlocking apps and using ApplePay. FaceID is the successor to the iPhone TouchID fingerprint scanner, and it is supposed to be 20 times more secure than the latter. The Russian FidFace website allows users to upload the picture of a person to link them to the person's profile on a social network called VK. In Shenzhen, jaywalkers are identified using a surveillance system. Photographs of the violators, their names and social identification numbers, are then displayed on LED screens installed at road junctions. The Chinese equivalent of Amazon, Alibaba, enables people to “pay with a smile” using facial recognition in stores.

3.1.3 Technical limitations of facial expressions recognition systems

Technical issues related to these techniques include camera performance, lighting, and the presence of makeup, glasses and more. Sometimes a variation of these conditions can affect the success of the systems. The accuracy of this process depends upon factors such as the exact methodology applied, the number of available training images, the quality of the photos, and the visibility of the individuals within those photos. However, while early face detection systems were barely able to recognize a single face from a frontal view, current technologies are able to identify individuals from various angles and distinguish faces from cluttered backgrounds (Welinder, 2012).

The experiments conducted so far in Affective Computing aimed to *understand* and identify whether computers can detect specific emotions. This means that the participants involved in the experiments were aware of the main goal and behaved accordingly. However, the next step for Affective Computing would be to apply these systems in several different situations, and ideally to the everyday life, in which affective state recognition should be a support rather than a goal. The user should ideally be able to concentrate on other tasks rather than constantly ensuring they are sitting in the correct position or worrying about technical issues.

3.1.4 Ethical issues of facial expressions recognition systems

During a panel discussion let in 2009 at the Information Systems Educators Conference in Washington DC, Lee Jonathan Steen asked participants about their feelings and opinions about several affective devices (Steen & Kim, 2010). One of the proposed devices, based on facial expression recognition, was a *wearable computer*, namely "computational devices that are *worn* as an article of clothing or jewellery" (Picard, 1997, p. 227). The device in question was a camera implemented on eye glasses and connected to a computer. The camera records a video of what and whom the person is facing and sends the images to the computer, which tries to determine the affective states of the people in the image. Steen discusses the positive effects that such a device (called Detection Expression, DE) could have, such as helping people who have expressive communication disorders, or parents to understand the needs and emotional states of their children. However, the use of DE could be expanded to other areas. Steen and Kim (2010, p. 580) ask: "But what about a car salesman who uses the device to help him sell a car to customers who are unaware DE is in use?".

Steen describes a scenario in which a salesman can instantly and easily recognise a buyer's feeling about a car, using a DE to detect even the smallest changes in the client's facial expression and, therefore, his or her appreciation or disapproval of a colour, a car model and son on. After presenting different facets of the same scenario to the participants, and asking for their opinions, Steen concluded that, even though the majority was a little *scared* of the presence of a machine, participants would accept such a device on the condition that they are told about its use. This is, as stated by Steen and Kim (2010), what happens with telephone conversations. In fact, people would feel violated if they found out that a call has been recorded without their consent, even though a majority of people readily allow customer service callers to record their conversations. At this point, it is necessary to understand whether and how it is possible to ask people for permission to use devices such as DE devices. A second issue concerning wearable computers is the use of personal data. Apple's FaceID system has given rise to several questions about privacy protection. However, the company has ensured that the FaceID images are not sent to Apple, nor included in device backups. They are only available to the Secure Enclave (Apple, 2017). If we consider Steen's DE device, the questions would be the same: where will the data be stored? Who will have access to that information? The video of the clients and the emotion detection "results" connected to the images could be used by the sellers to conduct analyses about which details of cars people focus their attention, which words are best to use in front of the client, and so on. Such a device could be used in other scenarios. For instance, webpages could activate webcams to detect the individuals' emotional responses to the website content, or to examine which part of the page captures their attention.

The above examples are just assumptions about how emotion detection based on facial expression could be used in the near future, and so are the privacy issues related to it. Because Affective Computing devices are momentarily used for limited purposes only, it is not necessary at this stage to deeply analyse the consequences of the above scenarios. However, the examples assume that, besides the technical issues, privacy might be one of the barriers that limits the social acceptance of Affective Computing devices.

3.2 Emotional speech perception

Emotional information from vocal utterances and facial expressions often interacts:

In fact, "facial and/or vocal emotional cues that are not consciously attended to, still influence the processing of facial and/or vocal stimuli that are attended" (Jaywant & Pell, 2011, p. 1).

Individuals tend to respond in a more accurate way if the facial expression of someone speaking is coherent with the person's words. As stated by Picard, voice is the reason why, in some cases, we prefer talking on the phone instead of writing emails or sending letters. "Spoken communication transcends the message of the words-altering the listener to states such as anxiety, nervousness, or love" (Picard, 1997, p. 27). Also, sometimes the intonation can contribute greater meaning than the semantic content of speech itself. Picard (1997) makes the example of a dog, who will get off the couch regardless of whether the owner yells: "get down of the sofa!" or "get up on the sofa!". While recent technologies have been focussing on *what* is said and *who* is speaking, the aim of Affective Computing is to analyse *how* something is told (Picard, 1997). For this purpose, computer scientists had to transform speech into something that machines are able to understand.

Petrushin's (2000) emotion recognition system was able to analyse telephone speech signals and distinguish between states of agitation and calmness with 77% accuracy. However, emotion detection through vocal behaviour presents some limits that are similar to those analysed in the *facial expressions* section. An important parameter in speech is the pitch level, namely "the average level of pitch extracted that forms the prosodic contour" (Rodero, 2011, p. 1). High pitch levels are characteristics of joy, anxiety, or fear, and they usually capture the listener's attention by establishing a contrast with neutral attitudes, marked by medium pitch levels. Instead, low pitch levels are related to sober emotions, such as sadness. According to Scherer (1986) researchers noted that the pitch levels associated with anger and happiness presented very similar characteristics, which makes attempts to distinguish between the two more difficult. Pitch levels are not sufficient to make a detailed analysis of a person's affective state. Mozziconacci (2002) states that combining pitch and contour type yields better results. Contour type refers to a function or curve that keeps tracks of pitch levels over time. Emotions such as joy, anxiety, and fear are usually characterised by greater inflexions, while fewer variations are typical of sober emotions (Rodero, 2011). Another issue related to vocal emotion recognition is the fact that the communication between humans and computers is generally spontaneous, which makes it difficult to distinguish the acoustic features. Schwark (2015) also highlights the limits of traditional computer systems and software, that are designed for use with the mouse and keyboard as input, instead of speech. Speakers tend to suppress their emotional behaviour when they are aware they are being recorded (McIntyre & Göcke, 2006). For this reason, experiments require actors that can portray emotions. The problem with this approach is that, as stated by McIntyre and Göcke, "acted speech elicits how emotions should be portrayed, not necessarily how they are portrayed" (McIntyre & Göcke, 2006, p. 2). The issues of emotional speech recognition systems are, therefore, the "fake" nature of acted performances on the one hand, and the suppression of emotional behaviour in individuals who know that their words were being recorded, on the other hand.

3.2.1 Process: how to detect emotion through speech

The process of voice affect recognition is very similar to the process used for facial expressions recognition, and it can be analysed in three stages: *voice detection*, *voice feature extraction*, and *voice emotion classification*. The following description is based on a paper (Bahreini *et al.*, 2015) that presents the FILTWAM (framework for improving learning through webcams and microphones), a framework that is used in e-learning that aims to recognise students' affective state to offer a proper feedback. The system also uses facial expression as an input to examine the emotion, but in the following analysis, I will only consider the voice recognition process.

Voice detection is conducted to recognise the emotions of a voice by dividing the input speech into segments. After the voice has been detected, the *voice feature extraction* component extracts a sufficient set of feature points of the learner, which are then put into vectors. By analysing the feature sequences in each vector, it is finally possible to recognise the changes between the vectors. The last stage is based on the already mentioned emotion classification.

3.2.2 Implementation of voice recognition systems

Voice recognition technologies are already popular outside the field of Affective Computing. A common example is Siri, Apple's intelligent personal assistant that uses voice requests to answer questions, make suggestions, and perform actions related to the requests. Alexa is a virtual assistant developed by Amazon that is able to set alarms, play music and audiobooks as well as provide information about the weather, the traffic, and so on. The mentioned applications are based on systems that take an input (the voice) and act (output) according to *what* is said. This process is generally called *speech recognition*, which should not be confused with *speaker recognition* systems, which are usually employed to provide access to secure systems, by extracting some features of the voice and comparing it to other voices to identify the best match.

3.2.3 Technical issues of voice recognition systems

At this stage, the database of emotional speech recognition comprises naturally occurring data sources, such as call centre recordings and news readings, which present some complications. Technical issues refer to background noise and overlapping utterances, which cannot easily be extract and separated from the voice. However, the progresses in signal processing, algorithms, and hardware achieved in the past decade, have enabled further improvements in speech recognition systems (Furui, 2000). A relevant theme in speech recognition is speaker independence. In fact, systems should ideally be able to function correctly and be applicable regardless of who is speaking to avoid a long and time-consuming training period for each new speaker. However, if on the one hand speaker-independent systems would offer the fastest service, then they would also be less precise and accurate (McIntyre & Göcke, 2006).

3.2.4 Ethical issues of voice recognition systems

Referring to Benjamin Cardozo's famous comment "Law never is, but is always about to be" (Cardozo, 1921), Lang and Benessere (2017, p. 20) state that "Yet, his law perfectly sums up the new frontier in law as more and more people integrate speech recognition technology into their everyday lives." When a person interacts with Alexa, the audio is streamed to the cloud. Amazon's Terms of Use for the Echo system states that "Alexa processes and retains your Alexa interactions, such as your voice inputs, music playlists, and your Alexa to-do and shopping lists, in the cloud to provide and improve our services" (Amazon, 2018).

The integration of a virtual assistants in the everyday life is helpful for the majority of the people, since it might simplify and accelerate some tasks. However, Lang and Benessere (2017) analysed the case of James Bates, who was investigated for the murder of his friend Victor Collis. When the prosecution asked Amazon to provide recordings from the suspect's Amazon Echo, the company refused because of privacy concerns. Bates voluntarily provided the recordings, and the case highlighted the fact that users have access to their recordings and, therefore, can willingly disclose them (Lang & Benessere, 2017). Virtual assistants can also set up timers and alarms and, thereby, provide information regarding a user’s life style, her appointments, working hours and so on.

“This data, like other documentary evidence, is likely to be more accurate and informative than deposition testimony, which was taken only after a preparation session with an attorney and relies on a person’s memories of events that may have occurred years before the deposition” (Lang & Benessere, 2017, p. 22). Lang and Benessere state that a possible, temporary self-help solution is to disconnect one’s assistant when one wants to avoid being recorded. In the context of Affective Computing, the questions and issues about privacy are basically the same as those concerning virtual assistants. Indeed, the potential privacy issues are significantly higher. Even though a part of the population uses virtual assistants on several occasions, such as to turn the phone light on, ask for weather conditions, or to call a friend, the use of these systems is generally limited to specific purposes. If (and when) Affective Computing succeeds in creating machines that use speech as an input, the possibility of turning the microphone off would probably be excluded. Otherwise, the computer might not be able to perform in the best way. In the situation in which computers react to a person’s voice (and therefore to his or her affective state), speech would not only be helpful, but necessary. Referring to Affective Computing, Picard (1997) states that such systems would lead to a more natural and efficient communication between humans and computers. In the context of emotion detection from speech, “natural communication” might mean that the use of words in no longer limited to specific purposes, but used in the same way that individuals would communicate with each other. Personal opinions and personal stories might therefore be included in the information that users provide to the machine.

3.3 Gestures and body movements

"The ability to work out what is really happening with a person is simple - not easy, but simple. It is about matching what you see and hear in the environment in which it all happens and drawing probable conclusions. Most people, however, only see the things they think they are seeing" (Pease, 2004, p. 20). The development of the communication system has allowed speech to become the main means of expressing emotions and intentions. However, before spoken language, individuals were accustomed (and forced) to understand others' affective states and intents by observing their behaviour, and therefore their body movement. Along with facial expressions, gestures and body movements "are considered to reflect non-symbolic and non-propositional emotion communication" (Flaisch *et al.*, 2011, p. 109). So-called "body language" is the factor that establishes an evidence about or a contrast between the words pronounced by an individual, and the message that a listener receives. They include the movements of hands, the head and other parts of the body that enable individuals to communicate a variety of feelings, thoughts and emotions. Emotional states can be communicated through the modulation of a single movement or through a combination of movements. For instance, anger can be expressed by increasing one's walking speed, or by making a fist, and therefore through a specific movement (Karg *et al.*, 2013).

As evidenced by Ray Birdwhistell's experiments, the message of a sentence is only affected 35% by the words, and the remaining 65% by the nonverbal signs (Brow, 2005). This means that individuals tend to interpret the words of a sentence according to the meaning that the speaker attributes to it with his or her body language. Allan and Barbara Pease (2004) agree that there are three basic types of gestures: intrinsic (e.g., nodding to expresses consent), extrinsic (e.g., turning to the side to refuse something), or a result of natural selection (e.g., the expansion of the nostrils to oxygenate the body for instance).

Even though gestures are commonly culture-dependent, individual's exposure to mass media has led to a general globalization of gestures. In *The definitive book of body language*, Allan and Barbara Pease (2004) take the example of the thumb-up symbol, which is supposed to have different meanings in different cultures. They state that "This is 'good' to Westerners, 'one' to Italians, 'five' to Japanese and 'up yours' to the Greeks" (p. 25). However, nowadays, the thumb-up is widely used as a sign of agreement, consent or interest.

3.3.1 Process: how to detect emotion through body language

Within Affective Computing, the detection of body movements has not received as much attention as facial and vocal expressions. Although a certain number of researchers have developed methods for detecting and coding body movements, there are still no precise guidelines for understanding and categorizing human behaviours (Harrigan, 2005).
Examples of existing methods include the coding system developed by Mehrabian (1972) that aims to study interpersonal attitudes by limiting the analysis to body orientation and body posture. Tracy and Robins (2007) have developed a coding scheme for analysing the nonverbal expression of pride in static upper body posture. In a paper published in 2012, Dael *et al.* proposed the BAP coding system, which is based on a main distinction between body posture units (general alignment of articulations) and body action units, namely "excursion of one or a set of articulators (mostly the arms) outside a resting configuration" (p. 101), providing evidence of its coding reliability of occurrence, precision, and segmentation. Known body movement detection techniques include motion capture, which, however, requires very controlled environments and specific cameras able to track human bodies or objects, while video based automatic tracking techniques that are already used in the area of Affective Computing might be a more efficient way of detecting gross body

movement, but are inappropriate for identifying the details that are needed (Dael *et al.*, 2012). Despite the limited number and efficiency of methods, a traditional body movement recognition process can be analysed in three stages. The first step is determining a suitable model for input (human detection). This usually happens by detecting a whole body and removing the background of every frame that involves a human gesture. The main issues concerning this first step are the non-rigid, dynamic nature of the human body and the potential technical problems due to possible changes in the environment, such as the illumination changes. The procedure is similar to the one involved in the face detection process, and consists of extracting some regions and classifying these regions as human or non-human. Once human regions have been identified, the pose needs to be tracked to be able to classify the result in the final stage (Noroozi *et al.*, 2018).

3.3.2 Implementation of gestures and body movement recognition systems

EmoteMail has been developed to express some aspects of the writer's affective state while writing an email. It is able to detect the writer's facial expression and typing speed, and introduces them as design elements, in order to allow the recipient to analyse the tone of the email (Reynolds et al., 2004). Researchers at Samsung have developed a smart phone that can detect people's emotions by analysing how the user uses the mobile phone, in terms of typing speed, times when a specific button is pressed and so on. The system gets trained to know the user to be able to recognise whether the user is happy, sad, surprised, fearful, angry, or disgusted. Samsung's main purpose is to implement this functionality in social media, such as Twitter, and enable users to have information about the affective state of those who tweet something (Anthony, 2012). To analyse transformations in human behaviour, such as the increasing walking speed in relation to anger for instance, it might be necessary to consider the possibility of implementing webcams in new types of environments.

3.3.3 Technical issues of body movement recognition systems

The existing body-detection and body-tracking systems are generally addressed to other purposes and contexts. For instance, the mentioned motion tracking technique has attracted the attention of professionals in the healthcare field, in which there are attempts to understand the postures and movements of the human body for medical purposes. However, motion capture systems allow their users to create realistic character motions by recording the movements of objects and people with the aim of animating digital characters with highly realistic and complex human movements. Affective Computing systems that are able to detect a human body would work through the use of webcams. Because the possibility of redesigning computers is unlikely, it might be problematic for the current devices to detect a whole human body. For this reason, and due to the fact that people do not tend to exaggerate body movements while sitting in front of a computer, at this stage the detection of movement for analysing a person's affective state would probably be limited to minor and specific movements only.

3.3.4 Ethical issues of body movement recognition systems

Currently, the devices used for detecting and tracking individuals are surveillance cameras, which have become increasingly popular during the last decade, both in public and in private sectors, especially thanks to the decreasing cost of the hardware (Cavallaro, 2007).

The need to protect individuals and properties from harm has encouraged developed countries to install surveillance cameras in offices, shopping centres, streets, train and bus stations and so on. As stated by Bullington (2005), the implementation of affective systems in these types of scenarios would require cameras with massive amounts of storage capacity, as well as software that is able to process real-time data. Also, the objectives of systems that recognise individuals' emotional states have never included surveillance so far. However, assuming that the technical difficulties will be overcome, implementing such systems in crowded spaces would allow the collection of important information about human behaviour. Beyond the operation of matching people's faces to images stored in a database, surveillance devices may be able to identify an individual's affective state and, therefore, his or her behaviour and intentions. While facial expressions and speech inputs could be used to improve several fields, such as e-learning, the emotion detection through body movement might be used for other purposes. In fact, if we consider the scenario in which a system is detecting the affective state of people walking in a shopping centre, individuals themselves would not directly benefit from the situation. Again, this could be used for marketing purposes. Surveillance cameras might be used to measure people's interest in objects that are displayed in store front windows and eventually conduct market analysis to identify which target costumers is most likely to buy something. If affective state analysis through body movement detection enables the identification of individuals' intentions, then it might be possible to differentiate between people who walk around shopping centres out of boredom, and those who are willing and determined to buy something.

However, Bullington (2005, p. 97) states that "Our behaviour and outward expressions of psychological states are a function of social and cultural context, as well as of individual differences in the way we respond to these contexts." In fact, besides the privacy issues that the implementation of such systems would cause, Bullington identifies the problem of the different habits and expressions of individuals. This means that a system could misinterpret a person's behaviour and, therefore, his or her affective state and intentions. According to the Bullington, the problem can be solved if the systems gets trained to learn the behaviour of individual people in a limited context, which, in turn, would give rise to further issues.

"Will human observers, along with the problems that they introduce, be necessary to oversee the monitoring process for a certain amount of time until the system is trained? How will such system 'training' be undertaken, and for how long? Are the self-reports of the person whose state is being observed to be trusted?" (Bullington, 2005, p. 98).

3.4 Physiological measurements

Affective states are generated by physiological changes that are regulated by the autonomous nervous system. Beyond the already analysed visible transformations, such as facial expressions, changes in speech, gestures and body movements, there are some changes of values that cannot usually be noticed or observed by humans, such as blood pressure, heart rate, and electrodermal activity.

Even though emotion recognition through facial expression or voice analysis generally achieves positive results, the problem of these methods is, as already mentioned, that they are usually tested on actors, and the results are often "performed expressions". Instead, physiological reactions cannot be controlled nor performed. For this reason, "they provide undisguised information on the user's current affective state. Having reliable access to physiological data would hence provide for an information source that might even prove to be more trustworthy than those relied on today" (Peter *et al.*, 2008, p. 294).

Also, as stated by Reynolds and his colleagues (2004), while facial expression, voice and body movement recognition systems do not grant users primary control (because it is not easy to disable these systems), the latest physiological measurement devices can be worn and, therefore, activated and deactivated at any time.

3.4.1 Process: how to detect emotion through physiological measurements

Among the methods used to induce a specific mood, one of the most popular is the Velten Procedure (Velten, 1968), which consist of giving a set of statements to participants to arouse a certain affective state. However, the main limit of this techniques is that participants are usually induced to an emotion for no more than approximately ten minutes. To overcome this issue and evoke enduring moods, researchers sometimes use film clips and sounds in their experiments (Zimmermann *et al.*, 2003). Zimmermann and his colleagues used various sensors to measure and analysed respiration, pulse, skin conductance level, and the corrugator activity of ninety-six students who were watching six different kinds of film clips and after completing the task of shopping on an e-commerce website. Ark, Dryer and Lu (1999) detected users' affective states through a mouse, Marrin and Picard's (1998) "Conductor's Jacket" was able to collect both physiological and motion data from musicians with the aim of analysing their affective states during performances, Scheirer and his colleagues (1999) developed glasses that can recognise moods such as confusion or interest. While facial expressions, speech, and body movements first have to be recognised (separated from close objects or from background noise), an "emotional mouse" or an "emotional keyboard" can directly read physiological information. This usually happens after electrodes, are applied to specific body parts to measure skin temperature, galvanic skin response and so on. The physiological data is then sent to a server, that analyses the data using algorithms to identify an emotion (Whang & Lim, 2008).

3.4.2 Implementation of physiological measurement systems

It is fairly common to use physiological measures for marketing purposes. The term neuro-marketing refers to the field that combines psychology, economics, and neuroscience, with the aim of analysing how potential consumers react to changes in advertising, packaging-design and so on (Bakardjieva & Kimmel, 2017). Other than that, the existing commercial systems that are used to analyse affective states through physiological measurements have usually been developed for recreational or medical purposes, and are, therefore, designed to work in specialised laboratories rather than to be applied in everyday life. However, physiological sensing devices will be slowly incorporated into devices that people are naturally in contact with. The Affectiva Q Sensor has been developed to identify and record stress levels by measuring moisture and electrodermal skin activity. The main purpose of the developer is to help parents and doctors to understand and monitor individuals with cognitive disabilities, such as autistic children. The unobtrusive device works by sending out a signal to parent or caregiver every time the sensor detects stress, excitement, or anxiety, and it also keeps track of repetitive movements. Even though this kind of devices may sound futuristic, physiological measurement technology has a quite long history. The polygraph, commonly called as the "lie detector", was developed in 1921 in Berkley, California. The polygraph machine was generally used as an interrogation tool with criminal suspects because it was able to identify physiological changes associated with increased stress in a person while answering questions. The term "lie detector" refers to the machine's presumed capacity to distinguish between the truth and a lie, and the device was, therefore, used to distinguish deceptive answers from non-deceptive answers. Regardless of the fact that there are many doubts about whether the polygraph is a reliable device or not, its functioning underlies the modern physiological measurement devices (Steen & Kim, 2010).

3.4.3 Technical limitations of physiological measurement systems

The main issues concerning physiological measurements are the expensive and invasive equipment needed, on the one side, and the difficulty of precisely analysing which external or internal factors actually influence the affective state, on the other side. For example, an increased heart rate might have several causes. In *Physiological Sensing for Affective Computing* (2009), Peter and his colleagues produced a list of features that a functional sensor system in the field of Affective Computing should have. First of all, the device cannot be susceptible to movements. The user should focus on his or her tasks rather than on the cable and wires. Otherwise, "users may find that the benefits from affective computing might not outweigh the costs of being hooked up to a wall of equipment while using the system" (Shwark, 2015, p. 763). At the same time, the cables and wires should not be obstructive for the user. Ideally, the communication between the sensor and the processing host should be wireless. Other features include easy usage, immediate access, ease of integration, and standard conformance.

3.4.4 Ethical Issues of physiological measurement systems

As mentioned above, physiological measurement might appear to be highly "secure" thanks to the possibility of wearing or taking off the devices. However, while facial expressions, vocal behaviour and body movements would probably be used to get an instant reaction from the computer, the future of physiological measurement devices might be to keep track of the emotional changes in individuals. One of the reasons behind this assumption is that, as already mentioned, physiological reactions cannot be performed nor faked, which means that they could represent a faithful picture of a person's inner world. While talking about privacy, people might prefer to have physiological measurements recorded and stored somewhere, instead of knowing that their face or conversations can be seen by others. Again, the main questions are: where will the data be stored? Who will have access to this information? And is it possible to filter the information? If a person that suffers from stress voluntarily asks professionals to monitor his or her stress level, then will the professionals also be able to monitor other affective states, such as happiness, sadness, or boredom?

4. Discussion

This analysis has examined the four methods used in Affective Computing to detect emotions. Given the technical and ethical issues, there is one method that might be successful in improving the interaction between humans and computers more than the other. As already mentioned, the main limit of facial expressions recognition systems is the fact that technologies are forced to use highly stereotyped personality types and emotional responses. Even though in some experiments computers were able to detect an affective state with a high accuracy, individuals tend to express their emotions in different ways. The probability that a machine will misunderstand a facial expression is therefore high. Also, actors and participants in the experiments are usually asked to exaggerate facial expressions, while it is uncommon for users to blatantly smile or cry while using a computer. The fact that this method requires the use of a webcam also calls into question the acceptance of individuals. For these reasons, emotion detection through facial expressions might be useful to identify interest or boredom, and therefore improve fields such as e-learning, in which a user is focused on a task and can be sure that his or her face is detected for that specific purpose only, but its utility in other fields might be limited. Using voice recognition systems to improve the human-computer interaction means redesigning computers in order to use speech as the main input. Excluding this possibility, a system that detects emotions though vocal behaviour might be needed when dictating an email to a computer, in order to communicate the "intentions" behind the words to the recipient. However, as already mentioned, the ones involved in Affective Computing experiments are asked to exaggerate affective states, while individuals commonly tend to communicate their emotions in a spontaneous way to computers, which makes it difficult for systems to distinguish acoustic features. A solution is to train the computer to *learn* how an individual person expresses his or her emotions, but this requires long training periods, and the main objective of voice recognition systems in Affective Computing is speaker independence.

Again, devices that recognise an individual's affective state through speech might be useful in e-learning, and in those scenarios in which the system only needs to detect a limited number of emotions, such as boredom and interest. The limits of voice recognition systems are also dictated by the fact that computers and smartphones are used in social contexts as well, and it is hard to image a scenario in which all users talk to a computer instead of using a mouse and keyboard. Affective states might be recognised in voice messages and communicated to the recipient of the message. If emotion detection through vocal behaviour is applied in situations in which users already use speech as an input, there is no need to redesign computers, and people are more likely to accept such devices without concerning about privacy issues. Therefore, voice recognition systems might not revolutionise the human-computer interaction, but they can improve some aspects of it.

In the analysis, I have mentioned the possibility of implementing body movement detection systems in surveillance cameras, but this fact would not directly improve the communication between individuals and machines in the sense that we understand human-computer interaction in this analysis. Instead, implementing emotion recognition in surveillance cameras could be used to analyse individuals' behaviours and eventually improve customers' experiences. We can use the example of a person walking in a shop to present an extreme scenario. If a surveillance camera analyses the body language of a person that goes anywhere near a certain type of product, and detects that the person is interested in buying it, a screen could make suggestions on how to use that product. For example, if the system detects that a client is willing to buy a food, the screen could display a video that shows how to cook with that product. Body movement detection systems work in specialised environments, but in order to implement such technologies in the everyday life, computers and smartphones need to be redesigned. Specific gestures such as finger movements are likely to be used to detect an emotion while using a smartphone, as evidenced by EmoteMail and the smartphone developed by Samsung, but it is unlikely that current machines are able to detect an entire body and react in real-time. Furthermore, people tend to stay in the same position while working at a computer, which makes it hard for a system to detect changes in emotional state. As has already been said, the main issue concerning systems that detect emotions through facial expressions, speech, and body movement, is the fact that people tend to communicate affective state in a spontaneous way, and, therefore, do not exaggerate their emotions while sitting in front of a computer. Also, these methods have to deal with "fake emotions", and, therefore, with the possibility that a system could misinterpret an affective state. Physiological measurements are capable of detecting the slightest changes in blood pressure, heart rate, and electrodermal activity, and these changes cannot be controlled nor manipulated. Even though the main limit of physiological measurement devices seemed to be the invasive and expensive equipment needed, this analysis has presented some systems that have been implemented in devices that individuals are naturally in contact with. Physiological sensing devices can be worn, which means that the user has the primary control over it. Devices such as Affectiva Q Sensor have been developed to help individuals with cognitive disabilities in the first place, and their success has been proved. Physiological sensing systems arise from the need of improving the interaction between humans and humans, rather than the human-computer interaction. For this reason, it is highly probable that researchers will continue to better these devices and extend their scope to other fields as well. Emotion detection through physiological data does not require the activation of a webcam or the use of a recorder, which means that there is a high probability that such devices will be easily accepted by individuals. However, the need to understand where the personal data will be stored and who will have access to it still persists.

The *Nosedive* episode of Charlie Brooker's *Black Mirror* presents a world in which people can rate other individuals by assigning each other from one to five stars. The personal rating affects almost every aspect of people's lives, from buying train tickets to renting a house. Even though these kinds of scenarios are generally only associated with science fiction movies, they might actually reflect reality more than we think. The episode reflects a Chinese governmental initiative based on a Social Credit System, which is supposed to be tested in 2020. Infractions, such as bad driving, smoking in non-smoking zones, or buying to many videogames, will lead to lower scores and, therefore, denied access to some kinds of services (Ma, 2018).

Combining these kinds of system with Affective Computing devices would allow individuals to acquire full pictures of each other. In fact, social credit systems like the one described above, might deal with a person's behaviour and, therefore, define what the person does and how she acts, while affective systems are based on evidence about who the person *is*. We can use the example of an employer, who not only checks the Facebook, LinkedIn, and Instagram profile of a job applicant, but also analyses his or her "average emotional state", by accessing a device that keeps track of the person's emotional changes. He might notice that the person is often stressed, or sad, or angry, and might not hire him for this reason, even though he has all the necessary requirements for the job.

Even though the above example represents an extreme case, it is essential to consider such scenarios in order to figure out if and how it is possible to define the limits of Affective Computing technologies.

Although this analysis includes evidence about the potentials of physiological sensing devices, the ways such systems will affect individuals' everyday life are still unknown, and so are the problems that they will cause. Considering current, rapid technological progress, it is highly probable that the technical issues regarding those systems will soon be overcome. This means that the limits of such devices mainly depend of whether rules and limitations to their use will be established or not.

In *Affective Sensors, Privacy, and Ethical Contracts*, Picard and Reynolds (2004, p. 1103) explain the following:

> Computers, as they are currently designed, do not have the capability to be ethical on their own. Lacking free will, machines cannot make moral choices between "good" and "bad." Instead, they largely carry out their designer's choices. This means that if a designer makes "bad" choices from the user's perspective, the resulting interaction could be viewed as unethical.

Even though designers do not generally aim to develop unethical devices, the ultimate way in which technology is used (including unethical uses) cannot be controlled. However, Affective Computing researchers have tried to generate rules to protect individuals.

The Concerns for Information Privacy (CFIP) model developed by Smith et al. (1996) refers to four issues, namely *collection*, *secondary use*, *errors*, and *improper access*. In fact, privacy threats arise once the data about the affective state is made available, and can therefore be collected and stored somewhere. *Secondary use* and improper access refer to the fact that the data can be analysed and used by other individuals without authorisation, and *errors* expresses the concern about the protection of both deliberate and accidental errors. Researchers who deal with privacy issues in Affective Computing systems agree that users should have the primary control over the system. First, exposing affective states can have negative psychological impacts on some individuals, such as by forcing someone to reveal emotions that he or she prefers to keep secret, such as depression or obsessive-compulsive disorders. For this reason, emotions should only be showed with the user's consent, and individuals should have the option to choose who will have access to their affective information. Cooney and his colleagues (2018) also propose a "mutual exchange", namely a scenario in which a person can visualise someone else's emotions only if he or she offers his or her own feelings as well. Experiments such as those conducted by Steen and Kim (2010) have proved that individuals are more likely to accept a type of technology if it is clearly stated that the collected information may be used for other purposes. Also, people feel protected from privacy issues if they are offered a contract.

However, rules and restrictions might not be enough to avoid future issues. Technology has never had the aim of harming people, the biggest danger is generally how people use a specific technology. Despite regulations and limitations in Affective Computing systems it is, therefore, necessary to establish "unwritten rules", namely limits about the use that humans can make of other humans' emotional information. People should respect other persons' decisions about who has access to information about their emotional states. Also, if an individual gets the privilege to know about someone's affective state, he must be responsible for it and not misuse the information. However, the question whether these regulations will be respected or not cannot be answered until Affective Computing devices will actually become part of individuals' lives. Despite individuals' awareness of the privacy issues related to the use of technology, and even though people express their desire to protect personal information, their action are often contrary to their intentions (Belanger & Crossler, 2011). Referring to what he calls “privacy paradox”, Castano (2015, p. 33) states that “people divulge personal information for a small benefit, even though people expressed the desire to keep their information private”. Social media are, for instance, a demonstration of this privacy paradox. Even though the majority of users are aware of the fact that uploading a picture on a social media basically means making that picture available to a huge number of people, this awareness and concern do not seem to affect their behaviour. It is therefore possible (and likely) that the privacy concerns of Affective Computing will not prevent such devices from becoming part of individuals’ lives.

4.1 Future research

In this research paper, I focused on the process of Affective Understanding. However, the main purpose of Affective Computing devices is to create machines that are able to react to individuals’ emotional states. The process of Affective Generation includes further technical and ethical issues that need to be analysed in order to have a complete picture of the Affective Computing technology. If Affective Computing systems are applied in the everyday life, computers will be able to filter and adapt any kind of information to users’ emotional states, and, therefore, suggest songs, movies, or advertisements according to our moods. Picard (1997) highlights the so-called “synthesis problem”, that arises because “some models of emotional recognition within artificial intelligence do not necessarily involve the recognition of emotion but attempt to simulate the mechanisms responsible for the production of emotion” (Schwark 2015, p. 763). This means that, once the computer *gets to know* a user, it will also know what makes him or her happy. Therefore, if it identifies only “happy” factors, the system could expect the user to be happy without being able to identify happiness in his or her voice, face, or body language. Picard offers the example of a computer that knows its user has won the lottery and therefore expects to detect happiness without being able to recognise sadness in the user's face, due to fact that he is unable to find the winning lottery ticket. Also, as already mentioned, it is possible that a computer can misunderstand an emotion, and therefore respond in an improper way. It might not be a big problem if a music player misunderstands an affective state and suggest a "wrong" song, but errors in emotion detection and emotion generation could cause major problems in other situations.

5. Conclusion

Affective Computing systems have the potential to improve the interaction between humans and computers, and they are likely to transform some aspects of our lives, as well as improve several fields, such as e-learning and marketing. Affective Computing systems have been compared to existing devices that detect facial expressions, speech, body movements and physiological data to analyse whether current technologies are advanced enough to enable affective devices to improve human-computer interactions, and establish the ethical issues that the use of such devices might cause. I argued that, among the four methods used to detect an individual's affective state, physiological information detection retains the highest potential because physiological data cannot be performed, controlled, or manipulated, and because physiological sensing devices can be worn, and they therefore enable the user to have the primary control over them. In the research paper, I also dealt with the ethical issues regarding Affective Computing devices, providing evidence that technology that is designed to generate efficient and symmetrical human-computer interaction can actually be transformed into an asymmetrical tool. Privacy is the main issue concerning emotional systems, and researchers have established regulations that might be applied to future

Affective Computing devices. However, I claim that the ultimate consideration of emotion detection systems will only be clear once individuals have started using the systems. In fact, technology has never had the purpose of harming people, the biggest danger is how individuals use technology. It is, therefore, essential to establish limits about the use that humans can make of other humans' emotional information. At the same time, the privacy paradox shows that concerns about the use that others' can make of one's personal data do not necessarily prevent users from sharing personal information. For this reason, I claim that as long as Affective Computing devices are governed by laws that protect individuals' privacy, they are likely to be accepted by people and, therefore, to be implemented in the everyday life.

6. Bibliography

Amazon (2018) *Alexa Terms of Use*.

Anthony S. (2012). Samsung develops emotion-sensing smartphone. *Extreme Tech.* 9th January.

Apple (2017). *About Face ID advanced technology.*

Ark W., Dryer C. & Lu D. J. (1999) The emotion mouse In *Human-Computer Interaction: Ergonomics and User Interfaces* (Bullinger H. & Ziegler J., eds), Lawrence Erlbaum, New Jersey, 818-823.

Bahreini K., Nadolski R. & Westera W. (2016) Towards real-time speech emotion recognition for affective e-learning. *Education and Information Technologies* **21**(5), 1367–1386.

Bakardjieva E. & Kimmel A. J. (2017) Neuromarketing Research Practices: Attitudes, Ethics, and Behavioral Intentions. *Ethics & Behavior* **27**(3), 179-200.

Belanger F. & Crossler R. E. (2011) Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly* **35**(4), 1017-1041.

Belhumeur P. N., Hespanha J. P., & Kriegman D. J. (1997) Eigenfaces vs. Fisherfaces: Recognition Using Class Specic Linear Projection. *IEEE Transactions on pattern analysis and machine intelligence* **19**(7), 711-720.

Brow K. (2005) *Encyclopedia of Language and Linguistics 2nd Edition*, Elsevier Science.

Bullington J (2005) Affective Computing and Emotional Recognition Systems: The Future of Biometric Surveillance? *Proceedings of Information Security Curriculum Development Conference* 2005, 95-99.

Calder A. J., Burton A. M., Miller P, Young A. W. & Akamatsu S. (2001) A principal component analysis of facial expression. *Vision Research* **41**(9), 1179–1208.

Cannon W. B. (1927) The James-Lange Theory of Emotions: A Critical Examination and an Alternative Theory. *The American Journal of Psychology* **39**(1), 106-124.

Cardozo B. N. (1921) The Nature of the Judicial Process. Yale University Press, USA.

Castano D. C. (2015) *Affect and Online Privacy Concerns.* Unpublished Doctoral dissertation. Nova Southeastern University.

Cavallaro A. (2007). Privacy in video surveillance. *IEEE Signal Processing Magazine* **24**(2), 166–168.

Cristiani M., Murino V. & Vinciarelli A. (2010) Socially Intelligent Surveillance and Monitoring: Analysing Social Dimensions of Physical Space. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp.51-58.

Cooney M., Pashami S., Sant'Anna A., Fan Y. & Nowaczyk S. (2018) Pitfalls of Affective Computing. *WWW '18 Companion Proceedings of the The Web Conference*, 1563-1566.

Dael N., Mortillaro M. & Scherer K. R. (2012) The Body Action and Posture Coding System (BAP): Development and Reliability. *Journal of Nonverbal Behavior* **36**(2), 97-121.

de Gelder B., Teunisse J. P. & Benson P. J. (1997) Categorical perception of facial expressions: Categories and their internal structure. *Cognition and Emotion* **11**(1), 1–23.

Duchenne G. B. (1962) The Mechanisms of Human Facial Expression. Paris: *Archives générales de médecine* **1**, 152-174.

Elfenbein H. A. & Ambady N. (2003) Universals and cultural differences in understanding emotions. *Current Directions in Psychological Science* **12**(5), 159-164.

Ekman P., & Friesen W. V. (1971) Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology* **17**(2), 124-129.

Ekman P. (1993) Facial expression and emotion. *American Psychologist* **48**(4), 384-392.

Flaisch T., Häcker F., Renner B. & Schupp H. T. (2011) Emotion and the processing of symbolic gestures: an event-related brain potential study. *Social Cognitive and Affective Neuroscience* **6**(1), 109-18.

Freund Y. & Schapire R. E. (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* **55**(1), 119-139.

Furui S. (2000) *Digital Speech Processing, Synthesis, and Recognition.* CRC Press, Boca Raton, Florida, US.

Guerrero L., DeVito J. & Hecht M. (1999) *The nonverbal communication reader: Classic and contemporary readings.* Waveland Press Inc., USA.

Gupta O. & Gupta P. (2015) Affective computing: facial recognition technology. *36th IRF International conference. 48-52.*

Harrigan JA. (2005) Proxemics, kinesics, and gaze. In *The new handbook of methods in nonverbal behavior research* (J. A. Harrigan, R. Rosenthal & K. Scherer, eds.), Oxford University Press, New York, pp. 137–198.

James W. (1884) What is an emotion? *Mind* **9**(34), 188-205.

Jaywant A. & Pell M. D. (2012) Categorical processing of negative emotions from speech prosody. *Speech Communication* **54**(1), 1-10.

Karg M., Samadani A., Gorbet R., Kuhnlenz K., Hoey J. & Kuli D. (2013) Body Movements for Affective Expression: A Survey of Automatic Recognition and Generation. *IEEE Transactions on affective computing* **4**(4), 341-359.

Kohonen T. (1989) *Self-organization and Associative Memory*. Springer-Verlag, Berlin, Heidelberg.

Labarre W. (1947) The cultural basis of emotions and gestures. *Journal of Personality* **16**(1), 49–68.

Lang R. D. & Benessere L. E. (2017). Alexa, Siri, Bixby, Google's assistant, and Cortana testifying the court. *Journal of the Missouri Bar* **74**(1), 20-23.

Ma A. (2018). China ranks citizens with a social credit system: here's what you can do wrong and how you can be punished. *Business Insider.* 10th April.

Marrin T., Picard R. W. (1998) The conductor's jacket: a device for recording expressive musical gestures. Appeared in *Proceedings of the International Computer Music Conference.*

McIntyre G. & Göcke R. (2006) Researching Emotions in Speech. In *Proceedings of the Eleventh Australasian International Conference on Speech Science and Technology SST2006*. ASSTA, Auckland, New Zealand, pp. 264-369.

Mehrabian A. (1972) *Nonverbal communication*. Aldine-Atherton, Chicago.

Mozziconacci S. (2002) Prosody and emotions. *Speech Prosody 2002*. Aix-en-Provence, France (April 11–13), 1–9.

Noroozi F., Corneanu C. A., Kaminska D., Sapinski T., Escalera S. & Anbarjafari G. (2018) Survey on Emotional Body Gesture Recognition. *Journal of IEEE transactions on affective computing* **20**(10).

Pantic M. (2014) Facial Expression Recognition. In *Encyclopedia of Biometrics* (Li S. & Jain A., eds). Springer, Boston, MA, pp. 1-8.

Pease A. & Pease B. (2004) *The definitive book of body language*. Pease International, Buderim, Australia.

Peter C., Ebert E. & Beikirch H. (2008) Physiological Sensing for Affective Computing. In *Affective Information Processing* (Tao J. & Tan T., eds), Springer, London, pp. 293-310.

Petrushin V.A. (2000). Emotion recognition in speech signal. Experimental study, delevelopment and application. *ICSLP 2000*, 222-225.

Picard R. W. (1997) *Affective computing.* MIT Press Ltd, Cambridge, Massachusetts.

Reynolds C. & Picard R. W. (2004) Affective sensors, privacy, and ethical contracts. *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, 1103-1106.

Rodero E. (2011) Intonation and Emotion: Influence of Pitch Levels and Contour Type on Creating Emotions. *Journal of Voice* **25**(1), 25-34.

Schachter S. & Singer J. (1962) Cognitive, social, and physiological determinants of emotional state. *Psychological Review* **69**(5), 379-399.

Scheirer J., Fernandez R., Picard R. W. (1999) Expression glasses: a wearable device for facial expression recognition. *Proceedings of CHI'99*, Pittsburgh, PA, 262-263.

Scherer K. R. (1986) Vocal affect expression: A review and model for future research. *Psychological Bulletin* 99, 143-165.

Schwark J. (2015) Toward a Taxonomy of Affective Computing. *International Journal of Human–Computer Interaction* **31,** 761-786.

Smith J. H., Milberg S. J. & Burke S. J. (1996) Information privacy: Measuring individuals' concerns about organizational practices. *MIS Quarterly* **20**(2), 167- 196.

Steen L. J. & Kim F. (2010). Affective computing: invasive technology and legal considerations to protect consumers. *Issues in Information Systems* **10**(1), 577-584.

Tao J. & Tan T. (2005) Affective Computing: A Review. In *Affective Computing and Intelligent Interaction. First International Conference, ACII 2005, Beijing, China, October 22-24, 2005, Proceedings* (Tao J., Tan T. & Picard R.W., eds). Springer, Berlin, Heidelberg, pp. 981–995.

Tracy J. L. & Robins R. W. (2007) The prototypical pride expression: Development of a nonverbal behavior coding system. *Emotion* **7**(4), 789–801.
Velten E. (1968). A laboratory task for induction of mood states. *Behaviour Research and Therapy* 6, 473–482.

Viola P. & Jones M. J. (2004) Robust Real-Time Face Detection. *International Journal of Computer Vision* **57**(2), 137–154.

Welinder Y. (2012) A Face Tells More than a Thousand Posts: Developing Face Recognition Privacy in Social Networks. *Harvard Journal of Law and Technology* **26**(1), 165-239.

Whang M. & Lim J. (2008) A Physiological Approach to Affective Computing. In *Affective Computing* (Or J. ed.), IntechOpen, pp. 309-318.

Wittig H. (1999) Introduction. In *Intelligent Media Agents. Multimedia Engineering.* (Wittig, Effelsberg & Steinmetz), Vieweg+Teubner Verlag, Wiesbaden, pp. 1-14.

Zimmermann P., Guttormsen S., Danuser B. & Gomez P. (2003) Affective Computing—A Rationale for Measuring Mood with Mouse and Keyboard. *International Journal of Occupational Safety and Ergonomics* **9**(4), 539-551.