

Document downloaded from:

<http://hdl.handle.net/10251/82818>

This paper must be cited as:

Rincón Arango, JA.; Julian Inglada, VJ.; Carrascosa Casamayor, C. (2016). Developing an emotional-based application for human-agent societies. *Soft Computing*. 20(11):4217-4228. doi:10.1007/s00500-016-2289-5.



The final publication is available at

<https://link.springer.com/article/10.1007/s00500-016-2289-5>

Copyright Springer Verlag (Germany)

Additional Information

The final publication is available at Springer via <http://dx.doi.org/10.1007/s00500-016-2289-5>

Developing an Emotional-based Application for Human-Agent Societies

J. A. Rincon · V. Julian · C. Carrascosa

Received: date / Accepted: date

Abstract The purpose of this paper is to present an emotional-based application for human-agent societies. This kind of applications are those where virtual agents and humans coexist and interact transparently into a fully integrated environment. Specifically, the paper presents an application where humans are immersed into a system that extracts and analyzes the emotional states of a human group trying to maximize the welfare of those humans by playing the most appropriate music in every moment. This system can be used not only online, calculating the emotional reaction of people in a bar to a new song, but also in simulation, to predict the people's reaction to changes in music or in the bar layout.

1 Introduction

Ubiquitous Computing and Ambient Intelligence (AmI) [33], [21] have changed the concept of smart homes, introducing devices that improve people's quality of life. Among other applications, we can highlight smart devices that learn our tastes, smart homes that help reducing energy consumption [13], or safer homes for elderly [15]. To achieve this, developers may use different Artificial Intelligence (AI) tools, sensor networks, and new and sophisticated embedded devices.

The development of applications which employ this new technology, as well as the introduction of multi-agent systems in many of these applications, has allowed the creation of systems in which the human is completely involved in the system. Currently there exist applications in which intelligent agents offer different kinds of complex services to humans in an environment of whole

Universitat Politècnica de València
Departamento de Sistemas Informáticos y Computación (DSIC)
Camino de Vera s/n, Valencia, Spain
{jrincon, vinglada, carrasco}@dsic.upv.es

integration. This kind of applications are what we call a *Human-Agent Society* [4], which can be defined as a new computing paradigm where the traditional notion of application disappears. This new paradigm is based on an immersion of the users in a complex environment that enables computation.

The main challenge to achieve real human-agent societies lies in the design and construction of intelligent environments, in which humans interact with autonomous, intelligent entities, through different input and output devices. This means that there are two layers in which humans interact within the environmental and ubiquitous computing intelligence. The first layer is the real world where the human being interacts with other humans and with real objects. The second layer is a virtual layer in which humans interact with virtual entities and objects. This latter layer will be inhabited by intelligent entities (agents), which must be able to perform the different human orders. The virtual environments where agents are involved, are known in the literature as intelligent virtual environments or *IVE*. An *IVE* [12] is a 3D space that provides the user with a collaboration, simulation and interaction with software entities, so the user can experience a high immersion level. This immersion is achieved through detailed graphics, realistic physics, AI techniques and a set of devices that obtain information from the real world. As an example, the *JaCalIVE* framework [30] enables the design, programming and deployment of systems of this kind.

On the other hand, the development of applications in which people's emotions are taken into account has become very popular over the last few years, but it is a hard and complex process. Humans use their emotions in their decision making. Human beings manage themselves in different environments, either in the working place, at home or in public places. At each one of these places people perceive a wide range of stimuli, that interfere in our commodity levels by modifying our emotional levels. These emotional levels can be expressed in many different ways. The most common is the use of the face to express emotions. The expressed emotion can be detected using an algorithm which analyzes the facial features [36]. The information extracted by this algorithm allows us to identify the detected persons and estimate their emotional state. This emotional information has been very useful in the development of applications of AmI [20] and Robotics[6]. Nevertheless, it is needed for the machines to have the capability of interpreting or recognizing emotional variations. This is the reason for the design of emotional models that interpret or represent the different emotions. These emotional models can be embedded into intelligent agents, which have the capability of perceiving the world, acting and communicating between them. Therefore, humans could be represented in a virtual environment as intelligent agents, which can detect and analyze the emotional state of a specific person and act based on that emotional analysis.

This paper presents a new way of interacting between humans and intelligent autonomous entities living within intelligent virtual environments using emotions as a form of communication. To do this a number of AI tools were used, thus creating a hybrid application which is composed of a multi-agent

system, machine learning and statistical classification mechanisms. All this integration of different AI tools is tested in an emotional-based application which is presented in this article. This application controls the music played in a bar with the goal of making all the individuals inside it as happy as possible. In order to achieve this, we use different algorithms for capturing the environment, for detecting faces and music, for converting inputs into emotional values and for analyzing the decision making of the entities. The application has been developed using the *JaCalIVE* framework [30], which is a framework for the design and simulation of intelligent virtual environments. This framework differs from other works in the sense that it integrates the concepts of agents, humans, artifacts and a physical simulation. The main reason to employ *JaCalIVE* is that it allows an easy integration of human beings into the system.

The rest of the paper is organized as follows: Section 2 presents the related work; Section 3 describes the proposed problem; the application design is presented in Section 4, and Section 5 explains implementation details. Finally, some conclusions are presented in Section 6.

2 Related Work

Over the last few years different approaches have been proposed in order to define the emotional state of an entity. Among them, the best known are the *OCC* [25] and *PAD* [24] emotional models. The *OCC* model designed by Ortony, Clore & Collins is a model frequently used in applications where an emotional state must be simulated. This has allowed to create applications which emulate emotions in virtual humans [3] or applications which incorporate agents reacting to stress situations [17]. The *OCC* model specifies 22 emotional categories, which are divided into five processes. These processes define the whole system, where the emotional states represent the way of perceiving our environment (objects, persons, places) and, at the same time, influencing in our behavior positively or negatively [1]. However, the *OCC* model utilization presents one important complication due mainly to its high dimensionality which greatly difficult computation.

On the other hand, the *PAD* is a simplified view of the *OCC* model. This model allows to represent the different emotional states using only three values. These three values are usually normalized in $[-1, 1]$, and correspond to the three components conforming the emotional model (*Pleasure*, *Arousal*, *Dominance*). These components can be represented in a \mathbb{R}^3 space. Each one of the components conforming the *PAD* model allows to influence the emotional state of an individual in a positive or negative way. This influence evaluates the emotional predisposition of such individual, modifying in this way his/her emotional state. For instance, if the *Pleasure* parameter is modified so that its value is positive (the same can be achieved modifying its opposite *Displeasure* with negative values), then such an individual would tend to prefer pleasant stimuli rather than unpleasant ones. In the same way, if the *Arousal* parameter

is modified so that its value is positive (or *Calm* is modified to have negative values) then the excitation levels and the duration of it will change. Last, the *Dominance* parameter can be modified (or its opposite *Submissiveness*) to indicate the usual inclination to feel in control facing different situations, and the relation existing with feeling controlled in different circumstances.

The emotional model is only a tool for an intelligent agent in order to understand and represent the human emotion. But, the detection of human emotions by the agent implies the use of different recognition techniques. For the recognition of human emotions, biosignals [7] and body gestures [26] are typically employed. This last feature is maybe the most common technique for the classification of the human emotions. To do this, it is necessary the employ different machine learning techniques. In the literature can be found different applications in which authors use neural networks [16], support vector machines [41] and anthropometric measurements of the face [19]. All these techniques require a preliminary step that previously extracts the face information. This information is a series of points called Facial Feature Points (Figure 1).

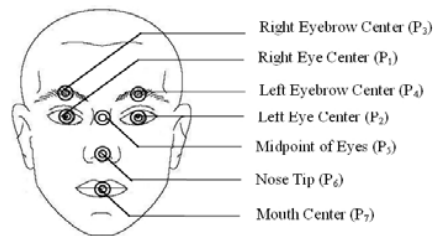


Fig. 1: Facial Feature Points

Based on these points, the distance between them is calculated (see Figure2). The next step consists of using these distances in order to train a neural network, a support vector machine or another machine learning technique for emotion recognition.

The existing emotional models and methods for detecting human emotions are usually employed in applications where multi-agent systems (MAS) are involved, allowing the simulation and the recognition of human emotions for a lonely entity. Emotional states turn into valuable information, allowing to develop applications that help to improve the human being life quality.

On the other hand, over the last few years, there have been different approaches for using MAS as a paradigm for modeling and engineering IVEs and how to introduce the human into this IVE. The use of MAS is mainly due to the characteristics that agents have, such as autonomy, proactivity, reactivity and sociability. Different approaches have been devised in order to develop MAS. All approaches for creating multi-agent systems that exist to

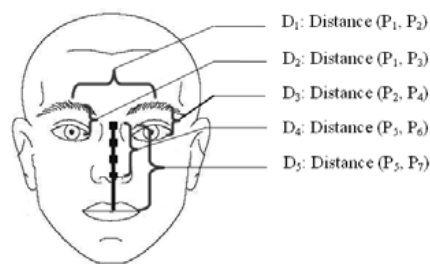


Fig. 2: Distances between Facial Feature Points

date should be studied according to two principal categories: those that simply support the creation and interaction of agents (JADE, April, Jason, JIAC), and those that permit the creation of virtual organizations (VO) with such key concepts as norms and roles (MadKIT, Jack, S-MOISE+, AMELI, Janus, Magentix2, Pangea). Until now, these platforms can create agents (some with different models), follow their life cycle and manage communication and services. And in the case of VO support, the platforms take into account the normative and organizational aspects that the platform itself should provide.

However, when modeling an IVE as a MAS it is necessary to take into account that not all the entities are agents. In this case, the main goal is to immerse human beings into IVEs. When a human being is completely immersed into a system of this kind, he/she can interact with the system in a natural way. Moreover, agents immersed in the system can learn about human actions adapting its behaviors and taking decisions about future situations. Examples of these systems can be domestic scenarios, production lines in an industry, entertainment industry, . . . This kind of applications is, as it has been commented in the introduction, a Human-Agent Society. Rather than developing software applications that accomplish computational tasks for specific purposes, this paradigm is based on an immersion of the users in a complex environment that enables computation [5]. Regarding the immersion of agents in dynamic physical, virtual or mixed environments, one of the main problems is to define what is the interface between the agents and their environment. One of the ways to solve this problem is to manage this interface as the agent body, i.e. a component that is attached to each agent to manage its interface with the environment [34]. Other works have introduced the use of mediators between agents and the environment [11][37], which can be viewed as functionally similar to bodies.

In order to achieve this kind of immersion, an IVE development process must employ a middleware specially designed for the execution and adaptation of Intelligent Virtual Environments, allowing an easy integration of human beings in the MAS. As far as we know, existing approaches do not allow an easy way to design systems of this kind. The tool that we propose is called JaCalIVE [28]. JaCalIVE specifically defines a type of entity which is a virtual

representation of a human. This entity can have the capability of detecting the human, associate itself to it and, moreover, associate many other characteristic that are considered for a better human-agent interaction into the IVE. Moreover, JaCalIVE also facilitates the interface between the agents and their environment allowing the definition of architectural building elements and patterns for agent-environment interactions

3 Problem description

The application example is developed in a bar, where there is a DJ agent in charge of automatically playing music and, also, there is a specific number of people listening to the music. The idea behind is that music may change the emotional state of a person or a group, as it influences in the human mood. But not any kind of music, but music that matches people interests. According to this, the main goal of the DJ agent is to play music making that all individuals within the bar are mostly *happy* as possible. Every time that a song is played, each one of the persons placed in the bar will have an emotional response according to his/her musical taste. That is, depending on the musical genre of the song, people will respond varying their emotional states. Moreover, varying emotions of each person will modify the social emotion of the group of people. If the DJ agent could have a way to evaluate the emotional state of the people which are in the bar, he could know the effect that the songs have over the audience. This will help the DJ agent to decide whether to continue with the same musical genre or not in order to improve the emotional state of the group. In some sense, this approach can be considered as a group recommendation technique, where the system tries to recommend the best song for the group of people [22]. Nevertheless, this proposal differs from typical group recommenders where typically the main issue arises how to combine individual user models in order to decide what would be optimal for the group. In this proposal it is not known what is good for each individual user, all the decisions are based on emotion recognition techniques.

This idea differs from other systems such as MUSICFX[23], where the users vote for the played songs, and they could even cheat to only play their favourite songs, in the sense that the users choose the songs unintentionally, avoiding the possibility of cheating for choosing only their favourite songs.

In the proposed application, each one of the individuals will be represented in the system as an intelligent agent, which will have an emotional response according to the person that represents. Moreover, these agents will try to recognize the real emotional states of people through the detection of changes in their facial expressions. With all of this, depending on the song that is being played in the bar, agents representing humans will be able to vary their emotional states in accordance with humans.

Specifically, the operating mode of the proposed application will be the following: in a specific moment the DJ agent will play a song in the bar. In such a way, the proposed application will seek to identify the different emotional

states using them as a tool of communication between humans and agents. To do this, agents that represent humans need to recognize the human associated to it, and moreover, recognize the musical genre of the song and, also, recognize the emotion. To perform these detection processes, we need to use pattern recognition algorithms and image and audio processing techniques in order to detect and classify the different emotional states of humans and try to modify the environment in which the humans are. Concretely, the proposed emotional detection uses image processing techniques as a way to capture information about changes in the facial expressions of the persons that are listening to the music into the bar. Moreover, the system can also capture the ambient sound in order to classify the music that is being played. Next subsections describe how the application has been designed and the description of the different employed techniques.

4 Application Design

Over the last few years, there have been different approaches for using MAS as a paradigm for modeling and engineering IVEs, but they have some open issues like low generality and then reusability, and weak support for handling full open and dynamic environments where objects are dynamically created and destroyed. As a way to tackle these open issues, and based on the MAM5 meta-model [2], the *JaCalIVE* framework was developed [30]. This framework provides a method to develop this kind of applications along with a supporting platform to execute them.

The presented work has used an extension of both the MAM5 meta-model along with the *JaCalIVE* framework to develop Human-Agent Societies, that is, to include humans in the loop. MAM5 allows to design an IVE as a set of entities that can be virtually situated or not. These entities are grouped inside Workspaces (*IVE Workspaces* in the case of virtually situated). Entities are divided into Artifacts (*IVE Artifacts* if situated) and Agents (*Inhabitant Agents* if situated). One new type of situated agents, that is, of *Inhabitant Agents* are *Human-Immersed Agents*, that model the human inside the system.

The application has been developed as a virtual multi-agent system using the *JaCalIVE* framework where there are will be different entities. Each one of these entities may represent not only real or simulated human beings, as the DJ agent, but also the furniture or the speakers with their location (Figure 3). The different types of agents and their main tasks are:

- DJ agent: the main goal of this agent is to achieve an emotional state of happiness for all of the people which are in the bar. When the DJ agent plays a song, it must analyze the emotional state of people. According to this analysis it will select the most appropriate songs in order to improve, if possible, the current emotional state of the audience.
- *Human-Immersed Agent*: it is in charge of detecting and calculating the emotional state of an individual which is in the bar, sending this information to the DJ agent. In order to accomplish its tasks, this agent must have

access to a variety of input/output information devices such as cameras, microphones, ...

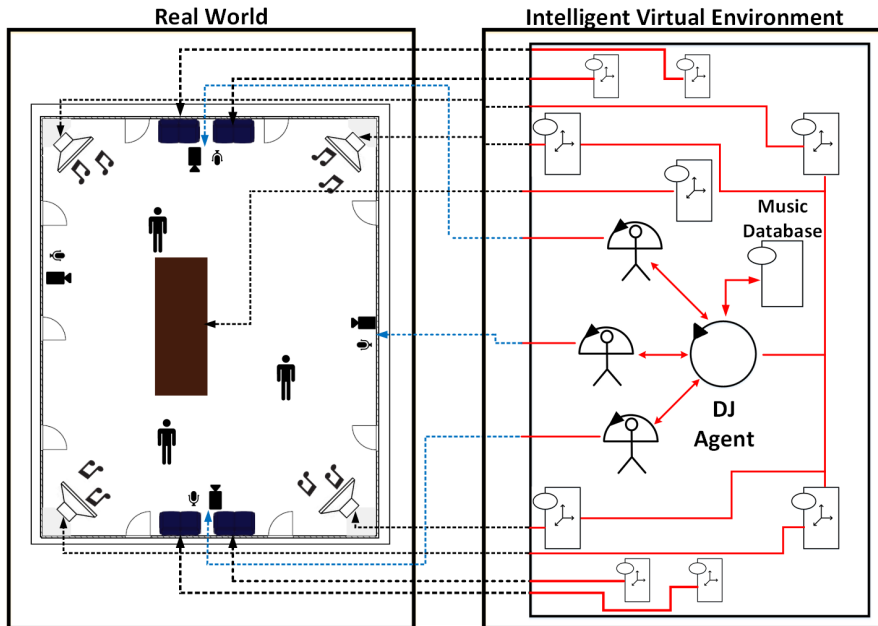


Fig. 3: General view of the different entities involved in the proposed application

In order to facilitate the access to this kind of devices, they have been modeled as artifacts (as can be seen in Figure 3)¹. Concretely, there has been designed an artifact for managing each camera which allow the face detection; each microphone is managed by an artifact which captures the ambient sound in order to classify the musical genre; the music DB has been designed as an artifact employed by the DJ agent (it stores around 1.000 songs classified by genres) and there is an artifact for controlling the multimedia player and the amplifiers for playing songs in the bar with the appropriated volume.

Each one of these entities has been designed using *JaCalIVE* through an XML file describing all its different properties (including physical ones). The XML file allows to describe if you need some kind of sensor to capture information, some type of actuator or simply an agent that does not need real-world information. It is also in this XML where humans are associated with each agent. These XML files are automatically translated into code templates using the *JaCalIVE* framework.

¹ In this Figure and in the following ones, we are using data from a simple example with only 3 Human-Immersed Agents

5 Implementation

As above commented, the proposed application includes two different types of agents: the DJ Agent and the Human-Immersed Agent. Next subsections explain the functionalities of these two agents.

5.1 DJ Agent

Regarding the DJ Agent, this agent uses the information sent by the Human-Immersed Agent to analyze the group's emotional state, using it to decide which is the next song to play. The goal of the DJ Agent is that all the humans feel as happy as possible. To analyze the current or to detect which emotion of the group, the DJ agent uses the social emotional model presented in [29]. This model employs the individual emotions sent by each *Human-Immersed Agent* and gives a way to calculate an aggregated emotional state for a group of agents. This social emotion allows the DJ agent to make a decision about next song to be played.

The process that follows the DJ agent is the following: in each cycle the DJ agent plays a song. Once the song has ended, the DJ evaluates the social emotion of the group of listeners that are within the bar (according to the individual emotional states received from the *Human-Immersed Agent*). In this way, the DJ agent can evaluate the effect that the song has had over the audience. This will help the DJ agent to decide whether to continue with the same musical genre or not in order to improve the emotional state of the group.

To do this, in our current implementation, the DJ agent has a connection with Spotify². In Spotify we have the songs classified into different musical genres (i.e. Rock, Pop, Classical, Country, Jazz and Techno). When the DJ agent must play the next song, it randomly chooses a song of the most appropriated musical genre. The DJ maintains a list of the last played songs (with a number of songs being configurable) to avoid repeating them. So, if the next song is inside this list, the DJ selects again other song. The chosen song will share the same musical genre of the previous song if the emotional state of the group tends to *happiness* or another musical genre if the emotional state moves away from *happiness*. It is very important to emphasize that the DJ does not have absolute knowledge of how the different human beings emotions may be affected by external factors, it only knows how the musical genre can make them happy.

5.2 Human-Immersed Agent

The Human-Immersed agent is an entity that is associated with a specific human and it is the representation of this human into the virtual environment.

² <https://www.spotify.com/>

This agent allows the data acquisition from the real world that are related with the human directly associated with the agent. This data will be used by the agent in order to estimate the emotional state of the human. After this estimation, the agent will communicate the estimated emotional state to the DJ Agent. These estimations and shippings of the emotional states are repeated in each execution cycle of the agent.

To do this, it is necessary to provide each Human-Immersed agent with a set of tools so that the DJ agent would be able to know the emotional state of each human. Those tools will help Human-Immersed agents to perceive the real environment, to interpret human it has associated and be able to classify the different emotional states this human expresses. So, each one of these agents contains both audio processing and image recognition mechanisms. In order to achieve the detection of the emotional states, each one of the Human-Immersed agents needs to perform a serie of four main processes in order to recognize the human and his/her emotional state. The first step is the data acquisition process which is divided into two subprocesses, a physical acquisition of the needed data and a face detection mechanism. The second process is in charge of the image processing in order to try to detect the emotional state of the human from his/her facial expressions. This process will be launched only if the face detection mechanism has properly ended. The third process is in charge of the audio processing. This process identifies the music that is being played and estimates the emotional state of the human according to his/her musical tastes. Finally, the last process is in charge of communicating the estimated emotional state to the DJ agent. It's important to remark that the image processing block and the audio processing block are two different ways to estimate the emotional state of a human. The agent will launch the image processing process only if the data acquisition process has been capable of detecting the human face, otherwise the agent will launch only the audio processing mechanism. Figure 4 shows all the processes and their relationships. These processes are explained in detail in the following subsections.

5.2.1 Data Acquisition

This first process to be executed by the agent corresponds with the acquisition of the needed information from the real world. Specifically, two physical devices are used for acquiring that information, a *WebCam* and a microphone. The information acquired by these two devices will be used to detect the emotional state of each human. Using these devices the *Human-Immersed agent* has the capability of listening and seeing the environment. The images captured by the webcam are processed using an identification mechanism which is based on the Viola and Jones algorithm [40]. This mechanism allows the face detection of the human that is associated with the agent. This process can fail in some occasions depending on the light conditions and the human movement. According to this, if the agent has been capable to identify the human face, the next step will try to estimate the human mood from the analysis of the facial gestures. On the contrary, if the agent has not identified the face, it will launch the audio

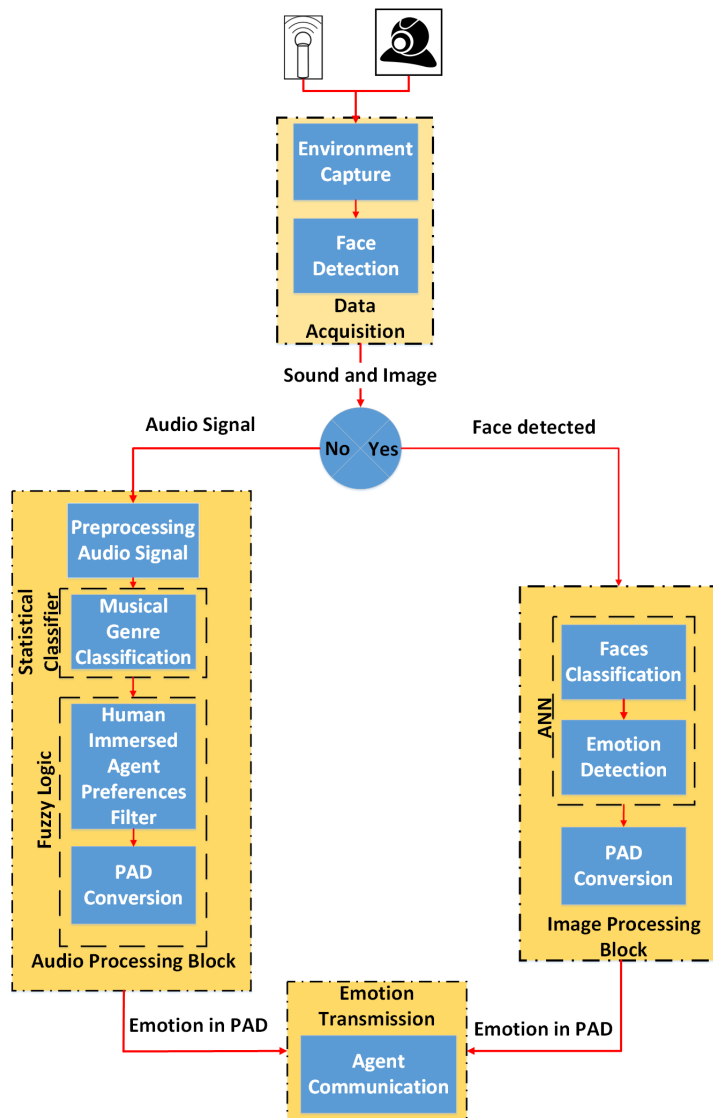


Fig. 4: Process sequence in the Human-Immersed agent

processing block. Regarding the audio acquisition, the Human-Immersed agent captures the song that is being played using a microphone. This microphone captures the environment sound which is stored as a *.wav* file. The recorded sound will be analyzed in a subsequent process.

5.2.2 Image Processing Block

In this block the *Human-Immersed Agent* divides the process in two sub-processes, *Face Classification* and *Emotion Recognition*. In the face classification process, the *Human-Immersed Agent* uses feature face points [42] to recognize the human associated to the agent. The API *Dlib*³ was used to extract these feature points of each one of the faces. These points were used to recognize the persons and to classify the emotions by calculating different distances between the points. This process is very important because is in which the *Human-Immersed Agent* learns how is the human that the agent has associated and it only captures information from that human. The face classification process uses feature faces points [42]. In our case twenty two points were used, and based on these points eighteen Euclidean distances were calculated. With these distances a vector for each person was created. This process was repeated in order to obtain seventy-four vectors. These vectors were grouped to create a matrix of 18x74, where 18 corresponding to eighteen distances between face points and 74 corresponding to the number of samples, that is, the number of photographs being used for a person.

This matrix is the face information for each human associated to a *Human-Immersed Agent*. In our experiments we have employed three different classes (persons) creating a matrix of 18x222. One of the problems of this matrix is its size, because adding a new person implies to add 74 new columns to the matrix. For this reason, it is very important to try to reduce the dimensionality of this matrix and to find a linear separation between classes. Thus, we have used Linear Discriminant Analysis (LDA) [8] to find linear separation between different classes. Figure 5 shows the distances between points before the use of *LDA* and Figure 6 shows the points after using *LDA*.

For the classification process, we have used a logistic regression. Figure 7 shows the classification map, with the delimitations for each class. To classify the images, the agent calculates the distance between points and do the test. An example of the obtained results after the test can be seen in Figure 8.

Once the face has been classified, the next step is to detect the emotion. To classify the emotion we use eight face feature points [41]. Based on these points, we calculate the displacements distance (Euclidean distances) of the feature points between a neutral facial expression and the six particular emotive expressions (Afraid, Angry, Disgusted, Happy, Sad and Surprised) (as it can be seen in Figure 9). It can establish a characteristic motion pattern for each expressed emotion. These characteristics were used to train a machine learning model. The training method used the database of *KDEF Emotional Lab at Karolinska Institute* [10] which is divided into 4900 pictures of 562 x 762 pixels and it specifies the related emotion (Afraid, Angry, Disgusted, Happy, Neutral, Sad or Surprised). In this way, in order to detect the emotions, it is necessary to extract some specific attributes from the images. These attributes

³ <http://dlib.net/>

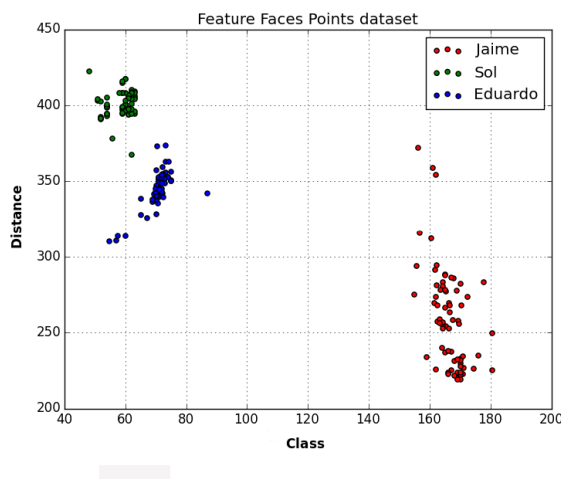


Fig. 5: Distance Feature Face Points without LDA

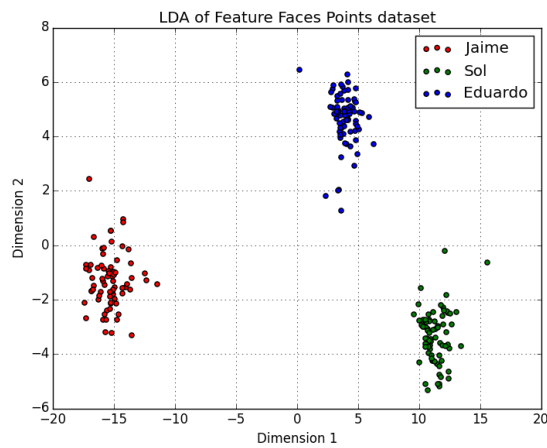


Fig. 6: Distance Feature Face Points with LDA

are the distance between some specific points of the face. Figure 10 specifies the eight points with which the displacement vector can be calculated.

In the training process, the ninety percent of the images from the database were used and the rest of the images were dedicated to the test stage. The machine learning algorithm used to recognize human emotions is an artificial neural network (*ANN*). The architecture of the neural network is shown in Figure11. The inputs are the eighth distances calculated using the database

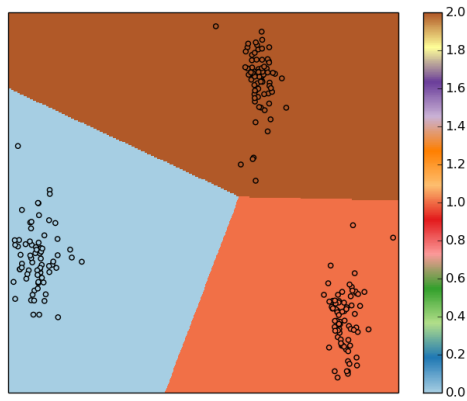


Fig. 7: Example of the classification map

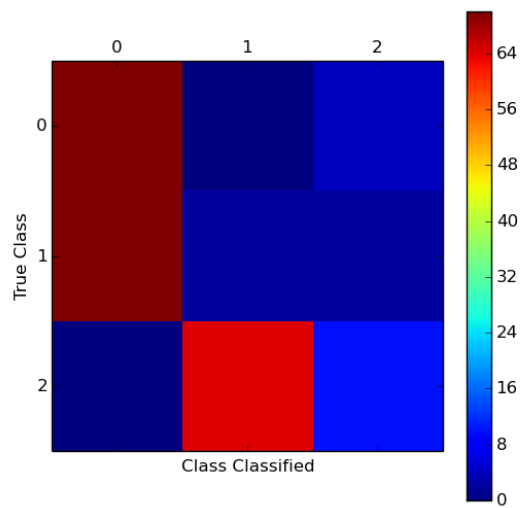


Fig. 8: Sample output of the Face Classification

and the output are the seven emotions to recognize (in fact, the ANN only classifies 6 emotions, the neutral emotion is obtained when the ANN does not return any value). After this process, the agent has an estimation of the emo-

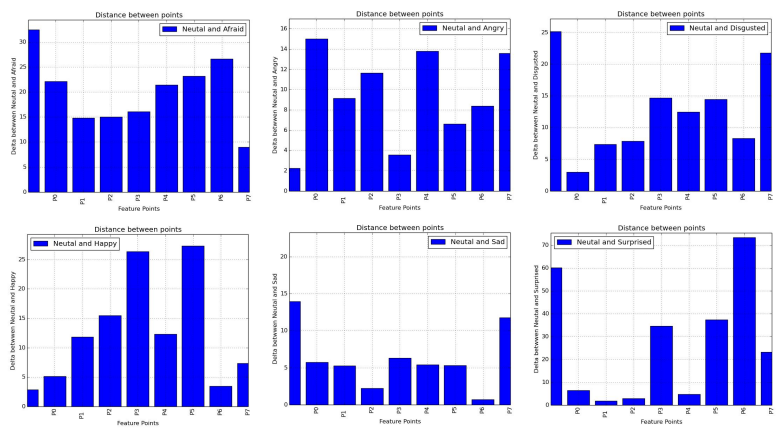


Fig. 9: Distance between Neutral Face Points and Face Points of Afraid, Angry, Disgusted, Happy, Sad and Surprised emotions



Fig. 10: Example of the Feature Face Points detection

tional state of the human through an analysis of the facial gestures. Following the equivalences shown in Table 1 (the values were extracted from [38]) the agent transforms the estimated emotion to a PAD value. For instance, the emotion *Happy* corresponds with the PAD values of [0.55, 0.24, 0.28]. This

	P	A	D
Afraid	-0.08	0.18	-0.39
Angry	-0.51	0.59	0.25
Disgusted	-0.36	0.08	0.13
Happy	0.55	0.24	0.28
Sad	-0.18	0.03	-0.14
Surprised	0.34	0.34	0.04

Table 1: Mapping between emotions and PAD values

emotional state expressed as a PAD value will be transmitted to the DJ agent in a subsequent process.

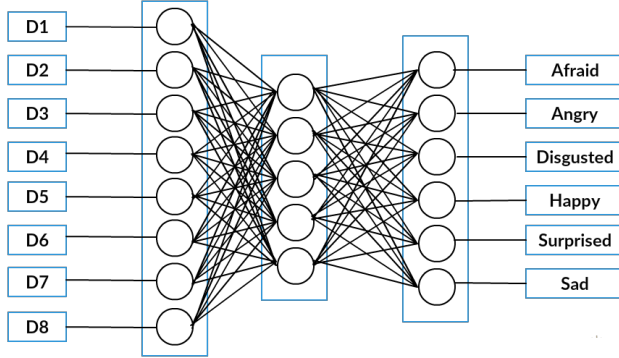


Fig. 11: Graphical view of the ANN Architecture used for Emotion Recognition

Moreover, some experiments have been made in order to evaluate how the light conditions into the bar can potentially affect the ML models. In this way we have test the proposed ANN in different situations, changing the luminosity (100%, 60%, 40%) and using different cameras such as RGB or infra-red (IR). Table 2 shows the obtained results, as can be seen, the classification rate gets worse when reducing luminosity. On the other hand, the RGB camera has been tested with normal luminosity levels without improving the results obtained with a normal camera. Nevertheless, the IR camera notably improves results obtained with low luminosity. According to these results, it's obvious that the best results are obtained with high luminosity levels, which can be complicated in some kind of bars. In these cases, IR camera can be used but with lower probability of success.

5.2.3 Audio Processing Block

This process will be executed if the agent is not able to detect the human face. In this case, the agent tries to detect the musical genre from the previously recorded sound.

	Classification Rate	MSE
Light 100%	93.5	0.063
Light 60%	74.1	0.112
Light 50%	40.5	0.450
RGB Camera%	93.1	0.062
IR Camera%	84.5	0.091

Table 2: Experiment results with different luminosity levels and cameras.

This block is divided in two phases, the first one is an offline training the second one is the execution of the machine learning model into each *Human-Immersed Agent*, along with the transformation to an emotion expressed as a PAD value:

- *Phase 1: Off-line training:* The offline process consists of an statistical model training [27][14][39] for the recognition of six musical genres (classical music, jazz, country, pop, rock and techno music). This model is the same for all the agents, that will use it to identify the musical genre.
- *Phase 2: On-line execution:* This phase consists of the following steps:
 1. *Musical genre classification:* Each *Human-Immersed Agent* executes its musical genre classifier, to identify the song being played by the DJ.
 2. *Human Immersed Agent preferences filter:* Once the agent has obtained the musical genre, it is necessary to extract the emotion associated to this genre. As each agent is associated to a human, it knows the musical preference of its associated human. This knowledge is a list of musical preferences that was given for its associated human to him. Based on this information, it was used a fuzzy-logic algorithm to get the *PAD* value corresponding to the emotion produced by the musical genre in the human associated to the agent. To get the best membership function several experiments were performed by modifying the functions of fuzzy logic, using triangular functions, Gaussian and trapezoidal. Modifying each of these functions in the model, the answer could vary the input *PAD* values. In our experiment we used the triangular membership.

It is important to highlight that the ambient noise is a serious problem. For this reason, it is important to have a series of filters in the *Human-Immersed Agent* as *Butterworth* with different configurations [35], [18].

The variation of the membership function depends on the corresponding human musical preferences, e.g. a human can respond favorably to pop music but not to the blues music. This is the reason why it is needed that each human configure its *Human-Immersed Agent* before using the system, varying the membership function of the fuzzy logic module.

5.2.4 Emotion Transmission

Finally, once the emotion has been estimated, it is necessary to send this emotional value to the *DJ Agent* in order to calculate the social emotion [29]

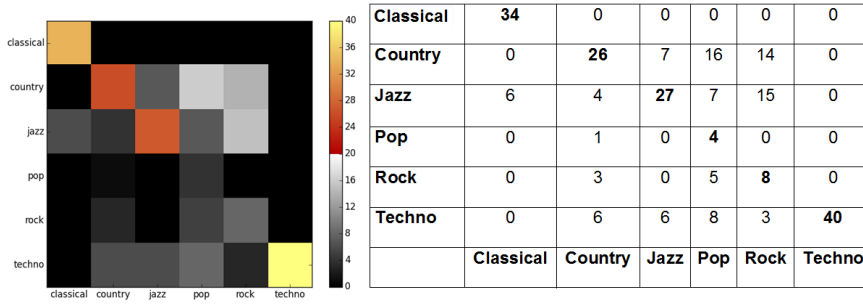


Fig. 12: Example of a Numerical and Graphical Confusion Matrix.

of all of the people which are in the bar. This social emotion model employs a representation of the emotional value according to the *PAD* model where the emotion is represented as a vector in the \mathbb{R}^3 space and each component have a scale range from $[-1.0, 1.0]$. These *PAD* values are sent to the *DJ agent* which receives all the emotional values from all the Human-Immersed agents that are active.

6 Conclusions and future work

Multi-agent systems allow the design and implementation of applications where the main components can be humans and software agents. These agents must be able to interact and communicate with humans in order to help them in their daily activities. In this sense, this paper presents an ambient intelligence application where humans and agents must coexist in a framework of maximum integration. The application has been developed over the *JaCalIVE* framework allowing an easy integration of the human in the multi-agent system and a visualization of the system in a virtual environment. Thus allowing the agents to learn and detect the human emotion and recognize the human associated to each *Human-Immersed Agent*. The proposed system is able to extract (in a non-invasive way) and to analyze the social emotion of a group of persons and it can take decisions according to that emotional state. Moreover, it can be used in an online fashion, where the system is reflecting what is happening in the real world at the same time, or it can be used to simulate - predict what would happen in the conditions that are used in the simulation, where this conditions can be about the number of agents and their preferences and also about the layout of the environment. As a future work we are applying this system to other application domains, as it can be extracted out of the music domain, and carried out to any other ambient intelligence domain, even an industrial one, where it can monitor and simulate the conditions inside a factory.

On the other hand, another aspect to be considered as future work is the improvement of users' satisfaction including other elements in the decision-making such as the opinion of the user. The idea is to include automated negotiation techniques, like employed in [9], in order to select a subset of songs or genres that everyone likes in some degree. This will reduce the complexity of the song selection process made by the DJ agent. Moreover, groups of people with very similar musical tastes could be detected into the bar. These groups can play as negotiation teams in this process, as in done in [32] and [31]. We are also planning the deployment of our proposal in the MEDERI⁴ living lab in our university. This living lab is a multidisciplinary environment, mainly focused on health technology, that will give us the needed tools for a real involvement of users in order to improve the experience and robustness of the proposed techniques.

7 Compliance with Ethical Standards

Funding: This work is partially supported by the MINECO/FEDER TIN2015-65515-C4-1-R and the FPI grant AP2013-01276 awarded to Jaime-Andres Rincon.

Conflict of Interest: Jaime Rincon declares that he has no conflict of interest. Vicente Julian declares that he has no conflict of interest. Carlos Carrascosa declares that he has no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors. All the data used in the experiments were obtained from public datasets.

References

1. Ali F, Amin M (2013) The influence of physical environment on emotions, customer satisfaction and behavioural intentions in chinese resort hotel industry. In: KMITL-AGBA Conference Bangkok, pp 15–17
2. Barella A, Ricci A, Boissier O, Carrascosa C (2012) MAM5: Multi-agent model for intelligent virtual environments. In: 10th European Workshop on Multi-Agent Systems (EUMAS 2012), pp 16–30
3. Becker-Asano C, Wachsmuth I (2010) Affective computing with primary and secondary emotions in a virtual human. *Autonomous Agents and Multi-Agent Systems* 20(1):32–49, DOI 10.1007/s10458-009-9094-9, URL <http://link.springer.com/10.1007/s10458-009-9094-9>
4. Billhardt H, Julián V, Corchado J, Fernández A (2014) An architecture proposal for human-agent societies. In: *Highlights of Practical Applications of Heterogeneous Multi-Agent Systems. The PAAMS Collection, Communications in Computer and Information Science*, vol 430, Springer

⁴ <http://mederi.ai2.upv.es/>

- International Publishing, pp 344–357, DOI 10.1007/978-3-319-07767-3_31, URL http://dx.doi.org/10.1007/978-3-319-07767-3_31
5. Billhardt H, Julián V, Corchado JM, Fernández A (2015) Human-agent societies: challenges and issues. *International Journal of Artificial Intelligence*TM 13(1):28–44
 6. Broekens J (2007) Emotion and reinforcement: Affective facial expressions facilitate robot learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 4451 LNAI:113–132, DOI 10.1007/978-3-540-72348-6_6
 7. Canento F, Fred A, Silva H, Gamboa H, Lourenço A (2011) Multimodal biosignal sensor data handling for emotion recognition. In: *Sensors, 2011 IEEE*, IEEE, pp 647–650
 8. Delac K, Grgic M, Grgic S (2005) Statistics in face recognition: analyzing probability distributions of PCA, ICA and LDA performance results. *ISPA 2005 Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis, 2005* pp 289–294, DOI 10.1109/ISPA.2005.195425
 9. Esparcia S, Sánchez-Anguix V, Aydogan R (2013) A negotiation approach for energy-aware room allocation systems. In: *1st Workshop on Conflict Resolution in Decision Making (COREDEMA 2013)*, Springer, vol 365, pp 280–291
 10. GOELEVELN E, De Raedt R, LEYMAN L, Verschuere B (2008) The karolinska directed emotional faces: A validation study. *COGNITION & EMOTION* 22(6):1094–1118
 11. Gouaïch A, Michel F, Guiraud Y (2005) MIC*: A Deployment Environment for Autonomous Agents, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 109–126
 12. Hale K, Stanney K (2002) *Handbook of Virtual Environments: Design, Implementation, and Applications*. Human Factors and Ergonomics, Taylor & Francis
 13. Han DM, Lim JH (2010) Smart home energy management system using IEEE 802.15. 4 and zigbee. *Consumer Electronics, IEEE Transactions on* 56(3):1403–1410, URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5606276
 14. Holzapfel A, Stylianou Y (2007) A statistical approach to musical genre classification using non-negative matrix factorization. In: *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, IEEE, vol 2, pp II–693
 15. Intille SS (2002) Designing a home of the future. *IEEE pervasive computing* 1(2):76–82
 16. Ioannou SV, Raouzaïou AT, Tzouvaras VA, Mailis TP, Karpouzis KC, Kollias SD (2005) Emotion recognition through facial expression analysis based on a neurofuzzy network. *Neural Networks* 18(4):423–435, DOI 10.1016/j.neunet.2005.03.004
 17. Jain D, Kobti Z (2011) Simulating the effect of emotional stress on task performance using OCC. In: *Advances in Artificial Intelligence*,

- Springer, pp 204–209, URL http://link.springer.com/chapter/10.1007/978-3-642-21043-3_24
18. Journal I, Technological F, Singla K (2014) Audio noise reduction using different filters 1,2 1(11):1373–1375
 19. Kim MH, Joo YH, Park JB (2005) Emotion detection algorithm using frontal face image. In: International Conference on Computer Application in Shipbuilding
 20. Leon E, Clarke G, Callaghan V, Doctor F (2010) Affect-aware behaviour modelling and control inside an intelligent environment. *Pervasive Mob Comput* 6(5):559–574, DOI 10.1016/j.pmcj.2009.12.002, URL <http://dx.doi.org/10.1016/j.pmcj.2009.12.002>
 21. Mangina E, Carbo J, Molina JM (2009) Agent-based ubiquitous computing. Atlantis Press : World Scientific, Amsterdam; Paris, URL <http://dx.doi.org/10.2991/978-94-91216-31-2>
 22. Masthoff J (2011) Group recommender systems: Combining individual models. In: *Recommender systems handbook*, Springer, pp 677–702
 23. McCarthy JF, Anagnost TD (1998) Musicfx: An arbiter of group preferences for computer supported collaborative workouts. In: *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*, ACM, New York, NY, USA, CSCW '98, pp 363–372
 24. Mehrabian A (1997) Analysis of affiliation-related traits in terms of the PAD temperament model. *The Journal of Psychology* 131(1):101–117, DOI 10.1080/00223989709603508, URL <http://www.tandfonline.com/doi/abs/10.1080/00223989709603508>
 25. Ortony A (1990) *The Cognitive Structure of Emotions*. Cambridge University Press
 26. Piana S, Odone F, Verri A, Camurri A (2014) Real-time Automatic Emotion Recognition from Body Gestures. arXiv preprint arXiv:14025047 pp 1–7, DOI arXiv:1402.5047, arXiv:1402.5047v1
 27. Richert W, Coelho LP (2013) *Building Machine Learning Systems with Python*. Packt Publishing, Birmingham
 28. Rincon J, Carrascosa C, Garcia E (2014) Developing Intelligent Virtual Environments using MAM5 Meta-Model. In: *International Conference on Practical Applications of Agents and Multi-Agent Systems*, Springer, pp 379–382
 29. Rincon J, Julian V, Carrascosa C (2015) Social Emotional Model. In: *13th International Conference on Practical Applications of Agents and Multi-Agent Systems*, LNAI, vol 9086, pp 199–210
 30. Rincon JA, Garcia E, Julian V, Carrascosa C (2014) Developing adaptive agents situated in intelligent virtual environments. In: *Hybrid Artificial Intelligence Systems*, no. 8480 in LNCS, Springer, pp 98–109
 31. Sanchez-Anguix V, Julian V, Botti V, Garcia-Fornes A (2013) Studying the Impact of Negotiation Environments on Negotiation Teams' Performance. *Information Sciences* 219:17–40
 32. Sanchez-Anguix V, Aydogan R, Julian V, Jonker C (2014) Unanimously acceptable agreements for negotiation teams in unpredictable domains.

- Electronic Commerce Research and Applications 13(4):243 – 265
33. Satyanarayanan M (2002) A catalyst for mobile and ubiquitous computing. *Pervasive Computing*, IEEE 1(1):2–5, URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=993138
 34. Saunier J, Carrascosa C, Galland S, Kanmeugne PS (2015) Agent bodies: An interface between agent and environment. In: *Agent Environments for Multi-Agent Systems IV*, Springer, pp 25–40
 35. Scott T, Green WB, Stuart A (2005) Interactive effects of low-pass filtering and masking noise on word recognition. *Journal of the American Academy of Audiology* 114(11):867–878
 36. Senechal T, Rapp V, Prevost L (2011) Facial feature tracking for emotional dynamic analysis. *Advances Concepts for Intelligent Vision ...* pp 495–506, DOI 10.1007/978-3-642-23687-7_45, URL http://link.springer.com/chapter/10.1007/978-3-642-23687-7_{_}45
 37. Thalmann D, Musse SR, Braun A (2007) *Crowd simulation*, vol 1. Springer
 38. Tsonos D, Stavropoulou P, Kouroupetroglou G, Deligiorgi D, Papatheodorou N (2014) Emotional prosodic model evaluation for greek expressive text-to-speech synthesis. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8514 LNCS(PART 2):166–174, DOI 10.1007/978-3-319-07440-5_16
 39. Tzanetakis G, Cook P (2002) Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing* 10(5):293–302, DOI 10.1109/TSA.2002.800560
 40. Viola P, Jones MJ (2004) Robust real-time face detection. *International journal of computer vision* 57(2):137–154, URL <http://link.springer.com/article/10.1023/B:VISI.0000013087.49260.fb>
 41. Visutsak P (2012) Emotion Classification Using Adaptive SVMs 1(3):279–282
 42. Vukadinovic D, Pantic M (2005) Fully automatic facial feature point detection using Gabor feature based boosted classifiers. In: *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, IEEE, vol 2, pp 1692–1698