Decision-making content of an agent affected by emotional feedback provided by capture of human's emotions through a Bimodal System

Javier G. Rázuri
Dept. of Computer and Systems Sciences (DSV), Stockholm University, Stockholm, SE-164 07 Kista, Sweden

Abstract

Affective computing allows for widening the view of the complex world in human-machine interaction through the comprehension of emotions, which allows an enriched coexistence of natural interactions between them. Corporal features such as facial expression, kinetics, structural components of the voice or vision, to mention just a few, provide us with valid information of how a human behaves. Among all the carriers of emotional information we may point out two, voice and facial gestures as holders of an ample potential for identifying emotions with a high degree of accuracy. This paper focuses on the development of a system that will track a human's affective state using facial expressions and speech signals with the purpose of modifying the actions of an autonomous agent. The system uses a fusion of two baseline unimodal classifiers based on bayes Net giving rise to a multi-classifier. The union of the three classifiers forms a bimodal scheme of emotion classification. The outputs from the baseline unimodal classifiers are combined together through a probability fusion framework applied in the general multi-classifier. The system classifies six universal basic emotions using audiovisual data extracted from the eNTERFACE05 audiovisual emotion database. The emotional information obtained could provide an agent with the basis for taking an affective decision. It is shown by experimental results that the proposed system can detect emotions with good accuracy achieving the change of the emotional behavior of the agent faced with a human.
Keywords: *Affective Computing, Machine Learning, Adversarial Risk Analysis, Broaden and Build Theory, Facial Expression Recognition, Speech Emotion Recognition, Detection of Emotional Information, Emotional self-regulation.*

1. Introduction

Algorithms that representing or recreating basic human behavior are increasingly sophisticated, allowing machines to perform tasks such as playing a masterpiece on a keyboard instrument, administering drugs to patients, assisting in educational tutoring of schools, imitation of human whole-body motions, to mention just a few. But probably the same tasks affected by emotional variables provided by human beings could be afforded in an entirely different manner. It might be possible that "Emotional Machines" emerge, capable of adapting their behavior to the needs of humans through affective information emitted by them.

Recognition and expression of emotions are essential for an effective communication between humans [1]. What affective computing tries to reach is providing machines with an ability to recognize, model, and interpret human emotions. A machine that expresses, recognizes and understands emotions similar to human ones could be a good collaborator. This is indeed an area of growing interest [2]. Recently, one of the challenges of the affective computing community has been how to conceive brainier, humanlike machines, with the capability to emulate and interpret complex human emotional variables. Even leading authors [3] have raised doubts on whether machines can exhibit intelligent behavior without emotion. Thus, the ability to detect emotions in humans will be a huge step for machines, so as to increase the efficiency of intelligent systems.

Handling emotions could allow a machine to constantly adapt to its environment making social interactions with humans. Emotions are essential to thought and intelligent behavior. They play an important role in concern realization, allowing for goals and motivations [4]. Perhaps this is the key to generating genuinely intelligent synthetic agents, interacting properly with humans. Emotions appear to be essential for the autonomy of a machine. Including emotional elements in their architecture makes them more valuable, but this would not be sufficient, the designer should demonstrate that emotions improve the behavior and interaction capabilities of machines.

Social relations among humans are possible because they manage a multiple sensing biological toolbox. To infer a human's emotion, the signals resulting from an affective state are captured by the biological sensors and associated to this state. Sometimes we capture the emotion of an individual only with an image from the face, other times only with the sound of the voice or the body poses, all of them within the non-verbal communication. The multi-channel physiological sensing machinery of humans allow discovery of the underlying emotion; we often use information from two or more sources provided by the human senses. Many studies underline the relevance of multi-sensory information for judging someone's affective states [5]. In non-verbal communication, humans use facial expressions and speech to reveal emotions, e.g., intonations in speaker's voice quality such as tone, pitch, loudness, duration, speech clarity, intonation, tempo and accent can vary simultaneously with variety of facial gestures to convey the emotional state. We notice that the coexistence between these two emotional sources provide a wide array of mutual information, the data collected through different sources are combined in a complex thought process. The research on automatic emotion recognition has been focused on different emotional data sources provided by humans. In unimodal way, sources from speech in order to decode the emotional states have been studied in [6] facial expressions to decode the current emotion [7], [8], the kinetics of postures [9], physiological brain signals [10]. In the case of multimodal emotion recognition, the advantages are performance and robustness that may be because the data provided by various modalities are consolidated in only one source, giving rise to a classification placed among the best [11], [12]. Other studies have demonstrated that accuracy of the approach resides in fusion techniques that have applied at feature-level and decision-level [13]. The multimodal approach normally handles a different set of features and it implies effects over the training process, normally the emotions are treated as an aftereffect of the features' behaviors. Finding a way to reduce the informative features and selecting only those that are related to the emotion, will be imperative [14].

The literature shows that automatic emotion recognition within the framework of affective computing covers multiple techniques that are useful for validating emotion in humans. It is a complex problem, many individual efforts have been made to resolve this issue, it may still be possible to produce accurate results using other strategies and improve on the mixture of facial expressions and acoustic information. Nonetheless, a big amount of research related the integration of emotional information has been already placed squarely upon the table, however, few attempts to use two techniques in a bimodal system in order to modify a decision in an autonomous agent. The methodology developed in this paper addresses the fusion of information provided by emotional cues from an individual through facial gestures and speech signals. The information to be fused comes from two unimodal classifiers through the use of ensemble probabilistic techniques, in order to obtain an accurate emotion to support the decision-making process of an autonomous agent, capable of interacting with an individual taking into account the emotional feedback loop.

Support Vector Machine, k-Nearest Neighbors, Multilayer Perceptron, bayesNet and decision tree were tested in emotion recognition tasks of audio-visual cues in isolated groups of tests as baseline unimodal classifiers, for the purpose of choosing the best baseline unimodal classifiers applied to cues in speech and facial gestures. Taking into account the performance, the system uses fusion of the data from two baseline unimodal classifiers built on an emotional multi-classifier. The multi-classifier proposed used three types of combination rules at the output: Maximum Likelihood (ML), Probability-Based Product Rule and weighted average probability, being the best the Probability-Based Product Rule. The output resulting of the fusion scheme will contribute to a future model that supports the decision making process of a decision agent.

This paper is organized as follows. Section 2 presents a literature review related to emotions in machines, recognition of emotion from emotional cues (facial expression and speech) and recognition of emotion from audio-visual sources. Section 3 describes the overall statement of the problem within the general scenario where the interactions between the agent and human are performed. Section 4 describes the data set used in the research, the features extracted to represent the emotions from human speech and facial gestures, the machine learning techniques to perform the emotion classification experiments and the measures to evaluate the performance of unimodal classifiers and multi-classifier. Section 5 describes the overall system architecture with the three main parts: Bimodal System, Decision-making system and Self-regulation of Emotional Feedback. Section 6 describes the experimental results of simulations from the overall system. Conclusions and future work are presented in Section 7.

2. Related Work

2.1 Emotion in AI-driven machines

At the beginning of the 90's the vision of believable machinery capable of understanding and displaying emotions, was placed on the table [15]. Humans are emotionally bonded, they use emotions to learn of, alter and anticipate the future feeling of others. In light of this, endowing machines with the ability to recognize and generate synthetic emotions could contribute to achieving a fluent communication with humans. In order to achieve this goal, "Affective Computing" [16], also called emotional artificial intelligence, arises as the study and development of systems and devices that can identify, decipher, measure, and simulate human affective states. Recently, attempts have been made to enhance the relationships in human-machine interaction, specifically, simulating and interpreting emotions [17], [18]. The automatic decryption and synthesis of human emotional states is aimed at producing machines which are able to recognize, emulate and communicate emotions, as it may be a crucial step towards to a new framework of naturally interacting and emotionally with machines [1]. Such is the case, e.g., of the theory of anticipated pleasure named "Decision Affect Theory" [19] which provides empirical evidence of the effect of expectations on emotion generation, showing the link between emotion and decision-making. Consequently, emotions could affect a decision support system in a machine in a similar way as in humans. Extensive models such as TAME [20] create demand for research connecting different theories and their effects on emotions and decision making, as we shall do. Many more examples can be seen in: FLAME [21], EMA [22], Decision Field Theory [23] or AiSoy 1 [24].

We address the problem of emotion interpretation by machines to enhance their social communication skills, knowing that emotion is one of the main drivers behind the choices and decisions in humans. Interpretation of emotions is likely an ingredient of intelligent behavior, both individually and in relation with other intelligent entities, enabling them to handle information detected from individuals, reinforcing their social behavior. It is essential that in a context of social interaction with human beings, machines should be life-like and believable, showing a behavioral consistency and acquiring ways of expressing their internal states and perceive the status of others. A critical success factor in interaction with humans could be the ability of a machine to gather and process signals so as to recognize the situation at hand as a basis for decisions be they in their own interests or share tasks to accomplish a collective goal [25], [26].

2.2 Recognition of emotion from facial expression

To identify emotions by facial expressions is a normal task for humans. The human face has many indicators that conduct dynamic information about various subtle emotional cues. The development of machines capable of reproducing the vision sense in face perception has benefited areas as varied as computer science, cognitive science, mathematics, physics, psychology and neurobiology. The physical ability to display emotions on a human like face is both an important and necessary step in making machines more accessible to the general public. We see clearly that the face is the window where the emotions are showed. In a ground breaking work, the underlying concept of facial expression recognition was developed in [27], with an early attempt to automatically analyze facial expressions by tracking the motion spots on an image sequence. There exists a correlation between all the emotions that are experienced and expressed in a spontaneous manner [28], e.g., the expressions of the face are clear manifestations of the intensity of an involuntary emotion, without previous planning or intention.

Exploring new ways in which machines can understand the defined meaning of gestures and putting them into the context of the feelings, with which they are expressed, should be the key in a closed loop of communication with humans. Following the same line, several approaches to decrypt emotions from face have been reported, these approaches have tried to tackle the group of basic emotions based on simple static models with successful results [29]. This group of studies also analyzes facial emotional expression from diverse physical zones, in tune with mechanical movements of the face muscles. As for decryption of more subtle emotional expressions that are carried out over the face, it has been found that dynamic information has been important for detecting emotions, thus the analysis is based on more natural sequences of facial expressions rather than the isolated captures usually depicted in early databases [30], [31]. The emotion detection applied over natural sequences is more difficult to achieve than when is applied to isolated expressions [32].

Other roads to categorizing information from facial expressions are based in explicit coding systems related to their dynamics, the face movements are coded in set of action units (AUs) with a muscular basis named Facial Action Coding System (FACS) [33]. Several authors have worked taking into account the dynamics of changing faces, the automatic captures were covered from action units from facial motion sequences [34], [35].

The issue of face tracking has also been the topic of several face-based emotion recognition systems, many methods include measurement of optical flow [36], active contour models [37], face identification, recovery of face pose and facial expression [38], probabilistic approach to detecting and tracking human faces [39], active [40] and adaptive [41] appearance model, multiple eigenspaces-based methods [42], tracking by a robust appearance filter [43] and facial motion extraction based on optical flow [44]. The machine learning framework shows several classifiers used in several tasks related to facial expression recognition. Each classifier has advantages and disadvantages in order to deal with the recognition problem, such as Support Vector Machines [45], Bayesian Network Classifiers [46], Linear Discriminant Analysis [47], Hidden Markov Models [48], Neural Networks [49], more detail can be shown in [50]. We could highlight some approaches within the facial expression of emotion; parametric models to extract the shape and movements of the mouth, eye and eyebrows developed in [51], major directions of specific facial muscles were the input in an emotion recognition system treated in [7], permanent and transient facial features such as lip, nasolabial furrow and wrinkles, were taken as recurrent indicators of emotions [52]. We can also point out geometrical models that use heavily rectangular areas, which containing the appropriate muscles of the face and are very useful for synthesizing facial expressions. The presence of the features and their geometrical relationship with each other appears to be more important than the details [53]. Nonetheless, the common denominator in facial emotion detection is that it is always initiated with a detection of face zone, extraction and tracking of relevant facial information and finally the facial expression classification, then with all of this information the facial expressions are analyzed to estimate emotion-related activities.

2.3 Recognition of emotion from speech cues

Voice quality analysis from a phonetic perspective [54] and linguistic features show the strong correlation between voice, pronounced words and emotions [55], e.g., different levels of voice could be depicted by neutral, whispery, breathy, creaky, and harsh or falsetto voice. The emotional features located in chains of words are depicted by the affective states associated with each specific word; many of them are related to the probability of one emotion within a certain sequence of words [56]. The emotional decryption of human speech has been a challenging research issue for many years and one with growing importance with many publications on the topic [57]. A growing scientific field in which it is crucial having a robust speech emotion recognition system is that which comprises medical robotics, because of the emotional knowledge that the robots could learn and acquire in social environments and take care of human beings [58]. Research in acoustics has found connections between emotional cues provided through dialogue, being a potential value to learn about human emotions [59]. Theoretically after the decrypting of emotional cues, it is possible to recognize the emotions present in speech utterances, bearing in mind that the emotional content of speech is not influenced by speaker or the lexical content. The emotional decryption of speech has a multitude of different features that have been used in much research, but a rule to follow in order to manage the multiple features has not yet been established. Studies based on psychological and neurobiological factors have discovered how prosodic cues as fundamental frequencies and the intensity of the voice can show variable levels across different speakers [60], [61]. Emotional indicators are located along the speech chain in short-term spectral features and sound quality cues [62]. Prosodic features, like pitch, loudness, speaking rate, durations, pause and rhythm have strong correlations between them, providing valuable emotional information, they have been viewed as the most informative group that supply many emotional recognition cues [63]. Taking into account the analysis of the entire segment of voice, statistical functions like mean, median, minimum, maximum, standard deviation, or more seldom third or fourth standardized moment are applied to the fundamental frequency (F0) base contour [64], [65]. Human speech contains spectral features such as Mel Frequency Cepstral Coefficients (MFCCs), generally used in speech recognition with great accuracy in emotion detection [66]. Features as Predictive Cepstral Coefficients (LPCC) or Mel Filter Bank (MFB) have been commonly used to classify and depict emotions [67]. Parameterization techniques as MFCCs and RASTA-PLP (Relative Spectral Transform-Perceptual Linear Prediction) work with the same performance and they are commonly used for extracting features from speech by measuring the auditory frequency selectivity [68].

2.4 Emotion recognition using audio-visual cues

The human being uses various sources to acquire knowledge about emotion experienced within the social context as a result of interaction, e.g., within non-verbal communication, the mix of sounds and images can be emotional cues. The information fusion that humans collect between face gestures and the tone of voice may contain indicators about the emotion depicted at the precise moment, consciously or unconsciously, carrying specific messages [69]. The visual markers of the face could be influenced by emotional indicators in speech; this aspect would ensure a global impression formed about the emotion communicated between sensory modalities. In interpersonal communication human speech prosody can augment the interpretation of facial expressions and other social cues [70]. Emotional cues as audio and face images can be concatenated to construct joint feature vectors that are then used to develop an emotional automatic classifier [71]. Some approaches for automatic emotion recognition are improved through unimodal systems. The outputs from the unimodal systems could be fused at the decision-level, increasing the accuracy of emotional detection. The features used were prosodic features as characteristics of the fundamental frequency (F0) [72], [73]. In the case of video frames, differences of movements and geometrical distances between six different zones in a face, the data encrypted in audio-visual features have been used to resolve inaccuracies between the confusing emotion classes. A rule based approach to classify the emotion classes was used for considering one feature at a time, this technique of classification achieved the ability to detect and discriminate finer details [74]. A thermal distribution over a human face was used as additional source in a multimodal system, in which the data from speech, voice and thermal levels in skin were discriminated at decision-level using empirically determined weights. The union of three modalities showed more accuracy that using only two [75]. Differences in audio-visual emotion recognition between spontaneous displays (natural or unintentional) [76], and deliberate displays (posed or intentional) [77] are important to understand of the nature and role of spontaneous human expression in social interaction, spontaneous emotions send signals not included into the criteria of social norms that pursue the goals for communication. To decrypt the emotions in audio-visual data usually used three level of fusion are used; feature-level [78], decision-level [79] and model-level fusion [80]. The first one constructs joint feature vectors with the concatenation of prosodic features and facial features to build an emotional classifier. The second one combines in a total result the outputs coming from each modality. Finally the third one is a hybrid from the first two; it aims to combine the benefits of both the feature-level and decision-level fusion methods.

A wide variety of schemes for audiovisual emotion recognition have been reported as standard methods for related tasks supported by machine learning techniques [64], many of them using the auditory and visual features discriminated by classifiers. Some of the innovative researches that had been joined to the discrimination of emotional cues based on machine learning classifiers are: Bayesian classifiers [81], Decision trees [82], Support Vector Machine (SVM) [11], [83], [84], Artificial Neural Networks [85], [86], [87], K-nearest neighbor [88], Bayesian Networks [89], Hidden Markov Models (HMM) [48], [90], AdaBoost [91], Genetic Algorithms [92].

3. Overall Problem Statement

The goal that we pursue is to have an autonomous agent capable of managing its behavior through information on emotions and interactions provided by humans. The model is constructed over a machine learning framework supported by emotional classifiers and an Adversarial Risk Analysis (ARA) framework. The normal loop of the agent's takes into account three main activities as acquire, process and decide using data from the environment and the adversary. To maintain the normal course of the aforementioned activities the agent needs to be able to interact with the adversary at any time.

The adversary and the agent communicate by exchanging emotions: the adversary displays his current emotional state, e.g., by a specific facial expression or speech signal, these signals will be the emotional inputs recovered by the agent. In a hypothetical case to clarify this point, the adversary could be a person deprived of conventional modes of communication, e.g., an individual with disability of lower limbs but able to produce an emotional prosody with or without semantic information, having a normal motor-sensory development of the upper limbs. The agent will behave according to information of the human adversary and the information of the environment in which the agent evolves. The agent almost continuously forecasts a new action to interact with the adversary. To perform the action, the agent needs to base its decision on the emotional indicators that it has perceived. Part of the model is essentially multi-attribute decision analytic [93], but the agent also manages to forecast models of the evolution of its adversary taking into account the emotional feedback and the environment surrounding all of them. We assume that our agent resides within an environment in which there are other persons with whom it interacts and the possibility of other intelligent systems with which it also interacts in the ways described in Fig. 1, as in the normal course of a human-agent interaction. In order to discover how the system behaves in its different subsystems we develop an isolated group of simulations over the complete emotional feedback decision-making system.

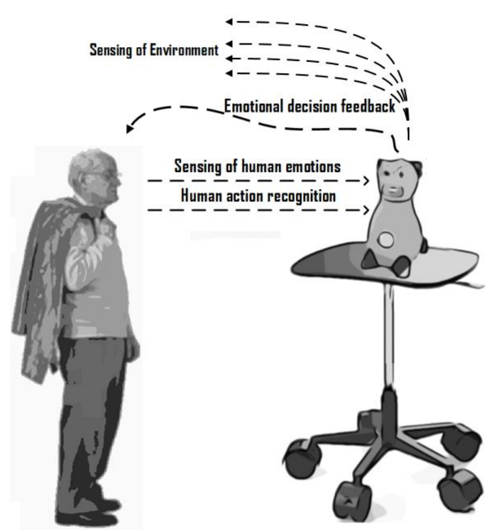

Fig. 1 Overall human-agent interaction.

4. Data and Methods

With the aim of implementing the system in an emotional robot the variable emotional charge that the agent could face and recognize within the possible interaction with humans needs to be investigated. To recreate the human-agent interaction environment, we needed audio-visual information from a real human and the information to feed the ARA (adversarial risk analysis) system (covered by past history of agent's actions, individual's actions and the evolution of the surroundings state in which the interrelation is carried out). Audio-visual information in the form of speech signals and facial expressions from individuals was used for finding emotional cues.

4.1 Dataset

The audio-visual data building the emotional database used in the experiments was collected from eNTERFACE05 audio-visual emotion database [94]. The six universal emotions are portrayed in a set of 1320 videos through five different spoken phrases and facial gestures with emotional connotation produced by 44 non-native English speakers from 14 nations. The research used only one sentence per each emotion which leads to a total of 264 videos. Once the set of videos are decided we proceed to acquire the data from the speech and facial gesture separately. Fig. 2 shows the process applied to each video to build the audio-visual data set.

In the case of facial gestures, to reduce the amount of data and yet retain the dynamics, a group of six static images were captured from each video. The collections of images in different video frames give us an idea of the associate emotion at the time. All the images were subject to the merging process developed in [95] that produces a fusion between the complete zone of the eyes and the mouth over the face zone. The area of the eyes and mouth are very sensitive areas to changes in human expressions and they are particularly relevant for the decoding of emotional expressions. The sets of images are matrices converted to the same dimensions that pass through a series of phases, using the Nearest Neighbor Interpolation method per image. For each individual image a new value is calculated from a neighborhood of samples and replaces these values in the minimized image; this technique is applied to all the images in the dataset. The final process is the addition all the images into a new matrix with the new size of 40 x 30 pixels named matrix-knowledge. The emotional features are depicted in the group of pixels contained in the matrix-knowledge. Each position of the matrix-knowledge (MK) will become a feature. The matrix-knowledge is converted to a row vector of features, whereby each position is a feature, in which the total amount of features will be 1200.

The matrix-knowledge and the vector-knowledge (VN) are observed in Eq. 1 and Eq. 2.

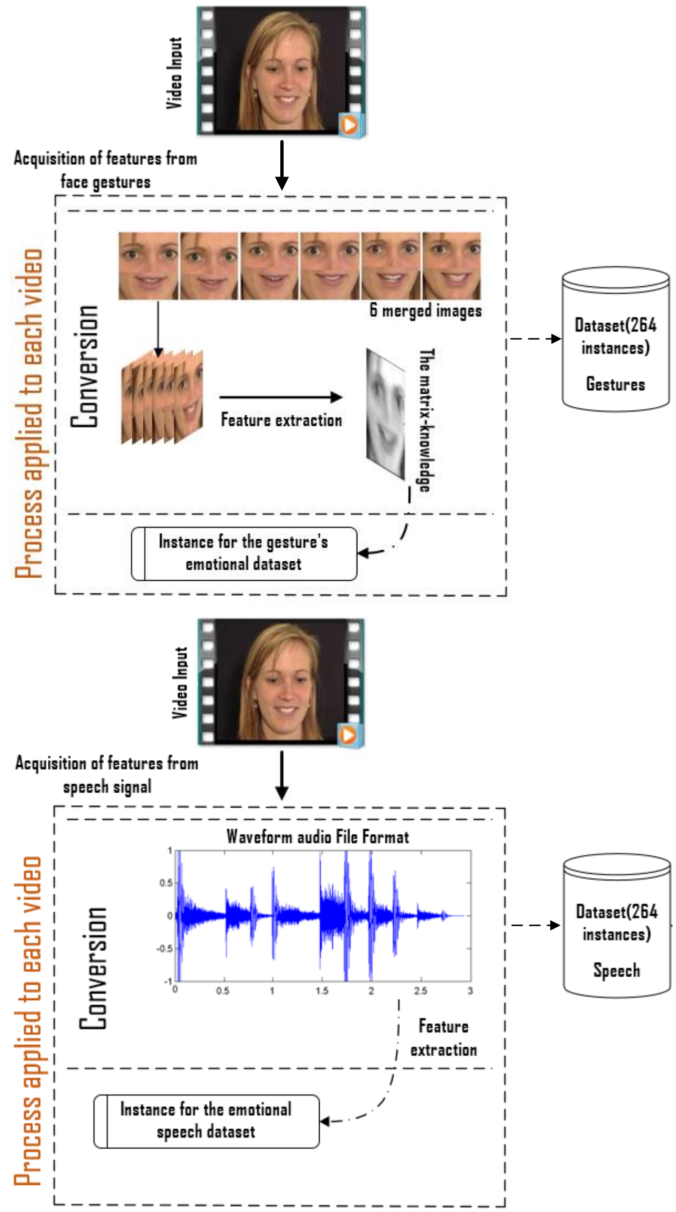

Fig. 2 Dataset construction of audio-visual information.

$$[MK] = \begin{bmatrix} \sum_{x=1}^{6} Im_{x_{11}} & \cdots & \sum_{x=1}^{6} Im_{x_{1n}} \\ \vdots & \cdots & \vdots \\ \sum_{x=1}^{6} Im_{x_{m1}} & \cdots & \sum_{x=1}^{6} Im_{x_{mn}} \end{bmatrix} \quad (1)$$

$$[VN] = [Vec(MK)]^T \quad (2)$$

To recover the information of the speaker's speech, each video from the group of 264 were converted to Waveform Audio File Format. The tool used to read the video in audio files was the MultimediaFileReader object from the DSP System Toolbox Library of MATLAB [96]. The features extracted from the speech signal are commonly used in the multi-disciplinary field of Music Information Retrieval (MIR), covering various tasks related to different music representation media. We know that the variability of emotions can be explained by a small set of acoustic features useful in the emotion estimation context, music sometimes exhibits a vast diversity of emotions that can be represented in some special features, and using the features applied in MIR we can classify some specific mix of cues that provide information about the speaker's mood [97]. MATLAB was used to identify the acoustic features, most of them included in [98]. The features extracted are: Spectral Flux (SF) feature, Spectral Centroid (SC), 2D Method of Moments of MFCCs, Root Mean Square (RMS), Spectral Centroid Variability (SCV), Zero Crossing rate (ZCR), Compactness, Mel-Frequency Cepstral Coefficients (MFCCs), Method of Moments, Strongest Frequency via FFT Maximum, Spectral Roll off Point, Strongest Frequency Via Zero Crossings, 2D Method of Moments of MFCCs, Fraction of Low Energy frame, Linear Prediction Cepstral Coefficients (LPCC) and Strongest Frequency Via Spectral Centroid. Some of the features have been subjected to linear transformations to build a final feature vector that contains 276 attributes, that will be evaluated by the classifiers, see more detail in [99].

4.2 Classifiers Methods

Many methods have been developed for classification tasks. Here we will review general purpose methods provided by WEKA [100] that we applied to classify emotional cues, located in facial expressions and speech signals.

4.2.1 *Support Vector Machine (SVM)*

SVM is a supervised learning method that can be applied to classification or regression, which originated in statistical learning theory. SVM is an extension to nonlinear models of the generalized portrait algorithm developed in [101], offering robust classification to a very large number of variables and small samples. SVM is capable to learn complex data from classification models applying mathematical principles to avoid overfitting. There are number of kernels that can be used in Support Vector Machines models. These include linear, polynomial, radial basis function (RBF) and sigmoid, being the most used the first two.

4.2.2 *k-Nearest Neighbors (kNN)*

k-Nearest Neighbors (kNN) is one of the simplest of classification algorithms available for supervised learning used in statistical estimation and pattern recognition from earliest 70's as a non-parametric technique. The kNN stores all available cases and classifies new cases based on a similarity measure (e.g., Euclidean, weighted cosine distance, etc.). Also, it can be explained as a lazy learning method that searches the closest unlabeled examples of the test data in the feature space, based on distance function [102].

4.2.3 Multilayer perceptron (MLP)

Within the connectionist techniques is also found the Artificial Neural Network (ANN). Multilayer Perceptron (MLP) is the feed-forward ANN more used to classifications. The most used training algorithm to MLP is Backpropagation [103]. The learning process follows two steps, the first is a forward processing of input data through the neurons that produces a forecasted output, the second is the adjustment of weights within the neuron layers to minimize the errors of the forecasted solution compared with the correct output.

4.2.4 *bayesNet (Bayesian Network)*

BayesNet (Bayesian Network) is a graphical model (GMs) for probabilistic relationships among a set of variables, it is used to represent knowledge the uncertainty [104]. The graph depicted in bayesNet is composed by nodes that represent propositional variables of interest and links that represent probabilistic dependencies among the corresponding random variables. Per each node there is a probability table specifying the conditional distribution of the variable given the values of its parents in the network. The network supports the computation of the probabilities of any subset of variables given evidence about any other subset. These conditional dependencies in the graph are generally calculated by using known statistical and computational methods.

4.2.5 *Decision tree*

The Decision tree is a tool for classification and prediction that makes possible the representation of rules, which represent the information in a tree based in a set of features. A classic decision tree is named ID3 (Iterative Dichotomiser 3) is a basic technique to construct a decision tree based on information gain/entropy, it is established on growing and pruning [105]. Other top–down decision trees inducers for continuous values is C45 [106].

C45 is named as J48 in WEKA [100] and it uses the information gain as measure to select and split the nodes.

4.3 Performance measures

Evaluation measures play an important role in machine learning in order to evaluate the performance of classifiers, which are principally focused on handling two-class learning problems. This research has faced a classification problem of six classes formed by six universal emotions.

4.3.1 *Confusion matrix*

The confusion matrix is a measure that contains predicted and actual information about the classification done by a classification system. The vast quantity of measures used in the performance evaluation of binary problems could also apply to multi-class problem. The performance level of classifiers can be assessed with a $m \times m$ confusion matrix in a problem with m classes, as depicted in Table 1. The rows that describe the matrix show the actual classes, while the columns are the predicted classes.

Table 1: Confusion Matrix

	$Predicted\ Class_1$	$\cdots$	$Predicted\ Class_m$
$True\ \ Class_1$	CM_{11}	$\cdots$	CM_{1m}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$True\ Class_m$	CM_{m1}	$\cdots$	CM_{mm}

The level of accuracy in the confusion matrix is computed as a sum of the main values from the diagonal of the matrix, representing the correctly classified cases divided by the total number of instances in the dataset as shown in Eq. 3

$$Accuracy = \frac{\sum_{i=1}^{m} CM_{ii}}{\sum_{i=1}^{m}\sum_{j=1}^{m} CM_{ij}} \quad (3)$$

where CM_{ij} represents the elements in the row i and column j of the confusion matrix.

4.4 Information retrieval measures

Some measures like accuracy do not represent the reality of the number of cases correctly classified per each class. In order to make a deeper analysis, the measure of recall has been calculated for each class. Recall of a class i refers to the percentage of correctly classified cases based on the overall number of cases within class i. Eq. 4 represents the recall for class [107].

$$Recall_i = \frac{CM_{ii}}{\sum_{j=1}^{m} CM_{ij}} \quad (4)$$

Precision or positive predicted value, it denotes the proportion of predicted positive cases that are correctly classified. Precision is a measure of accuracy of predicted positives cases in contrast with rate of discovery of real positives. This measure assesses the predictive power of the algorithm. The precision for class is calculated using the equation Eq. 5.

$$Precision_i = \frac{CM_{ii}}{\sum_{j=1}^{m} C_{ji}} \quad (5)$$

F-measure also F-score or F1 score, it is defined as the harmonic mean of precision and recall. It is null whenever one of the two indices is null, the F-Measure increases proportionally when the values of precision and Recall increases. A high value of F-Measure indicates that the model performs better on the positive class. The F-measure for class formula is represented in Eq. 6

$$F_i = 2\frac{precision_i * recall_i}{precision_i + recall_i} \quad (6)$$

4.1.1 *k-fold cross-validation*

A *k*-fold cross-validation with $k = 10$ was used to make validations over the classifiers. This technique allowed the evaluation of the model facing an unknown dataset. The group of data is randomly divided in k equal parts, one part of the group is used as a validation set and the rest $k-1$ will be the training set. The process is repeated k times using a different group as a validation set, this process continues until each group can used once as validation test. Then, the k results obtained by groups can be averaged to a single result. The advantage of 10-fold cross-validation is that all examples of the database are used for both, training and testing stages [108].

4.5 Multi-classifiers

A multi-classifier is an ensemble of individual classifiers to build a consistently more accurate classifier than a single classifier. The combination of outputs is divided in two: selection and fusion. Selection technique uses the output of the "best classifier" as the general output of the ensemble, whereas fusion technique uses and strategy to mix the different results. The literature describes a big amount of ways of fusion of multi-classifier models [109]. The multi-classifier system needs a base of classifiers with different decision boundaries, in order to handle diverse classifiers with different misclassified regions. The diversity of outputs of multi-classifier is a vital requisite to reach its achievement. It is possible to achieve diversity classifiers through different ways. One approach is to use different training datasets to train individual classifiers, requiring the performance of each classifier. Bootstrap aggregating [110] and Boosting [111] are the most well-known in this category. The diversity can also be obtained by using different base classifiers as Vote [112], Stacking [113] and Mixture of experts [114]. On the other hand, the use of different features in the data training is another method to achieve diverse base classifiers, as the random subspace method [115] and more recently [116]. In this paper we use a multi-classifier base on different features to build the base data training.

5. Overall system architecture

We shall focus on the activities of emotional support and companionship that a social companion agent can carry out. The system works through social interactions taking into account the emotional information provided by human beings. In such a way, through the proposed system, we expect to be able to offer an alternative model for caregiving based on the principles generated by cognitive sciences where social interactions have to be an active process, oriented towards the acquisition of emotional knowledge from humans. To accomplish this goal, we developed three-layer architecture. This distribution will allow organizing the system which integrates the decision making planning of the agent and its emotional acquisition capabilities as shown in Fig. 3.

5.1 Bimodal System

The agent could be endowed with a sensorimotor system (say with vision, audition, touch sensors, temperature sensor, inclination sensor and so on), capturing information about psychophysiological indicators of an individual. Information on facial gestures and speech cues is passed on to three classifiers, two of them are unimodal and the other is a multi-classifier. The stage of unimodal classifiers is responsible for the selection of the audio-visual information cues. We use self-dependent feature vectors as data format (each one from different source) for implementing different types of classification methods. To simulate the bimodal classification stage we conducted several classification tests, using the WEKA toolbox [100].

In a pilot phase, each unimodal classifier was evaluated separately in a group of preliminary studies [99], [117], in order to select the best group of features that served to predict six universal emotions from an audiovisual source.

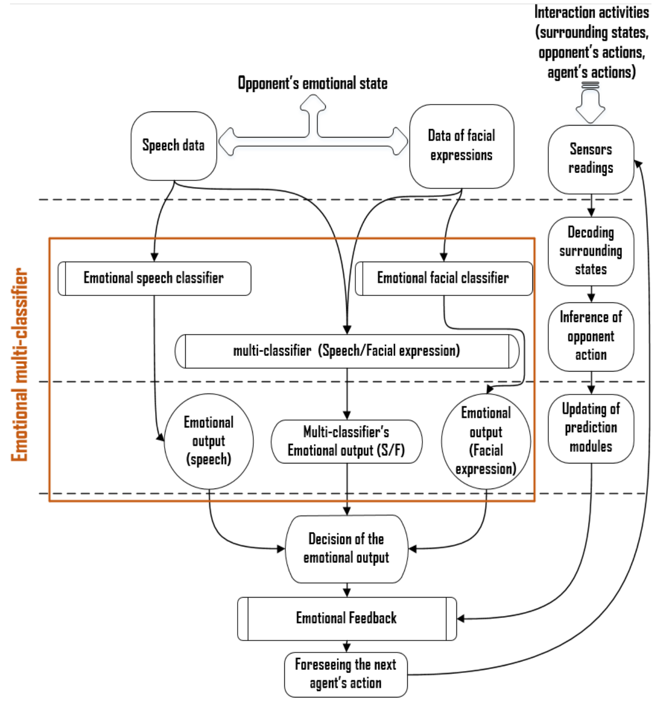

Fig. 3 Overall system architecture.

Given the respective selection of emotional cues, we decided about the best classifiers that gave accuracy results so that they can be used later as inputs to the multi-classifier. As can be seen in Fig. 3, the multi-classifier collects the information separately, the audio features go directly to the emotional speech classifier and the facial features to the emotional facial classifier. The multi-classifier uses the same database with different features to train the baseline unimodal classifiers. Further the multi-classifier fuses the outputs from the baseline unimodal classifiers based on the continuous output of them. In this research we used three matching algorithms from the Bayesian probability fusion framework: Maximum Likelihood (ML), Probability-Based Product Rule and weighted average probability. To this task, we implemented in Java a new multi-classifier with the probability-based fusion methods directly according to the WEKA interface. In a real case, an agent can face problems when capturing affective information through the sensors (e.g. because the robot lacks of the specific sensor to extract the emotional feature or it is within a noisy environment). The emotional features are not necessarily collected at the same time or have a high level of robustness. To deal with this problem the system has a

IJCSI International Journal of Computer Science Issues, Volume 12, Issue 6, November 2015
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

decision stage in which it decided how to combine its member outputs to feed the next level “The emotional feedback”.

5.2 Decision-making system

The right hand side of Figure 3 shows the interactive decision model. In this stage the agent captures information about its surrounding state and the individual, which is interpreted to infer the individual actions and surrounding. Such information is, on the one hand, evaluated to find out the impact on the agent’s objectives, and learn the evolution of the surroundings shared with the individual and to make forecasts. Forecasts and evaluations will be influenced by emotions and will have an impact on its decision making process, called emotional feedback. The prediction modules of the agent assess probabilities when the opponent (individual) has performed an action that triggers surrounding reactions, given the past history of agent’s actions, opponent’s actions and the evolution of the surroundings state. The simulation is implemented taking into account surrounding states such as Energy, Temperature, Position and Detection. Based on ARA framework [118] the system manages two modules, one represents the opponent’s own evolution and is called *opponent’s evolution module* and the other represents the opponent’s reactions to the agent’s actions and is called *classical conditioning module*.

The actualization of the data is based on the average of the posterior distributions, using the opponent’s evolution module and the classical conditioning module [119]. Assuming that $p(M_i)$ designates the probability that the agent follow the model i, with $p(M_1)+p(M_2)=1$, $p(M_i)\geq 0$, this model captures the agent’s reactive behavior facing the opponent’s actions subsequently influenced by the emotional feedback, in which $p(M_1)$ represents the individual’s own evolution and $p(M_2)$ represents the individual’s reactions to the agent’s actions. The agent’s behavior provides utilities and probabilities to select the optimal action based on all current information based on the Maximum Expected Utility (MEU) [120]. In order to compute the maximum expected utility, we plan one period ahead assuming additive utilities over time. The decision is probabilistic, based on a probability model that randomly generates alternatives with probability proportional to the expected utility [121], therefore, it increases the uncertain of the agent's decisions under similar circumstances. The agent aims at satisfying some hierarchically ordered self-objectives in order to support its own evolution [122]. Once it has reached the sufficient value in that level, it will redistribute its own resources to achieve the next level, and so on, e.g., if the agent needs to socialize with the opponent, it will carry out a salute action.

The hierarchy levels with which it evolves could have a high level of social connotation, in this case the simulations were developed with two levels “Energy and Security”, e.g., for the Security objective, the agent takes into account the level of injury (Atk) by a human (a negative emotional feedback by the part of caregiver’s patient), at the same time the agent verifies the appropriate level of temperature (Tp) to operate without problem. All self-objectives are constructed based on component utility functions that will evaluate in a Global utility function. To simulate the process, we have developed a simulator in python, as shown in Fig. 4, in which the agent makes decisions according to the process in the proposed system, for more detail see [24], [123].

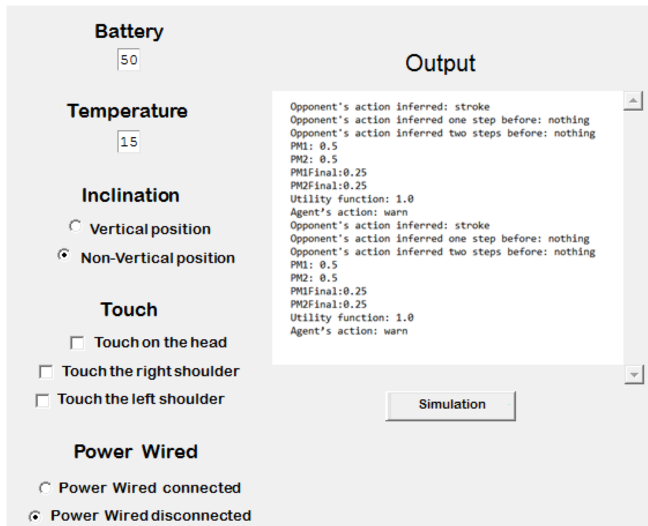

Fig. 4 Overall system architecture.

5.3 Self-regulation of Emotional Feedback

An early work was focused on positive evolution of an agent’s behavior, using the basis of the *broaden and build theory* [124]. An interaction with a cycle of positive emotions connected to the relevant actions is intended. In other words, a positive symbiosis between the individual’s emotions and the actions of an agent in charge of emotional caregiving. Some negative emotions tend to favour some preferences [125], e.g., anger tends to maintain a course of action or in the most extreme case attack, while fear can evade the immediate context or the course of action. To improve the interaction between agent and individual the system needs to have the support of auto regulation. The effects of *broaden and build theory* allow the agent’s decisions to evolve, whereas the *Emotional* *self-regulation* [126] will able to achieve more humanized decisions in the agent, as well as the ability to make spontaneous decisions with emotional connotation. For this research we used a modification of the rule-based algorithm developed in [127]. The new algorithm simply corrects the agent's behavior using the human biological emotions. According to the model of emotion-regulation developed in [128], an individual can handle the emotional charge based on the following points: situation selection, situation modification, attentional deployment, cognitive change and response modulation. Each step in the process is a trigger for regulating. Each level has consequences as a result of different emotion regulation forms. Taking the concept a step further as the emotional feedback, the emotion-regulation could affect the emotional decision process at different stages, making the simulation of the decision making behavior of the agent more realistic. Focusing on the agent's continuously made decisions, the period of time in which an action (e.g., *cry(-)* taking into account the sign) is kept constant could be critical. Because, the agent can maintain the same decision for a long time taking into account the risk of getting stuck in an endless loop, affecting some elements of humanity, often unpredictable, that we want to achieve with the agent. More specifically, if the agent has been noticed that one decision remains longer for a long period, with this information the agent can handle the choosing other decision randomly of the repertory of its decisions.

Table 2 shows six basic emotions [129], [130], represented in a set of rule-based action/response-pairs. The categorization that we assumed was arbitrarily chosen, taking into account that it is possible to modify the model with another order of emotions. For this first scheme, we have ordered the emotions according to the "Basic emotion theories", where we can differentiate two groups "positive" and "negative" based on the emotional stimuli organized in a categorical manner, with innate categories found in humans [131]. For more detail about the algorithm and the repertory of rules and decisions, see [127].

Table 2: Emotions, Actions, and Modified Actions

OS_{em}	R_1	R_2	NR_1	NR_2
Anger (-)	(+)	(-)	R_{upL}	R_{upL}
Sadness (-)	(+)	(-)	R_{upL}	R_{upL}
Fear (-)	(+)	(-)	R_{upL}	R_{upL}
Disgust (-)	(+)	(-)	R_{upL}	R_{upL}
Happiness (+)	(+)	(-)	R_{upL}	M_L
Surprised (+)	(+)	(-)	R_{upL}	M_L

Note: $R_a\,\{R_1, R_2\}$, $R_a\,\{(+), (-)\}$ are signs of the agent's actions before the emotional feedback. NR_1 and NR_2 are the modified actions of the agent taking into account the emotional feedback. R_{upL}: Reject and change the hierarchy level of agent's action (to positive action); M_L: Maintain the current hierarchy level of agent's action (maintain the negative action); OS_{em} : Signs of the opponent's emotion.

For the purpose of the simulations, we used only the first two steps of the process model of emotion-regulation, situation selection and situation modification. In situation selection it is taken into account that the performing of an act is intended to analyzing the expectations of emotions that the individual would be likely to have. Strictly speaking, the agent's decision could affect the emotions of the individual. The agent will assume a context in which the individual must not see a particular recurring decision, in this scenario the agent will be even more unpredictable as human behavior. The decision chosen by the agent could be modified to change the emotional future impact in the individual. Fig. 5 shows the flow chart of the agent's action modified by the positive emotion and the influence of the emotion-regulation stage. The counters (each by each decision rule) counts a number "N" (the time period in which an agent's decision has "N" occurrences), according to "N" the agent will change randomly the decision related its repertoire of decisions. Subsequently the agent will continue with the normal course of activities, e.g., an agent's decision counter in the stage of individual's anger emotion will be AD_c as shown in the first decision stage.

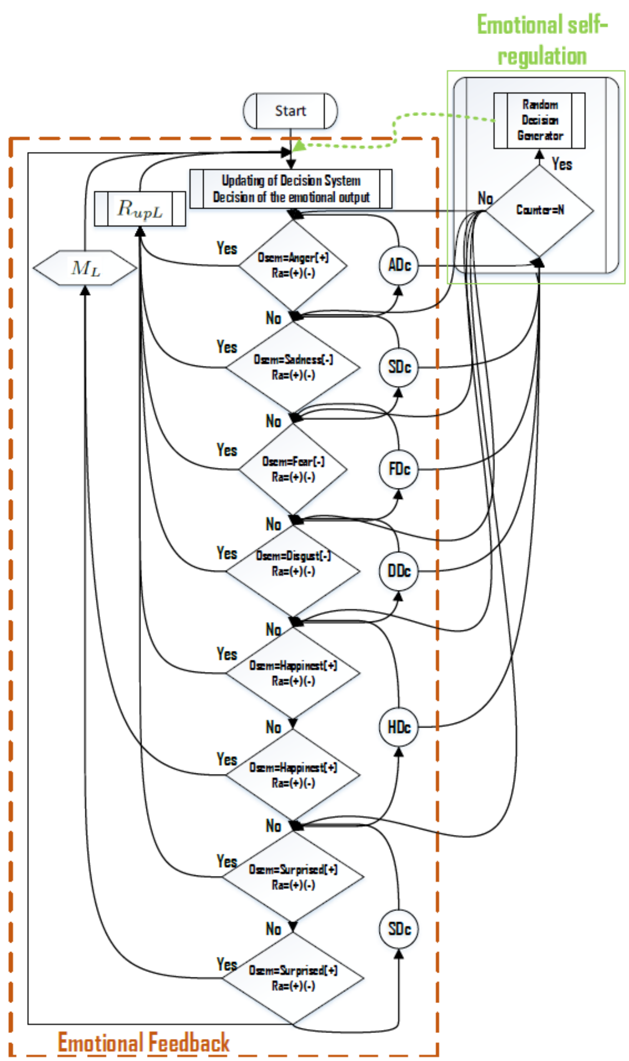

Fig. 5 Self-regulation of Emotional Feedback process.

6. Results

6.1 Bimodal System

We use the information gain criterion to select the features used by the speech classifier, according to the best results reached by a decision tree in [98]. Only a few features were selected, 2D Method of Moments and the 2D Method of Moments of MFCCs from the group of features of Music Information Retrieval (MIR). In the case of facial gestures the information was composed by the vector-knowledge (VN) constructed by 1200 pixels. A 10-fold cross-validation scheme was applied to the overall emotional classification scheme (multi-classifier and the baseline unimodal classifiers). Tests were made in order to select the two baseline unimodal classifiers that in turn are sent to the multi-classifier that will make a fusion between the two sources. Six classifiers were tested, the Support Vector Machine (SVM) has used three kernels, linear and polynomial (with degrees 2 and 3), *k* Nearest Neighbors (kNN) has used *k* from 1 to 15, Multilayer Perceptron (MLP) with hidden neurons from 2 to 20, bayes Net (BN) and decision tree (J48). We used the toolbox of WEKA [100] to implement the group of classification tests. A group of tests were carried out per each case (speech and facial gesture) from which bayesNet has been achieved the best results. The results per each classifier in recall measure are shown in Fig. 6 and Fig. 7, speech and facial gestures respectively. It should be noted that bayesNet shows the best results in any case.

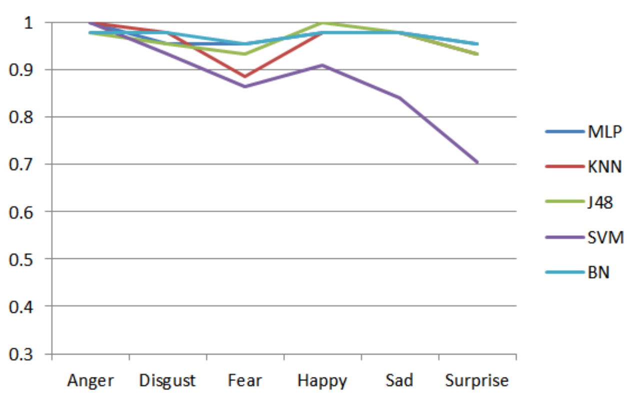

Fig. 6 Recall of different classifiers with feature selection for each emotion in speech.

BayesNet was the best candidate to build the baseline unimodal classifiers used by the multi-classifier. Once the multi-classifier is constructed, we compared the level of accuracy, as shown in Fig. 8, it can be seen clearly that the level of accuracy increased using the multi-classifier in its three levels of fusion compared to the baseline unimodal classifiers (used in speech and facial gestures depicted in the first two bars). Any fusion mode of the multi-classifier is higher than the two baseline unimodal classifiers. The best classification was achieved by the Probability-Based Product Rule fusion.

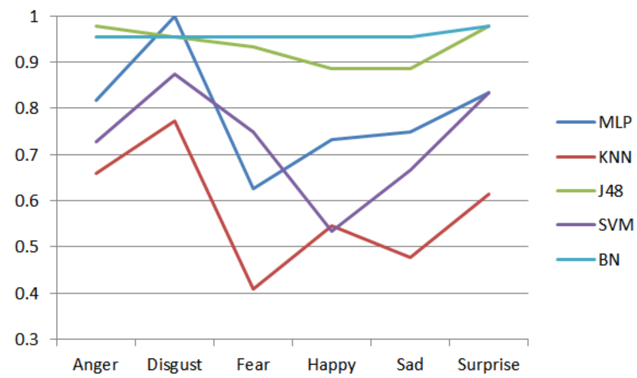

Fig. 7 Recall of different classifiers with feature selection for each emotion in facial gestures

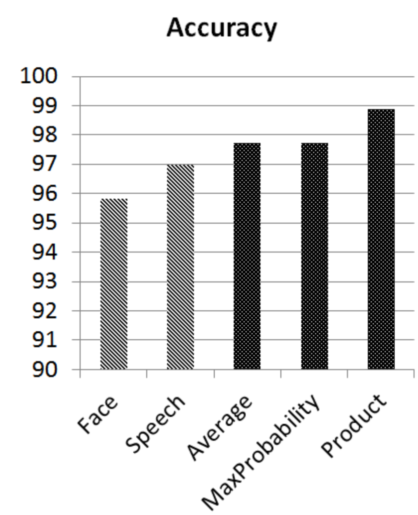

Fig. 8 Comparison of Accuracy in baseline unimodal classifiers and multi-classifier.

Recall and precision provide a detailed view of the results in emotion classification, for this reason we show the results previously depicted as well as the accuracy. Fig. 9 shows the Precision measure, in which all the results are superior to 95%, however, all the results in Probability-Based Product Rule are above 97%, being the best classification in happiness and surprise. Fig. 10 shows the recall measure, in which all of results are above 95.6%. In Probability-Based Product, the results reached the minimum of 97.8%, meaning that all the results are above to 97.8%.

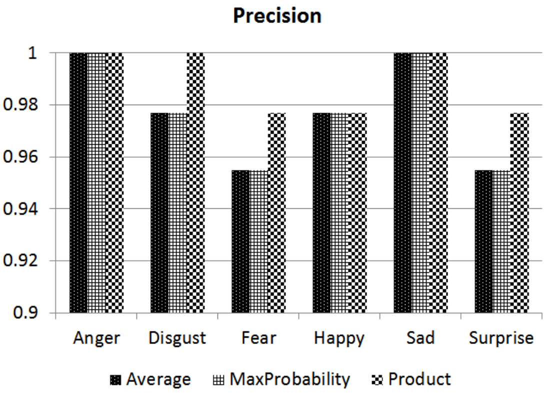

Fig. 9 Comparison of results from Precision per each emotion related to the multi-classifier's outputs (Average, MaxProbability and Product).

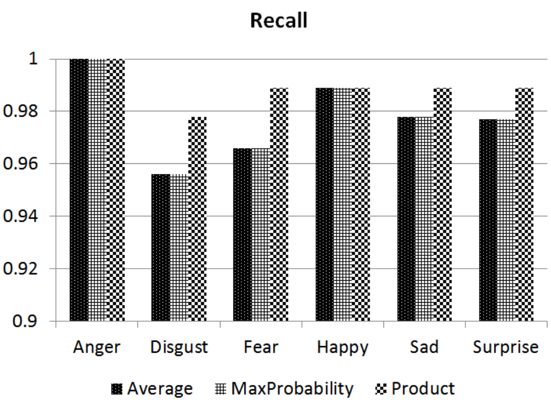

Fig. 10 Comparison of results from Recall per each emotion related to the multi-classifier's outputs (Average, MaxProbability and Product)

In Fig. 11, F-measure (the weighted average of the precision and recall) shows that emotions disgust and surprise not reached the 100%, however, precision and recall measures depict that disgust and surprise have achieved the best results. In short, any of the fusion methods (Maximum Likelihood (ML), Probability-Based Product Rule and weighted average probability.) have reached best results than the classification through the baseline unimodal classifiers. This is the reason we used this method fusion to build the multi-classifier used in the simulations.

6.2 Decision System and Self-regulation of Emotional Feedback

We carried out three different computational experiments in order to show the different behavior caused by the interaction between the individual and the agent, taking into account the Decision-making system and Self-regulation of Emotional Feedback stage working together. The decision-making system used the evolution of the expected utility just one period ahead, the agent's reactive behavior facing the individual's actions through the weighting of posterior model probability, obtaining the probably next action that the agent will select. The self-regulation of emotional feedback stage used the evolution of the emotional feedback provided by the face gestures and the voice from the individual in the bimodal system stage and the evolution of the agent's behavior using the rule-based design of emotional feedback with self-regulation. We selected only six agent's actions classified with negative and positive connotation as *alert(-), cry(-), claim for energy(+), salute(+), warn(+), do nothing(+).* The six actions attack, move, recharge, stroke and do nothing make up the set of individual's actions carried out through the simulator in python during the simulation with the Decision-making system. The individual's actions do not need the signs, because all the analysis will focus on the modification of the agent's behavior facing the individual during the interaction.

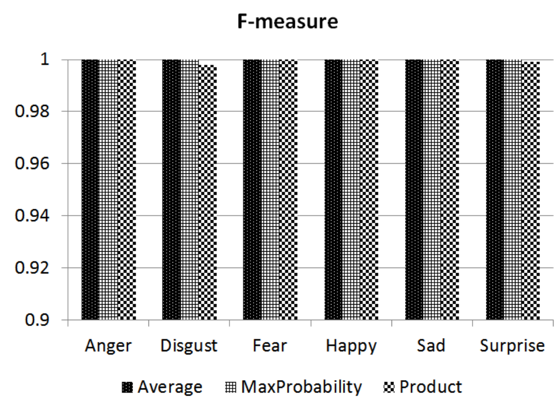

Fig. 11 Comparison of results from F-measure per each emotion related to the multi-classifier's outputs (Average, MaxProbability and Product)

Fig. 12 shows the first simulation, which was made with 44 individuals from the eNTERFACE 05 audio-visual emotion database. Each individual has faced to the agent continuously in order to show six universal emotions which are represented simultaneously, that gives a total amount of 264 cases where the interaction and the emotional classification took place. Taking into account a macro vision, the agent will face diverse individuals in a long period, e.g., the constant interaction would be a permanent training of the individuals' database that the agent has pre-recorded in its memory. This means that in the future the agent could learn about the interaction, adapting its behavior to a specific individual. As we can see, the behavior of the agent is clearly negative in actions as cry and alert before the emotional feedback provided by the individuals. The agent's behavior will suffer changes only in the slots in which it faces positive emotional charge of the individual. The emotional inputs as happiness and surprise displayed by the individuals allowed the agent to be more expressive. This will enable the performing of the agent's actions with negative connotation. The negative component in many actions was suppressed, which is why the interaction in positive way was favored.

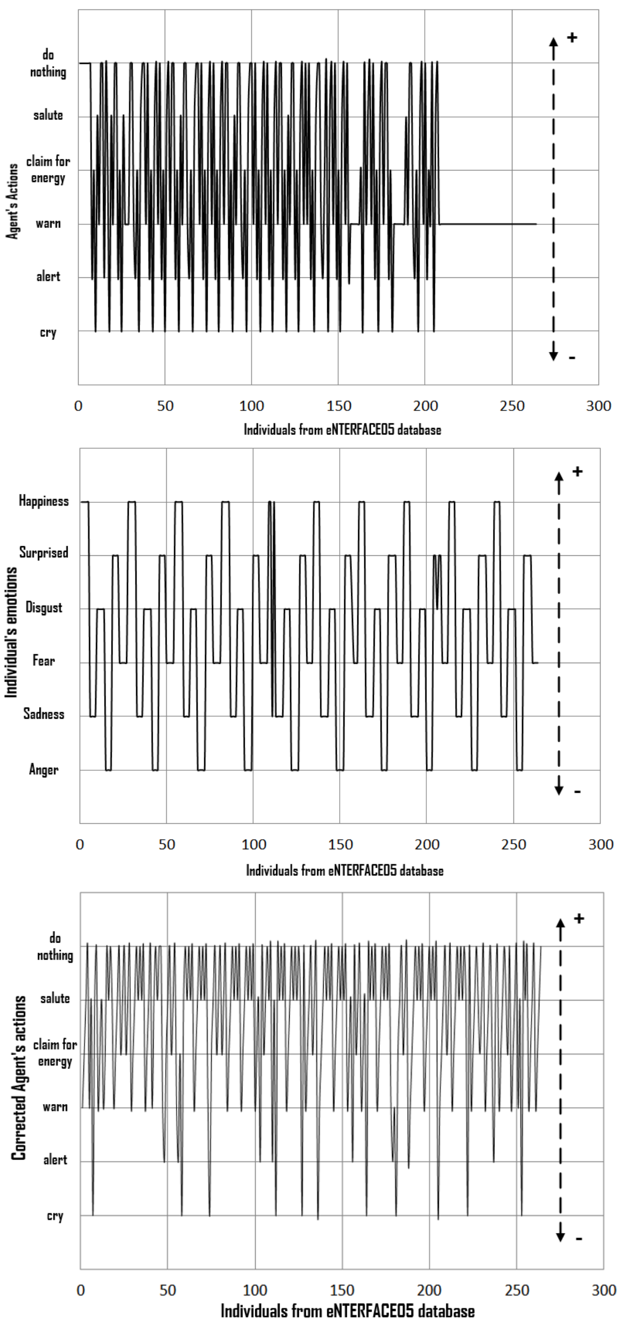

Fig. 12 Interaction test between the agent and all dataset in order

Fig. 13 shows the second simulation, which pursues the same objective, but in this case the agent will face diverse individuals in a random way. To achieve this, the categorized emotions from the individuals are chosen randomly from the eNTERFACE 05 audio-visual emotion database. The simulation aims to create a more realistic human-agent interaction. Here the individuals show a less interaction with the agent and it has particularly been reflected between the individual 22 and 223. The simulation took into account a less interaction using the simulator, only some variables were modified. We should also point out the expected utility of the different simulations varies depending on the interaction of the user, in this case for this long period, the group of individuals haven't shown interactions. After the emotional feedback a great quantity of agent's actions are allowed. Actions like *cry(-)* or *alert(-)* are more permissible when the opponent has a positive emotional charge. At the same time the agent's actions increase and were transferred to the positive zone, this means that after the emotional feedback the simulation shows more interaction between the individual and the agent (a more realistic socialization).

Fig. 14 shows the third simulation that represents an isolated case from the 44 individuals, in order to see with more detail the self-regulation over the agent's forecast decisions within the emotional feedback loop. The data belonging to a specific individual (speech and facial gestures cues) was randomly replicated as inputs to the bimodal system in order to decrypt the current emotion. At the same time, the probability model works with the selection of agent's probabilistic decisions. The counters per each agent's forecast decisions were fixed in 7. As can be appreciated on the figure, many agents' actions have now transferred to the positive zone, the level of interaction increases with the individual in comparison with the agent's action before the emotional feedback. As is reflected in some points of the iteration period, the decisional behavior of the agent changes because of its self-regulation, the agent changed randomly the decision related its repertoire of decisions.

In interactions points such as 7, 64, 71, 78, 85, 92, 99, 120, 127, 134 and 141, address changes were recorded because of self-regulation, the agent has randomly chosen decision alternatives within the positive range (e.g., *salute(+)*, *warn(+)*, *do nothing(+)*). In the case of iterations as 106 and 113, it should be concluded that the random choice has tended to be negative in the two cases, this is evidenced through the selection of the agent's decision with negative connotation during the negative emotional charge from the individual (e.g., *disgust (-)* and *anger (-)* faced *cry(-)*). We can observe in Fig. 15 how the expected utility of the simulation varies depending on the interaction with the user. If the user constantly interacts with the agent, its expected utility decreases (interaction's points between 0-

16 and 123-150) which is the result of the varying behavior of the user. The simulation shows as expected that the agent perceives the individual as very reactive to its actions. The figure also showed as expected that the agent perceives the individual as very reactive to its actions. The value of $p(M_1)$ (individual's own evolution) rapidly achieves the maximum and $p(M_2)$ (the individual's reactions to the agent's actions) the minimum for the two cases.

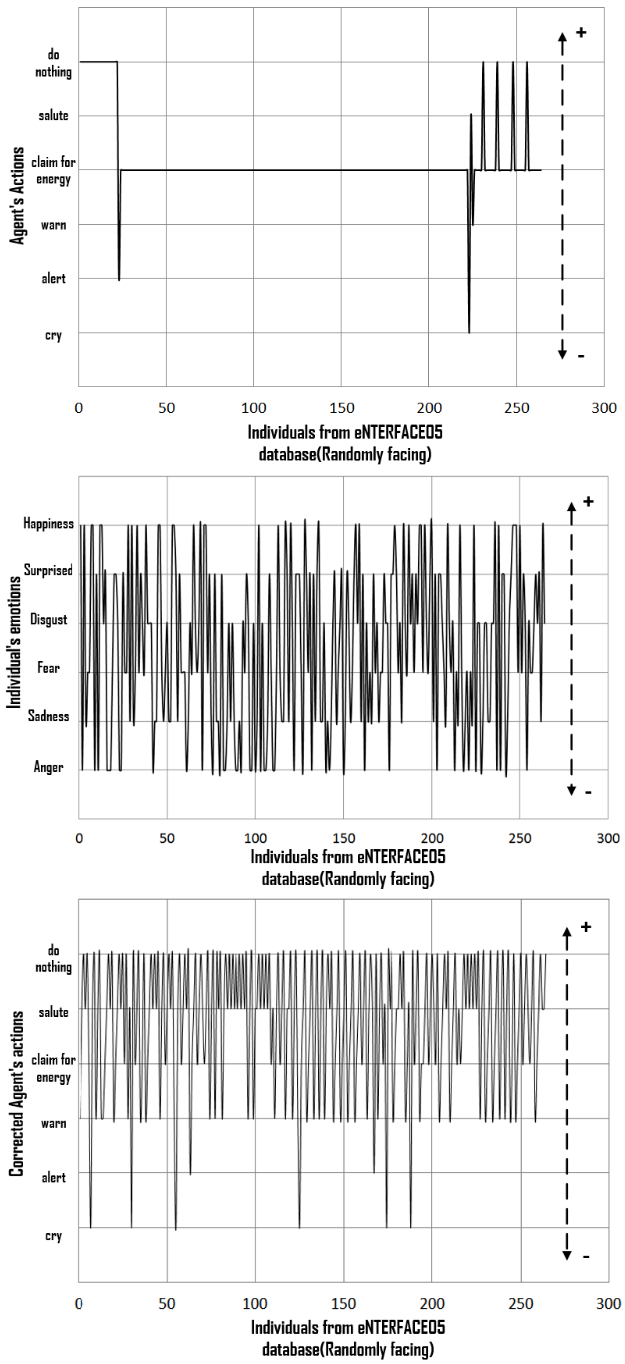

Fig. 13 Interaction test between the agent and all dataset randomly

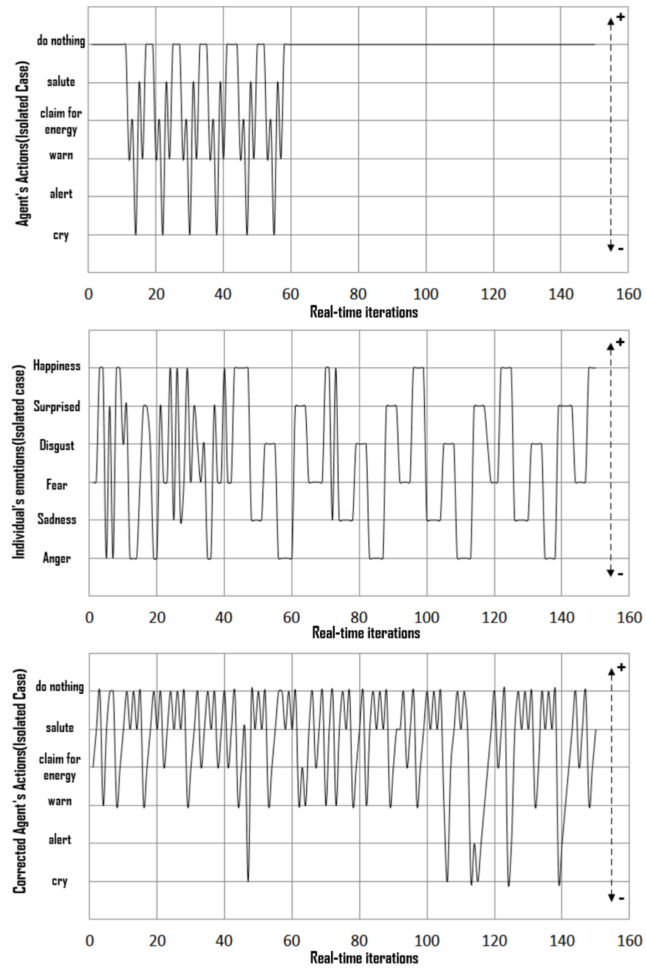

Fig. 14 Interaction test between the agent and one specific individual

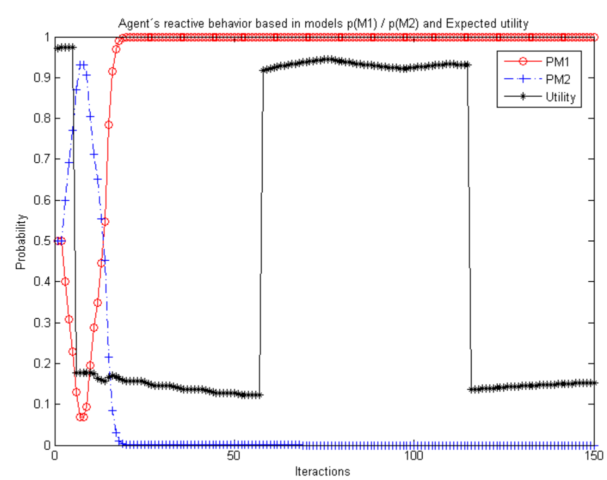

Fig. 15 Expected utility result of the interaction between the agent and the specific individual

7. Conclusions and Future Work

Designing emotional caregiving interfaces is crucial for successful support of humans in targeted rehabilitation schemes (such as assisting the societal participation of persons deprived of conventional modes of communication). In this paper, we proposed a concept of system for an agent in order to appropriately provide emotional support for disabled people. The agent's system has the ability to read information from its sensors and selecting appropriate actions based on the emotional signals expressed by a human. To apply the concept, we studied the learning of the individual's behavioral patterns in a real emotional interaction with an agent, in which we decided the group of agent's decisions pertinent for the individual. The model uses multi-attribute decision analysis, forecasting models of the adversary and emotional feedback provided by the opponent, all supporting the final decision of the agent. The proposed emotional acquisition stage consisted of preprocessing, feature extraction and pattern classification steps. Preprocessing and feature extraction methods were devised so that emotion-specific characteristics could be extracted in an audio-visual scheme. We proposed a bimodal system fuelled by emotional cues provided by the facial expressions and speech from an individual.

There is however a risk that the agent is located in the midst of a noisy environment or with the malfunction of one of its sensors. Regarding the latter, the system can react giving the emotional information of one or two or the fusion of the two sources. For this reason, within the emotional acquisition stage, we proposed a bimodal system composed of two baseline unimodal classifiers and a multi-classifier. To select the best emotion recognition, we compared 6 classification methods within the most used, where BayesNet was chosen because of its best results over the two sources (facial gestures and speech). Taking into account unimodal classifiers with a recall value greater than or equal to 95% for all the set of human emotions, we proposed a multi-classifier supported by two baseline unimodal classifiers (each one trained per one source). The multi-classifier used three methods of output fusion: Maximum Likelihood (ML), Probability-Based Product Rule and weighted average probability. The Probability-Based Product Rule achieved the best results with 98% of accuracy, precision ranges between 97-100% and 97.7-100% in recall. According to the results provided by the Decision System and Self-regulation of Emotional Feedback, the agent is capable of improving its social interaction by changing its behavioral decision state according to the affective state of the individuals that it is facing. We can consider that the final decision of the agent be turned into an affective decision, taking into account that the agent predicts the affective consequences of each available alternative to not disturbing the individual to which it is faced; it means a strengthening of relations between human-machine within an emotional loop. Following this, future work will address testing the model in a real robotic platform that makes decisions influenced by more accurate emotional factors, as the Self-regulation of Emotional Feedback provided by the individual with whom the robot generates empathizing.

Acknowledgments

This research was supported by VINNOVA Swedish Governmental Agency for Innovation Systems through the ICT project The Next Generation (TNG).

References

[1] R. Picard, Affective Computing, United States: the MIT Press, 1998.
[2] T. Ziemke and R. Lowe, "On the role of emotion in embodied cognitive architectures: From organisms to robots", Cognitive computation, Vol. 1, No. 1, 2009, pp. 104-117.
[3] M. Minsky, The Society of Mind: Simon and Schuster, 1985
[4] N.H. Frijda and J. Swagerman, "Can Computers Feel?, Theory and Design of an Emotional System", in Cognition and Emotion, Vol. 1, No. 3, 1987, pp. 235-257.
[5] P.N. Juslin and K.R. Scherer, "Vocal expression of affect", in Harrigan J, Rosenthal R, Scherer K (eds), The new handbook of methods in nonverbal behavior research. Oxford University Press, Oxford, 2005.
[6] J.D. Moore, L. Tian and C. Lai, "Word-level emotion recognition using high-level features", in Computational Linguistics and Intelligent Text Processing, Springer Berlin Heidelberg, pp. 17-31, 2014.
[7] M.N. Islam and C. Loo, "Geometric Feature-Based Facial Emotion Recognition Using Two-Stage Fuzzy Reasoning Model", in Neural Information Processing, Vol. 8835, C. Loo, K. Yap, K. Wong, A. Teoh, and K. Huang, Eds., ed: Springer International Publishing, 2014, pp. 344-351.
[8] S.V. Ioannou, A.T. Raouzaiou, V.A. Tzouvaras, T.P. Mailis, K.C. Karpouzis and S.D. Kollias, "Emotion recognition through facial expression analysis based on a neurofuzzy network", in Neural Networks, 2005, Vol. 18, No, 4, pp. 423-435.
[9] E.M.J. Huis In´t Veld, G.J.M. van Boxtel and B. de Gelder, "The Body Action Coding System II: muscle activations during the perception and expression of emotion. Frontiers in Behavioral Neuroscience", 2014, Vol. 8, pp. 330.
[10] R.W. Picard, E. Vyzas and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state", in IEEE Trans. on Pattern Analysis and Machine Intelligence, 2001, Vol. 23, No. 10, pp. 1175-1191.
[11] S. Zhang, X. Wang, G. Zhang and X. Zhao, "Multimodal Emotion Recognition Integrating Affective Speech with Facial Expression", in WSEAS Transactions on Signal Processing, 2014, Vol. 10.

[12] R. el Kaliouby and P. Robinson, "Generalization of a Vision-Based Computational Model of Mind-Reading" , in Proceedings of the First International Conference on Affective Computing and Intelligent Interfaces, 2005, pp. 582-589.
[13] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C.M Lee, A. Kazemzaeh, S. Lee, U. Neumann and S. Narayanan, "Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal information", in Proceedings of the ACM 6th int'l Conf. on Multimodal Interfaces (ICMI 2004), State College, PA, 2004 pp. 205-211.
[14] M. Pantic, N. Sebe, J.F. Cohn and T Huang, "Affective Multimodal Human-Computer Interaction", in ACM Multimedia, Singapore, 2005, pp. 669-676.
[15] E.T. Mueller. Daydreaming in humans and machines: a computer model of the stream of thought. Intellect Books, 1990.
[16] R. Picard. Affective Computing, United States, The MIT Press, 1998.
[17] K. Williams and C. Breazeal, "Reducing Driver Task Load and Promoting Sociability through an Affective Intelligent Driving Agent (AIDA)", INTERACT, 2013, Vol. 4, pp. 619-626.
[18] C. Breazeal, "Social robots for health applications", in Conference proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2011, pp. 5368-5371.
[19] B.A. Mellers, A. Schwatz, K. Ho and I. Ritov, "Decision affect theory: Emotional reactions to the outcomes of risky option", in Psychological Science, 1997, Vol. 8, no. 6, pp. 423-42.
[20] L. Moshkina, "An integrative framework for affective agent behavior", in Proceedings of the International Conference on Intelligent Systems and Control, IASTED, 2006.
[21] M.S. El-Nasr, J. Yen and T.R. Ioerger, "FLAME Fuzzy logic adaptive model of emotions" in Autonomous Agents and Multi-Agent Systems, 2000, Vol. 3, No. 3, pp. 219-257.
[22] J. Gratch and S. Marsella, "A Domain-independent Framework for Modeling Emotion" in Cognitive Systems Research, 2004, Vol. 5, no. 4, pp. 269-306.
[23] J.R. Busemeyer, E. Dimperio and R.K. Jessup "Integrating emotional processes into decision making models", in W. Gray, Eds. Integrated models of cognitive systems. Oxford University Press, 2007.
[24] J.G. Rázuri, P.G. Esteban and D.R. Insua, "An adversarial risk analysis model for an autonomous imperfect decision agent" in T.V. Guy, M. Kárný and D.H. Wolpert, Eds. Decision Making and Imperfection. SCI. Springer, Heidelberg, 2013, Vol. 474, pp. 165-190
[25] B.R. Duy, "The Social Robot", PhD Thesis, Department of Computer Science, University College Dublin (UCD), Dublin, Ireland, 2000.
[26] R.W. Collier, "Agent Factory: A Framework for the Engineering of Agent-Oriented Applications", Ph.D. Thesis, University College Dublin, Ireland, 2001.
[27] M. Suwa, N. Sugie and K. Fujimora, "A preliminary note on pattern recognition of human emotional expression" in International Joint Conference on Pattern Recognition, 1978, pp. 408-410.
[28] D. Krstic´. Psiholoski reénik, Beograd: Savremena, 1991.
[29] P.G. Enticott, H.A. Kennedy, P.J Johnston, N.J. Rinehart, B.J. Tonge, J.R. Taffe and P.B. Fitzgerald, "Emotion recognition of static and dynamic faces in autism spectrum disorder", in Cognition and Emotion, 2013, Vol. 28. pp. 1110-1118.
[30] M. Pantic and I. Patras, "Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences", in IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2006, Vol. 36, No. 2, pp. 433-449.
[31] B. Raducanu and F. Dornaika, "Dynamic vs. static recognition of facial expressions", in AmI'08: Proceedings of the European Conference on Ambient Intelligence, Berlin, Heidelberg, Springer-Verlag, 2008, pp. 1325.
[32] H. Fang, N.M. Parthalin, A.J. Aubrey, G.K.L. Tam, R. Borgo, P.L. Rosin, P.W. Grant, D. Marshall and M. Chen", Facial expression recognition in dynamic sequences: An integrated approach, Pattern Recognition, 2014, Vol. 47, No. 3. pp. 1271-1281.
[33] P. Lewinski, T.M. den Uyl and C. Butler, "Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader", in Journal of Neuroscience, Psychology, and Economics, 2014, Vol. 7, No. 4, pp. 227-236.
[34] M. Valstar and M. Pantic, "Fully automatic facial action unit detection and temporal analysis", in CVPRW06: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop, Washington, DC, USA, IEEE Computer Society, 2006, pp. 149.
[35] Y. Tong, W. Liao and Q. Ji, "Facial action unit recognition by exploiting their dynamic and semantic relationships", in IEEE Trans. Pattern Anal. Mach. Intell, 2007, Vol. 29, No. 10, pp. 1683-1699.
[36] K. He, G. Wang and Y. Yang, "Optical flow-based facial feature tracking using prior measurement", in Cognitive Informatics. ICCI2008. 7th IEEE International Conference on, IEEE, 2008, pp. 324-331.
[37] D. Freedman and T. Zhang, "Active contours for tracking distributions" in IEEE Transactions on Image Processing, 2004, Vol. 13, No. 4, pp. 518-526.
[38] Z. Zhiwei and Q. Ji, "Robust real-time face pose and facial expression recovery", in Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, Vol. 1, pp. 681-688.
[39] H.T. Chen, T.L. Liu and C.S. Fuh, "Probabilistic tracking with adaptive feature selection", in Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04), Washington, DC, USA, 2004, Vol. 2, pp. 736-739.
[40] A.U. Batur and M.H Hayes, "Adaptive active appearance models" in IEEE Transactions on Image Processing, 2005, Vol. 14, No. 11, pp. 1707-1721.
[41] P. Corcoran, M.C. Ionita and I. Bacivarov, "Next generation face tracking technology using AAM techniques", in Proceedings of the International Symposium on Signals, Circuits and Systems (ISSCS'07), Ias, i, Romania, 2007, Vol. 1, pp. 14.
[42] L. Marco - Giménez, M. Arevalillo – Herráez and C. Cuhna-Pérez, "Eigenexpressions: emotion recognition using multiple eigenspaces", in Pattern Recognition and Image Analysis, of Lecture Notes in Computer Science, Springer, Berlin, Germany, 2013, pp. 758-765.

[43] H.T. Nguyen and A.W.M. Smeulders, “Fast occluded object tracking by a robust appearance filter”, in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, Vol. 26, No. 8, pp. 1099-1104.

[44] W. Bao, H. Li, N. Li and W. Jiang, “A liveness detection method for face recognition based on optical flow field”, in Image Analysis and Signal Processing, 2009. IASP 2009. International Conference on, pp. 233-236.

[45] A. Geetha. V. Ramalingam, S. Palanivel and B. Palaniappan, “Facial expression recognition a real time approach” in Expert Systems with Applications, 2009, Vol. 36, No. 1, pp. 303-308.

[46] I. Cohen, N. Sebe, F.G. Cozman, M.C. Cirelo and T.S Huang, “Learning Bayesian network classifiers for facial expression recognition using both labeled and unlabeled data”, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003, pp. 595-601.

[47] X.W. Chen and T. Huang, “Facial expression recognition: a clustering-based approach”, in Pattern Recognition Letters, 2003, Vol. 24, No. 9, pp. 1295-1302.

[48] K. Lu and Y. Jia, “Audio-visual emotion recognition with boosted coupled HMM”, in Pattern Recognition (ICPR), 2012, 21st International Conference on, pp. 1148-1151.

[49] L. Ma and K. Khorasani, “Facial expression recognition using constructive feed forward neural networks” In: IEEE Transactions on Systems, Man, and Cybernetics B: Cybernetics, 2004, Vol. 34, No. 3, pp. 1588-1595.

[50] T. Wu, S. Fu and G. Yang, “Survey of the Facial Expression Recognition Research”, in Advances in Brain Inspired Cognitive Systems 736, H. Zhang, A. Hussain, D. Liu, and Z. Wang, Eds., ed: Springer Berlin Heidelberg, 2012, pp. 392-402.

[51] C. Bouvier, P.Y. Coulon and X. Maldague, “Unsupervised Lips Segmentation Based on ROI Optimization and Parametric Model”, in Image Processing, 2007. ICIP 2007. IEEE International Conference on, 2007, pp. 301-304.

[52] J. Cohn, T. Kanade, “Use of automated facial image analysis for measurement of emotion expression”, The handbook of emotion elicitation and assessment. Oxford University Press Series in Affective Science, J.A. Coan and J.B. Allen, ed, 2007, pp. 222-238.

[53] V. Bruce. Recognizing Faces, Hove, East Sussex: Lawrence Erlbaum Associates, Inc, 1988.

[54] P. Alku, “Glottal inverse filtering analysis of human voice production - A review of estimation and parameterization methods of the glottal excitation and their applications”, in Sadhana, 2011, Vol. 36, No. 5, pp. 623-650.

[55] S. Steidl, “Automatic Classification of Emotion- Related User States in Spontaneous Children’s Speech”, Germany: University of Erlangen-Nuremberg, 2009, pp. 1-250.

[56] S.E. Fotinea, S. Bakamidis, T. Athanaselis, I. Dologlou, G. Carayannis, R. Cowie, E. Douglas-Cowie and N. Fragopanagos, “Emotion in Speech: Towards an Integration of Linguistic, Paralinguistic, and Psychological Analysis”, in Artificial Neural Networks and Neural Information Processing- ICANN/ICONIP 2003, pp. 173-173.

[57] M.E Hoque, J.K Lane, R. el Kaliouby, M. Goodwin and R.W. Picard, “Exploring Speech Therapy Games with Children on the Autism Spectrum”, in Proceedings of InterSpeech, 2009, Brighton, UK

[58] Z. Zeng , M. Pantic, G.I Roisman and T.S Huang, “A survey of affect recognition methods: Audio, visual, and spontaneous expressions”, in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, Vol. 31, No. 1, pp. 39-58.

[59] E. Douglas-Cowie, N. Campbell, R. Cowie and P. Roach, “Emotional Speech: Towards a New Generation Of Databases”, in Speech Communication, 2003, Vol. 40, pp. 33-60.

[60] J. Ang, R. Dhillon, A. Krupski, E. Shriberg and A. Stolcke, “Prosody-based automatic detection of annoyance and frustration in human-computer dialog” in Proc. International Conference on Spoken Language Processing (ICSLP 2002), 2002, pp. 2037-2040.

[61] J. Kim and E. Andr, “Emotion recognition using physiological and speech signal in short-term observation”, in Perception and interactive technologies, LNAI 4201. Springer, Berlin, Heidelberg, pp 53-64.

[62] B. Schuller, D. Seppi, A. Batliner, A. Maier and S. Steidl, “Towards more reality in the recognition of emotional speech”, in International Conference on Acoustics, Speech, and Signal Processing, Honolulu, HI, USA, 2007, Vol. 4, pp. 941-944.

[63] Y. Zhou, Y. Sun, J. Zhang and Y. Yan, “Speech Emotion Recognition Using Both Spectral and Prosodic Features”, in International Conference on Information Engineering and Computer Science, 2009. ICIECS, pp. 1-4.

[64] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, S. Kollias, W. Fellen and J. Taylor, “Emotion recognition in human-computer interaction”, in IEEE Signal Process, Vol. 18, pp. 32-80.

[65] D. Ververidis and C. Kotropoulos, “Emotional speech recognition: Resources, features, and methods”, in Speech Communication, 2006, pp. 1162-1181

[66] D. Neiberg, K. Elenius, I. Karlsson and K. Laskowski, “Emotion recognition in spontaneous speech”, Working Papers in Linguistics, 2009, Vol. 52, pp. 101-104.

[67] C. Busso, S. Lee S and S.S. Narayanan, “Using neutral speech models for emotional speech analysis. In: Interspeech 2007-Eurospeech, 2007, pp. 2225-2228.

[68] K.P. Truong and D.A. van Leeuwen, “Automatic discrimination between laughter and speech” in Speech Commun, 2007, Vol. 49, No. 2, pp. 144-158.

[69] C. Busso, Multimodal analysis of expressive human communication: speech and gesture interplay, PhD Thesis, Viterbi School of Engineering, University of Southern California, Los Angeles, California, 2008.

[70] S. Rigoulot, M.D. Pell, “Emotion in the voice influences the way we scan emotional faces”, in Speech Communication, 2014, Vol. 65, pp. 36-49.

[71] S. Hoch, F. Althoff , G. McGlaun and G. Rigoll, “Bimodal fusion of emotional data in an automotive environment”, in Proc. 30th Int’l Conf. Acoustics, Speech, and Signal Processing (ICASSP’05), 2005, Vol. 2, pp. 1085-1088.

[72] L.C. De Silva and P.C. Ng, “Bimodal emotion recognition”, in Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on, 2000, pp. 332-335.

[73] L.C. De Silva, "Audiovisual emotion recognition", in Systems, Man and Cybernetics, 2004 IEEE International Conference on, 2004, Vol. 1, pp. 649-654.
[74] L. S. Chen, T. S. Huang, T. Miyasato and R. Nakatsu, "Multimodal Emotion/Expression Recognition", in Automatic Face and Gesture Recognition, third IEEE International Conference, 1998, pp. 366-371.
[75] Y. Yoshitomi, K. Sung-Ill, T. Kawano and T. Kilazoe T, "Effect of sensor fusion for recognition of emotional states using voice, face image and thermal image of face", in Robot and Human Interactive Communication, 2000. RO-MAN 2000. Proceedings. 9th IEEE International Workshop on, 2000, pp. 178-18.
[76] Z. Zeng, J. Tu, P. Pianfetti, M. Liu, T. Zhang, Z. Zhang, T.S. Huan and S. Levinson, "Audio-Visual Affect Recognition through Multi-Stream Fused HMM for HCI", in Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR '05), 2005, pp. 967-972.
[77] K. Karpouzis, G. Caridakis, L. Kessous, N. Amir, A. Raouzaiou, L. Malatesta and S. Kollias, "Modeling Naturalistic Affective States via Facial, Vocal, and Bodily Expression Recognition", in LNAI 4451, 2007, pp. 91-112.
[78] Z. Zeng, Z. Zhang, B. Pianfetti, J. Tu and T.S Huang, "Audio-Visual Affect Recognition in Activation-Evaluation Space" in Proc. 13th ACM Int'l Conf. Multimedia (Multimedia'05), 2005, pp. 828-831.
[79] S. Petridis and M. Pantic, "Audiovisual Discrimination between Laughter and Speech" in IEEE Int'l Conf. Acoustics, Speech, and Signal Processing (ICASSP), 2008, pp. 5117-5120.
[80] J.N. Gowdy, A. Subramanya, C. Bartels and J. Bilmes, "DBN based multi-stream models for audio-visual speech recognition" in Billene, M. (ed.): Proc. IEEE International Conf. Acoustics, Speech, and Signal Processing, IEEE, Canada, 2004, Vol. 1, pp. 993-996.
[81] L. Kessous , G. Castellano and G. Caridakis, "Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis" in Journal on Multimodal User Interfaces, 2010, Vol. 3, pp. 33-48.
[82] G.A. Ramirez, O. Fuentes, S.L. Crites, M. Jimenez, J. Ordonez, "Color Analysis of Facial Skin: Detection of Emotional State", in Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on, 2014, pp.474-479.
[83] K. Wang, "The Feature Extraction Based on Texture Image Information for Emotion Sensing in Speech", in Sensors, 2014, Vol. 14, No. 9, pp. 16692-16714.
[84] M. Sidorov, W. Minker, "Emotion Recognition and Depression Diagnosis by Acoustic and Visual Features: A Multimodal Approach", in Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge (AVEC'14). ACM, New York, NY, USA, 2014, pp. 81-86.
[85] M. Wöllmer, A. Metallinou, F. Eyben, B. Schuller, S.S. Narayanan, "Context-sensitive multimodal emotion recognition from speech and facial expression using bidirectional LSTM modeling" in INTERSPEECH, 2010, pp. 2362-2365.
[86] H. Boughrara, M. Chtourou, C. Ben Amar and L. Chen, "Facial expression recognition based on an mlp neural network using constructive training algorithm", in Multimedia Tools and Applications, 2014, pp. 1-23.
[87] M. Wöllme, M. Kaiser, F. Eyben, F. Weninger, B. Schuller and G. Rigoll, "Fully automatic audiovisual emotion recognition: Voice, words, and the face", in Speech Communication; 10. ITG Symposium; Proceedings of, VDE, 2012, pp. 1-4.
[88] Y.H. Lee, W. Han, Y. Kim, C.G. Kim, "Robust Emotion Recognition Algorithm for Ambiguous Facial Expression using Optimized AAM and k-NN", in International Journal of Security and Its Applications, 2014, Vol. 8, No. 5, pp. 203-212.
[89] R. KalaiSelvi, P. Kavitha and K.L Shunmuganathan, "Automatic emotion recognition in video", in Green Computing Communication and Electrical Engineering (ICGCCEE), 2014 International Conference on, 2014, pp.1-5.
[90] J.C. Lin, C.H. Wu and W.L. Wei, "A probabilistic fusion strategy for audiovisual emotion recognition of sparse and noisy data", in Orange Technologies (ICOT), 2013 International Conference on, IEEE, pp. 278-281.
[91] P. Shifeng, T. Jianhua, L. Ya, "The CASIA audio emotion recognition method for audio/visual emotion challenge 2011", in Affective computing and intelligent interaction. Springer Berlin Heidelberg, 2011, pp. 388-395
[92] M. Sidoro, C. Brester, W. Minker and E. Semenkin, "Speech-Based Emotion Recognition: Feature Selection by Self-Adapted Multi-Criteria Genetic Algorithm", in Proceedings of the 9th edition of the Language Resources and Evaluation Conference, Reykjavik, Iceland, 2014.
[93] R.T. Clemen and T. Reilly, Making Hard Decisions with Decision Tools. Duxbury, Pacific Grove, CA, 2001.
[94] O. Martin, I. Kotsia, B. Macq and I. Pitas, "The eNTERFACE' 05 Audio-Visual Emotion Database", in Data Engineering Workshops, Proceedings. 22nd International Conference, 2006, pp. 8-8.
[95] J.G. Rázuri, D. Sundgren, R. Rahmani, A.M. Cardenas, "Automatic Emotion Recognition through Facial Expression Analysis in Merged Images Based on an Artificial Neural Network" in Proceedings of 12th Mexican International Conference on Artificial Intelligence (MICAI), Mexico City, Mexico, 2013, pp. 85-96.
[96] MathWorks (2014) DSP System Toolbox: User's Guide (R2014b). http://fr.mathworks.com/help/pdf_doc/dsp/dsp_ug.pdf
[97] F. Weninger, B. Schuller, C. Liem, F. Kurth and A. Hanjalic, "Music information retrieval: Inspirational Guide to Transfer from Related Disciplines" in Multimodal Music Processing, 2012, Vol. 3, pp. 195-216.
[98] T. Giannakopoulos, A. Pikrakis, Introduction to Audio Analysis. Elsevier Academic Press, 2014.
[99] J.G. Rázuri, D. Sundgren, R. Rahmani, A. Larsson, A.M. Cardenas, I. Bonet, "Speech emotion recognition in emotional feedback for Human-Robot Interaction", in International Journal of Advanced Research in Artificial Intelligence (IJARAI) , 2015, Vol. 4, No. 2, pp. 20-27.
[100] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I.H. Witten, "The WEKA data mining software: an update", in ACM SIGKDD Explorations Newsletter, 2009, Vol. 11, No. 1, pp. 10-18.

[101] V. Vapnik, The Nature of Statistical Learning Theory, ed. Springer-Verlag, New York, 1995.
[102] T.M. Mitchell, “Machine Learning”, McGraw-Hill: New York, NY, 1997, p. 432.
[103] D.E. Rumelhart, G.E. Hinton and R.J. Williams, “Parallel distributed processing: explorations in the microstructure of cognition”, MIT Press, Cambridge, MA, USA, 1986, Vol. 1, pp. 318-362.
[104] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Mateo, CA, 1988.
[105] J.R. Quinlan, “Induction of decision trees”, in Mach. Learn, 1986, Vol. 1, No. 1, pp. 81-106.
[106] J.R. Quinlan, C4.5: Programs for Machine Learning, 1st ed.; Morgan Kaufmann Publishers: San Francisco, CA, USA, 1993.
[107] P. Flach, Machine Learning: The Art and Science of Algorithms that Make Sense of Data, Cambridge University Press, 2012.
[108] B. Efron and R.J. Tibshirani, An introduction to the Bootstrap, Chapman and Hall: New York, USA, 1993.
[109] R. Polikar, “Ensemble based systems in decision making”, IEEE Circuits and Systems Magazine, 2006, Vol. 6, No. 3, pp. 21-44.
[110] L. Breiman, “Bagging predictors. Mach. Learn, 1996, Vol. 24, pp. 123-140.
[111] Schapire RE (1990) The strength of weak learnability. Mach. Learn 5(2):197-227.
[112] I. Witten and E. Frank, “Data Mining: Practical Machine Learning Tools and Techniques”, 2nd ed.; Diane Cerra: San Francisco, 2005, pp. 525.
[113] D. Wolpert, “Stacked generalization”, Neural networks, 1992, Vol. 5, No. 2, pp. 241-259.
[114] R.A. Jacobs, S.J. Nowlan and G.E. Hinton GE, “Adaptative mixtures of local experts”, Neural Comput, 1991, Vol. 3, pp. 79-87.
[115] T.K. Ho, “Random subspace method for constructing decision forests”, in IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, Vol. 20, No. 8, pp. 832-844.
[116] I. Bonet, P. Franco-Montero, V. Rivero, M. Teijeira, F. Borges, E. Uriarte and A. Morales Helguera, “Classifier ensemble based on feature selection and diversity measures for predicting the affinity of A2B adenosine receptor antagonists” in Journal of chemical information and modeling, 2013, Vol. 53, No. 12, pp. 3140-3155.
[117] J.G. Rázuri, A. Larsson, R. Rahmani, D. Sundgren, I. Bonet and A.M. Cardenas, “Recognition of emotions by the emotional feedback through behavioral human poses”, in International Journal of Computer Science Issues (IJCSI), 2015, Vol. 12, No.1, pp. 7.
[118] J. Rios and D.R. Insua, “Adversarial risk analysis for counterterrorism modeling”, Risk analysis, 2012, Vol. 32, No. 5, pp. 894-915.
[119] J. Hoeting, D. Madigan, A. Raftery and C. Volinsky, “Bayesian model averaging: A tutorial”, in Statistical Science, 1999, Vol. 4, pp. 382-417.
[120] S. French and D.R. Ríos, Statistical Decision Theory. Oxford University Press. New York, NY, USA, 2000.
[121] C. Bielza, P. Müller, D.R. Ríos, “Monte Carlo Methods for Decision Analysis with Applications to Influence Diagrams”, in Management Science, 1999, Vol. 45, No. 7, pp. 995-1007.
[122] A.H. Maslow, “A theory of human motivation”, Psychological Review, New York: Harper, 1943.Vol. 50, No. 4, pp. 370-396.
[123] J.G. Rázuri, D. Sundgren, R. Rahmani, A. Larsson and A.M. Cardenas, “Effect of Emotional Feedback in a Decision- Making System for an Autonomous Agent”, in International Journal of Computer Science Issues (IJCSI), 2014, Vol. 11, No.5, pp. 180-191.
[124] B.L. Fredrickson and R.W. Levenson, “Positive emotions speed recovery from the cardiovascular sequelae of negative emotions”, in Cognition and Emotion, 1998, Vol. 12, pp. 191-220.
[125] R.S. Lazarus, Emotion and adaptation. New York: Oxford University Press, 1991.
[126] P.M. Cole, M.K. Michel and L.O. Teti LO, “The development of emotion regulation and dysregulation: A clinical perspective”, Wiley-Blackwell, 1994, Vol. 59, pp. 73-100.
[127] J.G. Rázuri, D. Sundgren, R. Rahmani, A. Larsson A, “Effect of emotional feedback in a decision-making system for an autonomous agent” in Advances in Artificial Intelligence IBERAMIA 2014, A.L.C. Bazzan and K. Pichara, Eds., ed: Springer International Publishing, 2014, pp. 613-624.
[128] J.J. Gross, “Emotion regulation: taking stock and moving forward”, Emotion, 2013, Vol. 13No. 3, pp. 359.
[129] P. Ekman, “An argument for basic emotions”, in Cognition and Emotion, 1992, Vol. 6, pp. 169-200.
[130] R. Plutchik, “The nature of emotions”, in American Scientist, 2001, Vol. 89, No. 4, pp. 344-350.
[131] A. Öhman and S. Mineka, “Fears, phobias, an preparedness: Toward an evolved module of fear and fear learning”, in Psychological Review, 2001, Vol. 108, pp. 483-522.

Javier G. Rázuri is currently a researcher at the Department of Computer and System Sciences, Stockholm University. He was a researcher and Ph.D. student of Department of Statistics and Operations Research at Universidad Rey Juan Carlos and acting developer of the Affective computing department in AISoy Robotics. In 2003, he received the Bachelor of electronic engineering degree at Peruvian University of Applied Sciences (UPC) in Perú. In 2007, he received a Master’s Degree in Business Management at IEDE - Business School of European University of Madrid. In 2008, he received a Master's Degree in Decision Systems Engineering at Universidad Rey Juan Carlos. From 2008, he was a researcher in several projects financed for the university and the European Union focused on Robotics and decision-making related to EU higher education. His research interests are related with, Affective computing, emotional decision making, Human-machine interaction, Robotics, Human Robotics, Robotics and Autonomous Systems, Neuroscience, Sentiment analysis, in order to reproducing behavior patterns similar to human and provide to the agents a type of emotional intelligence and improve the interaction experience making more close the loop human-robot emotional interaction.