Evaluation of Emotional Agents' Architectures: an Approach Based on Quality Metrics and the Influence of Emotions on Users

M. O. Meira, A. M. P. Canuto

Abstract – **There is a need for multi-agent system designers in determining the quality of systems in the earliest phases of the development process. The architectures of the agents are also part of the design of these systems, and therefore also need to have their quality evaluated. Motivated by the important role that emotions play in our daily lives, embodied agents researchers have aimed to create agents capable of producing affective and natural interaction with users that produces a beneficial or desirable result. For this, several studies proposing architectures of agents with emotions arose without the accompaniment of appropriate methods for the assessment of these architectures. The objective of this study is to propose a methodology for evaluating emotional agents' architectures, which evaluates the quality's attributes of the design of architectures, besides to evaluation of human-computer interaction, the effects on the subjective experience of users of applications that implement it. In assessing the quality of architectural design, the attributes assessed were: extensibility, modularity and complexity. In assessing the effects on users' subjective experience, which involves the implementation of the architectures in applications, and we suggest the domain of computer games, metrics chosen were: enjoyment, felt support, warm, caring, trust, cooperation, intelligence, interestingness, naturalness of emotional reactions, believability, reducing of frustration, likeability, and the average time and average attempts. We have experimented our approach and evaluated five emotional agents' architectures: BDIE, DETT, Camurri-Coglio, EBDI and Emotional-BDI.**

Index Terms – **architecture evaluation, emotional agents, human-computer interaction, multi-agents systems, quality metrics.**

I. INTRODUCTION

Rosalind [1] defined Affective Computing expression as "computing that is related to, arises from, or deliberately influences emotions". Since then, studies have been developed trying to provide machines and computers with similar emotional skills to human users, such as the recognition of skills and emotional expression [2].

Moreover, emotions are considered key elements for more effective human computer interaction (HCI), "machines may never need all of the emotional skills that people need; however, there is evidence that machines will require at least some of these skills to appear intelligent when interacting with people" [3].

Manuscript received December 03, 2014; revised March 11, 2015.

Marcílio de Oliveira Meira. Federal Institute of Education, Science and Technology of Rio Grande do Norte – IFRN. Brazil. M.Sc. in Computer Science (e-mail address: marcilio.meira@ifrn.edu.br)

Anne Magaly de Paula Canuto. Federal University of Rio Grande do Norte – UFRN. Brazil. Ph.D in Electronics Engineering (e-mail address: anne.canuto@dimap.ufrn.br).

Thus, following the pioneering work of Bates [4] and Picard [1], agents that exhibit human emotions now become a common subject for both industrial and academic research circles [6], for example, Oz Project [4], Comic Chat [7], and Virtual Theater [8], and the humanoid robot Kismet [9].

According to Padgham and Taylor [10], there are many reasons to incorporate emotions in agents, one of them being that emotions can make most engaging and believable agents. Wehrle [11] also offered some motivations for modeling emotions in agents: (i) as a way to improve our knowledge about the nature of emotion and its implications; (ii) in the field of engineering, observing the performance criterion, to build good artifacts to specific tasks; (iii) in human computer interaction, taking into account acceptance issues, performance and usability, the use of agents with emotions to improve the interaction of people with computers.

In general, the reason to analyze software architecture is to learn about the system that is to be implemented, and thus have the benefits of a clear and universally defined semantic on the descriptions of the software architecture. Another reason is the architecture has a great impact on the quality of a software system, and so it is important to be able to make informed decisions about such architecture in a number of situations [12]. In large software systems, software quality attributes such as performance, security and changeability depend not only on the code level practices, but also the software architecture as a whole [13].

According to García-Magariño et al. [14], projects of multi-agent systems (MAS) also lack adequate systems to determine their quality at early stages of the development process. Likewise, in the area of emotional agents' architectures, we still can observe the lack of studies to provide results of experiments, and especially a specific methodology to evaluate agents' architectures with emotions, which considers at least two points: quality criteria and the effects of user interaction with the emotional agent.

Therefore, given these needs for determining the quality of agents' architectures in the early stages of the development of applications, and given studies on the impact of emotional agents on human computer interaction, this work aimed to propose an assessment methodology over architecture of agents with emotions that supports all these needs. The methodology is based on a well-defined metrics model, from the definition of objectives, which are to evaluate the quality attributes of architectures and the effects of user interaction with the emotional agents.

In next sections, we will review some theories and architectures about emotional agents, will discuss methods for measuring agents' architectures, present our proposed methodology in assessing emotional agents' architectures, and finally discuss our results.

II. EMOTIONAL AGENTS

When we incorporate emotions in agents, we can make them most engaging and believable, so they can better perform in various interactive systems involving simulation, as in entertainment applications, education and training [10].

According Padgham and Taylor [10], emotions can affect the goals of an agent, and consequently affect your actions, and goals can also affect emotional states. However, for the agents "experiment" some emotion, it is necessary that its designers give them some mechanism or subsystem for processing these emotions, which has usually been done by choosing an emotional model among various structural emotions theories [15]-[17], such as Ortony et al. [18], known as OCC model, which has been experienced in agents' architectures with emotions [19].

There are another studies attempting to match various other emotional theories, such as *Cathexis* system [20] which seeks to identify and create explicit models for six different families of emotions. Despite all of this, there is still no clear definition of which model to use [17].

A. BDIE

The BDIE approach [16] was motivated by the need to work with human interaction robot and the interest of creating artificial agents for computer games, and its goal is to capture the effects of emotions in a human mind and play them in artificial agents.

The differences between BDI [21] and BDIE architectures are the first one has only three data repositories: beliefs, desires and intentions, and three data processing algorithms: perceptual process, desire generator, and the planning and execution algorithm. The BDIE architecture has new modules containing both data repositories and processes: perceptual, motivational, behavioral and emotional systems, where all have inter-dependencies among them, and we highlight the emotional system, which maintains the affective state of the agent, and works with three levels of emotions: primary emotions, secondary and tertiary.

B. EBDI

The EBDI architecture proposed by Jiang et al. [22] separates the practical reasoning of the specific emotion mechanism, extends the traditional BDI architecture and is influenced by primary and secondary emotions in the decision-making process.

Comparing the traditional BDI architecture with EBDI, the last has four modules: emotion, beliefs, desires, and intentions, in contrary the first has only the last three. In EBDI, beliefs are influenced by both the environment and the emotional situation. In addition, it were added methods that allow acquire beliefs not only of perceptual form, but also through communication and contemplation, which is a reconsideration of beliefs based on the current situation of the emotions and intentions. So, EBDI architecture can be summarized as a loop of a BDI agent that manages and integrates emotions.

C. DETT

Observing emotions as essential to human behavior, considering them as important as rational analysis in stressful situations such as combat, Parunak et al. [19] proposed a model of emotions for situated agents using as a guide the OCC model [18], they called DETT (Disposition, Emotion, Trigger, Tendency). That architecture captures the essential aspects of the OCC emotions model, was designed for situated agents, and handles your emotions as triggered by perception, instead of being purely a result of the internal reasoning.

The DETT architecture also adds a new component, the *Disposition* which, as well as the desires, is persistent and have a mapping with the Emotions. *Disposition* modulates *Appraisal* to determine the extent to which a given belief triggers the corresponding emotion. *Emotion* in turn modulates the analysis to impose a trend the resulting intention. We highlight the *Perception* process, which updates *Beliefs* and is based on a pheromone infrastructure [23], which are spread spatially and evaporate over time.

D. Emotional BDI

Pereira et al. [24] also proposed an extension of the BDI architecture by adding emotions, which incorporates an accurate model of practical reasoning [25] through the interconnection of the mechanisms that are responsible for managing the emotional state, resources and capabilities of the agent, and all the mechanisms that make up the classic BDI architecture.

According to Pereira et al., the Emotional-BDI architecture solves three problems encountered in the original BDI. First, it solves the lack of information about the limits of resources, using the concepts of *Capabilities* and *Resources*, called *Effective,* introduced to the original BDI by Padgham and Lambrix [26]. Second, the problem of time for reconsideration of the environment, using *Sensing and Perception Module*. The last problem, the lack of other human mental states, that despite the fact that mental state, was addressed with the *Emotional State Manager*, which is responsible for controlling the use of all *Resources* and *Capabilities*.

E. Camurri-Coglio

Camurri and Coglio [27] suggested an agent architecture with emotional state, which interacts with the outside world, receiving input stimuli and sending headed back. His emotional state is developed over time, influences the output, and its evolution responds to inputs and the upgrade process is flexible and not restricted to a single stream possible implementation.

In their architecture, there are the concepts of buffer among components, data repositories, and flow of information among components, that acting as producers or consumers of data. We can observe five main components: Input, Rational, Emotional, Reactive, and Output, where the input has the responsibility to provide the sensors, processing the input data and distribute them to the Reactive, Emotional and Rational components. According to the authors, this input component, to have all these functions, needs to be built with various modules each responsible for a certain type of processing.

III. METHODS FOR MEASUING AGENTS' ARCHITECTURES

In this section we will discuss various approaches regarding the evaluation of agents' architectures, starting with the evaluation methods, then describing metrics adopted in the evaluation of other architectures of multi-agent systems and agents, and finally discussing specific studies on the evaluation of agents with emotions.

A. Analytical and Experimental Evaluation Methods

Hayes-Roth [28] presented several goals and issues that we want to address with an evaluation of integrated agent architecture. She ranked formal evaluation methods to address these questions and similar others in two main categories: analytical methods and experimental methods.

Regarding the analytical category, Hayes-Roth discussed the mathematical evaluation methods, that provide the advantages of having a mathematical certainty but also has drawbacks that is the complexity of the systems being analyzed, which can lead to intractable mathematical models. Finally, she concludes that formal methods often must be supplemented by other techniques.

On the other category, the experimental evaluation method measures the properties of particular instances of runs of experiments in particular IA systems, and in the end the results are used as evidence for or against certain conclusions, using inductive inference of general relations between the IA design concepts and the performance of observations-based systems. The goal is to find general relations which can be expected always when the appropriate conditions are maintained.

B. Evaluation of Agents' Architectures

About the quality of agents' architecture, Hexmoor [29] reported that the comparison of features and architectures attributes are futile because they are domain dependent. Thus, he emphasizes the empirical assessments of architectures on their own qualities, besides the utility of developing domain independent metrics such as the measure of the number of missed opportunities due to slowness of agents' reactions, the impact of intentionality on the opportunities, autonomy, robustness and fault tolerance.

The metrics used by Hexmoor did not apply to this project since his assumption about the constancy of appearance rate of opportunities and reaction time, *we believe those are also domain dependent*. In addition, his other metrics require interactions among agents, which involve coordination, cooperation and communication issues, and also do not apply to our methodology, because we do not address these issues.

On the other hand, in the field of formal assessments, Alechina and Logan [30] argued that despite the informal descriptions are useful, but they can be difficult to be confirmed when trying to determine whether a given system has a particular property as a result of its architecture. Alechina and Logan further explained the classification of agents' architectures by specifying both sets of concepts or possible states (beliefs, goals, emotions) and skills or possible transitions (actions, perceptions, learning, planning) carried out by the system. Using a logical called SimpleAPL, Alechina et al. [31] presented a formal agents' architectures assessment methodology, which consists in defining a set of transition systems corresponding to an architecture of interest, and check the properties of this set of transition systems.

Rose et al. [32] reported an agent architecture that incorporates social and philosophical layers, whose evaluation in test scenarios, allowed them to make comparisons between the performances of agents' architectures with different combinations and orders of precedence of philosophical principles.

Other metrics of architectures evaluation were presented by Lee et al. [33], who identified them to measure the attributes that characterize the quality of systems based on agents. These quality metrics are useful to analyze the quality of system's characteristics based on agents where. Just as most Hexmoor's metrics [29], metrics used by Lee et al. evaluate attributes associated with interactions among agents, which are outside the scope of this research.

On evaluation of SMA architectures, García-Magariño et al. [14] proposed a suite of metrics to measure certain quality attributes of these systems, considering the agents and their organization. Most metrics were inspired by object-oriented metrics [34], and also in addressing *factors-criteria-metrics* of McCall et al. [35], so that had to be adapted to the agents oriented concepts. The suite aims to evaluate the quality of architectures following attributes: *extensibility*, *modularity* and *complexity*.

Dobrica et al. [36] showed in their research eight of the most representative methods of software architecture analysis, the IEEE Std standard. 1061 [37] presents a list of examples of these attributes (quality factors), and the research of García-Magariño et al. [14] is a good example of application of quality factors in the evaluation of architectures of multi-agent systems.

In particular, the method was tested by Woods and Barbacci [13] in agent-based systems. This method focuses on finding points of trade-off (advantages and disadvantages) in architecture from the perspective of product quality requirements [12]. The method entries consist of a system architecture and perspectives of stakeholders involved in that system. The output is an understanding of what architectural decisions will be used to achieve the specific qualities goals and the implications of these decisions.

The SAEM method has a distinction as comparing to other methods, which is the adopted evaluation technique: metrics and the GQM paradigm [38]. The metrics are considered by domain experts as a more accurate method for the assessment of attributes in terms of architecture [39]-[40].

Dobrica et al. [36] considers the GQM as a good technique to create new metrics from a certain sequence of reasoning, where the result of applying this paradigm is the specification of a measurement system with a target set specific questions and a set of rules for the interpretation of measured data. This resulting measurement model has three levels: conceptual, operational, and quantitative, which respectively refer to the objectives, questions and metrics. The **GQM paradigm was also adopted in this research**.

C. Evaluation of Emotional Agents

Specifically for the architectures of agents with emotions, aspects of appearance and behavior credibility are studied and measured, discussed the taxonomy of conversation agents in Isbister and Doyle [41]. This is the believability of agents, in which agents provide an "illusion of life", and that according to Isbister and Doyle can be evaluated by survey technique, useful to measure the satisfaction of a certain audience about the "illusion of life" of the behavior of agents. An example of this behavior is the *chatterbot* [42], which has emotional component to improve the realism of the conversation.

Many articles were published on aspects of the interaction

ISBN: 978-988-19253-4-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

between emotional agents and users. Beale and Creed [43] gathered and reviewed the results of several studies on the emotional agents affect users. They explored works focused on the areas of education and learning, video games, behavioral changes, and other collaboration.

Another study analyzed the collaboration area [44], which tested the influence of the incorporation of a character, for both on screen and in robotic form, and their emotional expressions. In the experiment, subjects were asked to complete a task with the character on the screen, which involved negotiating value stamps. The subjects rated the agent on a number of variables, and it was concluded that the inclusion of emotion had no significant impact on the perceptions and behaviors of the subjects.

Note that the embodiment of the agent was not human and consisted of a round face, with a line to the brow, a single large eye, and a line for the mouth, and no initial check was made to see if the participants recognized the expressions emotional.

The general conclusion of Beale and Creed was that the results of research in all areas, are often inconclusive and contradictory, as several studies highlight the potential of the simulated emotion to influence users in various ways, sometimes better and sometimes difficult interactions with agents. Nevertheless, the researchers emphasize that there is no strong evidence to suggest that emotional agents hinder interaction.

Based on the discussions of this review, Beale and Creed also provided general guidelines for conducting research on the subject, such as: (i) always validate the emotional expressions; (ii) promote a fair comparison between emotional and non emotional agents; (iii) provide explicit descriptions of the emotional expressions; (iv) provide statistics which report all mean values and standard deviations; (v) pursue a more refined approach to study the types of agents, in which types of domain, cause what kinds of effects on attitudes of users.

From this survey of Beale and Creed, from their comparative tables, taking only those studies of the highest importance, and selecting only the variables that showed positive results, **we captured twelve metrics** related to evaluate the influence of simulated emotions on subjective experience of users. In the next section will be described all the details about the selected metrics and issues that helped answer.

IV. PROPOSAL OF METHODOLOGY FOR ASSESSING EMOTIONAL AGENT ARCHITECTURES

Now we will explain the proposed methodology of this work, which aims to address the problem of lack of methodologies for evaluating emotional agents' architectures. Thus, about the two research methods mentioned in the previous sections, the method which could be chosen for this work was the experimental one, due to the difficulties of the analytical method raised by Hayes-Roth [28] and confirmed by Wooldridge [45].

However, to drive the choice of the metrics, the use of the GQM paradigm [38] has been very useful, from the definition of evaluation objectives, questions were derived, and based on related works, we have chosen the necessary metrics to achieve the answers on a methodology for evaluating emotional agents' architectures.

A. Proposed Model

Using GQM approach and inspired in the researches of García-Magariño et al. [14] and Beale and Creed [43], in table I we have summarized into three level the proposed model of metrics for evaluating architectures, where we can track the metrics, from the two evaluation goals, through the four questions used to respond and determine whether goals were achieved. In the first column of table I, we placed the goals considered essential in the evaluation of an emotional agents' architecture:

- assess the quality of architectural design;
- assess the human-computer interaction using the incorporation of agents.

In the second column, we write the questions that were derived from goals, and in the third column, the metrics responsible for answering each question of evaluation.

About the metrics related to quality attributes, it is believed that are of great importance for being able to assess the architectural aspects of the agents, before implementing them. It is also believed that such metrics can help identify architectural problems [14] and still rank which architectures have better quality attributes.

TABLE I. MEASUREMENT MODEL TO EVALUATE ARCHITECTURES OF AGENTS WITH EMOTIONS

Conceptual Level Goals	Operational Level Questions	Quantitative Level Metrics
(i) Evaluate the Quality of Architectural Design	What is the *extensibility* level of architecture?	Cohesion and Coupling
	What is the *modularity* level of architecture?	Cohesion, Coupling, Fan-in, Fan-out,
	What is the *complexity* level of architecture?	Communications per Module, Services per Module, Components per Module, Knowledge per Module, and Size
(ii) Evaluate the Human-computer Interaction using embodied emotional agents	What is the influence level of implemented emotional agent on *subjective experience* of application users?	Cooperation, Likeability, Felt support, Warm, Enjoyment, Trust, Caring, Reducing of frustration, Intelligence, Naturalness of emotional reactions, Believability, and Interestingness
	What is the influence level of emotional agent on *performance* of application users?	Elapsed Time for user win, and Necessary Attempts for user win

Based on the survey conducted by Beale and Creed [43], since there is much interest in research of embedded agents [46]-[53], and the potential to improve human-computer interaction (HCI) with agents remains uncertain [6], then we defend these HCI aspects should also be assessed in an agent with emotions architecture. Thus, the metrics for the second objective were chosen from Beale and Creed research, noting only the works considered highly relevant.

B. Evaluation Steps

We can summarize the methodology in two main steps that are grouped around the two goals explained above for evaluate the quality attributes and the effects of interaction with emotional agents.

ISBN: 978-988-19253-4-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

IV.B.i. Quality Attributes Evaluation

Based on suite of metrics and equations proposed by García-Magariño et al. [14], we must solve equations (1) and (2) in tables II, calculate metrics in table III, to get the measurements of quality attributes. However, in this work, variables of García-Magariño et al. had to be adapted to the specific assessment of emotional agents' architectures, since García-Magariño et al. evaluated SMA architectures.

According to García-Magariño et al., the metrics for *extensibility* help verify whether the architectures are designed to include hooks and mechanisms to expand or improve them with new skills without having to make major changes in its infrastructure. That metric is closely related to the concepts of modules and components.

For the *modularity*, whose metric help verify whether modules can be separated from one another and recombined without great cost, beyond metrics *cohesion* and *coupling* shared with *extensibility*, also apply the *fan-in* and *fan-out* metrics, respectively measuring the number of arrivals and outputs dependencies of a module.

TABLE II. METRICS, EQUATIONS AND VARIABLES FOR QUALITY ATTRIBUTES. GARCÍA-MAGARIÑO ET AL., 2010

Metrics	Definition	Equations	Variables
Ch_a Architecture Cohesion	Average cohesion of each module of an architecture	$\begin{cases} \frac{DepComp}{MaxDepComp} & \text{If } NC >= 2 \\ 1 & \text{If } NC = 1 \end{cases}$ (1)	DepComp = Count of internal dependencies per module. MaxDepComp = max count of possible dependencies: $\frac{NC*(NC-1)}{2}$ NC = count of components per module
Cp Architecture Coupling	The real count of external dependencies divided by the count of possible external dependencies	$\begin{cases} \frac{DepMod}{MaxDepMod} & \text{If } N >= 2 \\ 1 & \text{If } N = 1 \end{cases}$ (2)	DepMod = Count of dependencies among modules. MaxDepMod = Max count of dependencies among modules: $\frac{N*(N-1)}{2}$ N = count of architecture components.

TABLE III. OTHER METRICS FOR QUALITY ATTRIBUTES. GARCÍA-MAGARIÑO ET AL., 2010

Metric	Title	Definition
Fi	Fan-in	Count of incoming communication dependencies per module
Fo	Fan-out	Count of outgoing communication dependencies per module
ACmM	Average communication per module	Count of communication protocols divided by the number of modules
ASM	Service average per module	Count of service divided by the number of modules
AKM	Average Knowledge Elements per module	Count of architectural components divided by the number of modules
ACM	Average components per module	Count of knowledge elements divided by the number of modules
Sz	Size	Count of model elements for each type of component

Noting that the only common module in all emotional agents' architectures is the module related to the processing or evaluation of emotions ([21], [19], [22], [24], [27]), and there is a tendency among researchers that the architecture should allow the combination of various theories of emotion in the reasoning process of agents, therefore *fan-in* and *fan-out* metrics will be analyzed *only for the emotional module*. Thus, we allow to be investigated whether the emotional module is responsible for the high coupling, and consequently the cause of low modularity of the architecture.

For the *complexity*, which measures the responsiveness of architecture, has several metrics that involve relationships among architectural elements, the amounts module of elements (e.g. knowledge elements, components and instances), and the total amount of elements (e.g. size). From the original list of metrics, about the *complexity* advocated by García-Magariño et al. [14], the only one not applied to this research was the *average of instances per module*.

Therefore, the evaluation of quality attributes is achieved as follows:

- **Analysis of architectural models**: an initial analysis of the architecture on its models and diagrams, to identify and count all modules, components, communications among modules, architecture services, and knowledge elements;
- **Evaluate the extensibility**: for each architecture, calculate Ch_a and Cp, solve equations (1) and (2);
- **Evaluate the modularity**: for each architecture, calculate Ch_a and Cp, solve equations (1) and (2), and also calculate Fi and;
- **Evaluate the complexity**: for each architecture, calculate $ACmM$, ASM, AKM, ACM and Sz;
- **Sort architectures**: in order to compare the architectures using calculated metrics above, we should calculate the arithmetic mean for all metrics, and then sort architectures by their relative position to the mean;
- **Quality attributes analysis**: to compare all architectures and discover what has the best architectural design, we made a variation of the method García-Magariño et al. [14], and we assign to each architecture a quantity of "+" or "-", the more an architecture moves away, above or below respectively, the arithmetical mean calculated above.

When sorting architectures, horizontal bar charts are very useful to make the comparison, putting the names of architectures as the ordinate and the calculated metric values on the abscissa. The values of the ordinate can be sorted considering the architecture position relatively to the average. It should make a chart for each metric. This makes it easier to see for example what architecture has the highest cohesion, or the lowest coupling.

Finally, for each architecture, we must count all the "+" made in the score, subtracting that sum all the "-" that have also won. The final number that algebraic operation is the overall rating as the level of quality of the architectural design of architectures. Thus, now it is possible to objectively point out which of emotional agents' architecture has the best architectural design in terms of quality attributes.

IV.B.ii. Computer-Human Interaction Evaluation

On the other hand, the second moment of assessment aims to clarify the second goal's questions, which is to assess the human-computer interaction using embodied emotional agents. In this case, we must perform specific steps with suitable procedures for this type of evaluation that, in our methodology, were based on general guidelines of Beale and Creed [43].

This evaluation involves implementing an application with the agents of architectures that are being assessed, and the participation of people in experiments, who will be consulted about their subjective experience after interacting with the application. In the end, based on the collected data, we can compare the architectures of agents with emotions in relation to human-computer interaction process.

Therefore, the evaluation of the human-computer evaluation is performed as follows:

- **Design applications**: (i) first choose a domain, we suggest the games because the richness and complexity of the characters of the current games, in [43] we can find other options; (ii) use specific methodology and development platform for agents, we suggest the Prometheus [54] and JADEX [55] respectively; (iii) define the "emotional expressions", that the embodied agent will use; (iv) finally, implement *one application version with emotional agents and a second version without emotional agents*;
- **Prepare questionnaires**: prepare one first questionnaire to validate the emotional expressions by application users, before using the application, and a second one, for answering after using the application, whose questions involve the metrics about subjective experience of the participants;
- **Define and organize the population**: (i) define a physical or geographical location, for the choice of population, balance the number of people of both sexes, as well as don't vary too much the age range of participants, where it is important considering the experience of participants in the game implemented as one of the survey variables; (ii) divide population into two groups: one for interacting with emotional agent, and another for interacting with agent without emotions;
- **Validate the understanding of emotional expressions**: by the specific questionnaire described above;
- **Perform Experiments and Apply Questionnaires**: put the population in contact with the developed applications; where besides the data of subjective user experience, who interacted with the emotional agent, the user performance information should also be collected and recorded in the same questionnaire;
- **Analyze results of emotional expressions validation**: data analysis should focus on the error rate and recognition hit for each emotional expression, where the result of this validation can confirm problems in emotional expressions used, for example, a confusion between expressions of sadness and concern;
- **Analyze results of subjective experience**: based on the second questionnaire answers, about the subjective experience of participants, the analysis of results can be divided into two phases: (i) analysis of subjective variables; (ii) analysis of the influence of emotional agents on the performance variables of users in easy and difficult levels of the game.

If the type of implemented game allow the option to choose the level of difficulty, we should also record the performance data of the users in the game for easy level and then register to a more difficult level. This procedure is essential to examine whether an architecture, which offers emotional agents, favors the performance improvement of the players. The players performance will also be measured both in the version of the game without emotional agents, as in the version with emotional agents.

The analysis of emotional expressions validation can be made easier with the help of a worksheet or chart. Indicate the use of a bar graph where the abscissa axis is the names of the emotional expression, and the ordinate the percentages. For each abscissa, two bars show both accuracy and error rates. The use of inappropriate emotional expressions may confuse participants and prejudice the final outcome of the evaluation of architectures by measuring the subjective experience of application users.

V. CONCLUSIONS

We have experimented our approach and evaluated five emotional agents' architectures: BDIE, DETT, Camurri-Coglio, EBDI, Emotional-BDI, all described previously. In the evaluation of quality attributes, the architecture with the highest *extensibility* was Emotional-BDI, the best *modularity* was DETT, and the less *complex* was Camurri-Coglio. Considering all the metrics of quality attributes, the architecture that achieved the best overall result was DETT followed by BDIE. The worst results were the Emotional-BDI followed by EBDI.

In the evaluation of subjective experience of users by the interaction with the embodied emotional agents, about the obtained results, there were no significant differences between the BDIE and EBDI architectures, which were both chosen to be implemented. However, the results showed that, about the *Intelligence* metric, the BDIE architecture overcame EBDI by almost 40%.

By using the method in the experiments, from the results obtained in the evaluation of BDIE EBDI and architectures, one can deduce that the BDIE architecture is preferred over the EBDI, at least in the field of computer games, because presented an architectural design with best quality attributes, achieved a positive influence on the user performance, and also presented the highest rates in most metrics on the subjective experience of users.

The potential benefits of the methodology proposed in this work, mainly is for designers of emotional agents' architectures and developers of applications involving agents with emotions. Application developers can have now objective criteria to analyze and choose the architecture that best contributes to achieve the desired qualities, and. designers of emotional agents' architecture can now use our proposed metrics for verification and validation of their own architectures.

We believe that there are still many points of improvements to our methodology that could be for example: (i) provide reference values for the metrics; (ii) the adoption of other metrics including favoring the evaluation of the social aspects of the agents; (iii) the availability of online applications and questionnaires over the Internet; and (iv) evaluate the results of subjective experience, by varying the emotional models employed in the emotional component architectures, or implement other applications in different fields of application as those used in this work.

REFERENCES

[1] Rosalind W. Picard. 1997. *Affective Computing*. MIT Press, Cambridge, MA, USA.

[2] Ciceri, M.R., Balzarotti, S., From signals to emotions: Applying emotion models to HM interactions, in Or Jimm, O.J. (ed.), Affective Computing. Emotion Modelling, Synthesis and Recognition, I-Tech Education and Publishing,, Vienna 2008: 271- 296.

[3] Picard, R.W.; Vyzas, E.; Healey, J., "Toward machine emotional intelligence: analysis of affective physiological state," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.23, no.10, pp.1175,1191, Oct 2001.

[4] Joseph Bates. 1994. The role of emotion in believable agents. *Commun. ACM* 37, 7 (July 1994), 122-125.

[5] Cañamero, L. D. 2001. Building emotional artifacts in social worlds: Challenges and perspectives. Paper presented at the AAAI Fall Symposium, North Falmouth, MA

[6] Scott Brave, Clifford Nass, and Kevin Hutchinson. 2005. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *Int. J. Hum.-Comput. Stud.* 62, 2 (February 2005), 161-178.

[7] David Kurlander, Tim Skelly, and David Salesin. 1996. Comic Chat. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (SIGGRAPH '96). ACM, New York, NY, USA, 225-236.

[8] Hayes-Roth, B., and van Gent, R. Story-Making with Improvisational Puppets and Actors. Tech. Report KSL-96-.05, Stanford, 1996.

[9] Cynthia Breazeal. 2002. *Designing Sociable Robots*. MIT Press, Cambridge, MA, USA.

[10] Lin Padgham and Guy Taylor. 1996. A System for Modelling Agents Having Emotion and Personality. In *Proceedings from the Workshop on Intelligent Agent Systems, Theoretical and Practical Issues* (PRICAI '96), Lawrence Cavedon, Anand S. Rao, and Wayne Wobcke (Eds.). Springer-Verlag, London, UK, UK, 59-71.

[11] Wehrle, T. (1998). Motivations behind modeling emotions agents: Whose emotions does your robot have? In Workshop W5: Grounding Emotions in Adaptive Systems (pp. 71-76). Workshop conducted at SAB'98: Fifth International Conference on Simulation of Adaptive Behavior.

[12] Bengtsson, P. (1998), "Towards Maintainability Metrics on Software Architecture: An Adaptation of Object-Oriented Metrics," In *Proceedings of First Nordic Workshop on Software Architecture*, Research Report 1998:14 ISSN: 1103-1581, University of Karlskrona/Ronneby, Ronneby, Sweden, pp. 87-91.

[13] Woods, S.G., Barbacci, M.: Architectural evaluation of collaborative agent-based systems. Technical Report CMU/SEI-99-TR-025, CMU/SEI (1999).

[14] Iván García-Magariño, Massimo Cossentino, and Valeria Seidita. 2010. A metrics suite for evaluating agent-oriented architectures. In *Proceedings of the 2010 ACM Symposium on Applied Computing* (SAC '10). ACM, New York, NY, USA, 912-919.

[15] Cynthia Breazeal. 2003. Emotion and sociable humanoid robots. *Int. J. Hum.-Comput. Stud.* 59, 1-2 (July 2003), 119-155.

[16] D. Herández, O. Déniz, J. Lorenzo, and M. Hernández. BDIE: a BDI like architecture with emotional capabilities. In Architectures for modelling emotion: Cross-Disciplinary foundations. From AAAI Spring Symposium, 2004.

[17] Crane, E. and Peter, C. A working definition for HCI specific emotion research. In C. Peter, R. Beale, E. Crane, L. Axelrod and G. Blyth (Eds.). *Emotion in HCI: Joint Proceedings of the 2005, 2006, and 2007 International Workshops*. Fraunhofer IRB Verlag, Stuttgart, 2008, pp. 54-61.

[18] Ortony, A., Clore, G. L. and Collins, A.: 1998, *The cognitive structure of emotions*. Cambridge University Press.

[19] H. Van Dyke Parunak, Robert Bisson, Sven Brueckner, Robert Matthews, and John Sauter. 2006. A model of emotions for situated agents. In *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems* (AAMAS '06). ACM, New York, NY, USA, 993-995.

[20] Juan D. Velásquez, When robots weep: emotional memories and decision-making, Proceedings of the fifteenth national/tenth conference on Artificial intelligence/Innovative applications of artificial intelligence, p.70-75, July 1998, Madison, Wisconsin, USA.

[21] Rao, A., and Georgeff, M. 1995. BDI agents: From theory to practice. In Proceedings of the First International Conference on Multi-Agent Systems, 312-319. San Francisco: AAAI Press.

[22] Hong Jiang, Jose M. Vidal, and Michael N. Huhns. 2007. EBDI: an architecture for emotional agents. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems* (AAMAS '07). ACM, New York, NY, USA, Article 11, 3 pages.

[23] Brueckner, S. Return from the Ant: Synthetic Ecosystems for Manufacturing Control. Thesis at Humboldt University Berlin, Department of Computer Science, 2000.

[24] David Pereira, Eugênio Oliveira, and Nelma Moreira. 2008. Formal Modelling of Emotions in BDI Agents. In *Computational Logic in Multi-Agent Systems*, Fariba Sadri and Ken Satoh (Eds.). Lecture Notes In Artificial Intelligence, Vol. 5056. Springer-Verlag, Berlin, Heidelberg 62-81.

[25] Bratman, ME, Israel, DJ and Pollack, ME, 1987. "Toward an architecture for resource-bounded agents" Technical Report CSLI-87-104, Center for the Study of Language and Information, SRI and Stanford University.

[26] Lin Padgham and Patrick Lambrix. 2005. Formalisations of Capabilities for BDI-Agents. *Autonomous Agents and Multi-Agent Systems* 10, 3 (May 2005), 249-271.

[27] Antonio Camurri and Alessandro Coglio. 1998. An Architecture for Emotional Agents. *IEEE MultiMedia* 5, 4 (October 1998), 24-33.

[28] Barbara Hayes-Roth. 1991. Evaluation of integrated agent architectures. *SIGART Bull.* 2, 4 (July 1991), 82-84.

[29] H. Hexmoor, M. Lafary, M. Trosen, 1999, Towards Empirical Evaluation of Agent Architecture Qualities, In Agent Theories, Architectures, and Languages, Orlando, Florida.

[30] Natasha Alechina and Brian Logan (2007). Formal Evaluation of Agent Architectures. Evaluating Architectures for Intelligence: Papers from the 2007 AAAI Workshop (pp. 1-4).

[31] Natasha Alechina, Mehdi Dastani, Brian Logan, and John-Jules Ch. Meyer (2007). A Logic of Agent Programs. Proceedings of the Twenty-Second National Conference on Artificial Intelligence (AAAI 2007) (pp. 795-800).

[32] John R. Rose, William H. Turkett, Michael N. Huhns, Soumik Sinha Roy. An Evaluation of Philosophical Agent Architectures for Mission Robustness. in Innovative Concepts for Agent-Based Systems: First International Workshop on Radical Agent Concepts WRAC, Walt Truszkowski, Mike Hinchey, and Chris Rouff, editors, Springer, vol. LNCS 2564, Berlin, 2003, pp. 201-214.

[33] S.K. Lee, C.S. Hwang. Architecture modeling and evaluation for design of agent-based system. In Proceedings of Journal of Systems and Software. 2004, 195-208.

[34] Genero, M., Piattini, M., & Calero, C. 2005. A survey of metrics for UML class diagrams. Journal of Object Technology, 49, 59-92.

[35] J. McCall, P. Richards, and G. Walters. Factors in Software Quality (Vol. 1, 2 and 3). Technical report, Nat'l Tech. Information Service, Springfield, Va. NTIS AD-AO49-014, 015, 055, Nov. 1977.

[36] Liliana Dobrica and Eila Niemelä. 2002. A survey on software architecture analysis methods. *IEEE Trans. Softw. Eng.* 28, 7 (July 2002), 638-653.

[37] IEEE. Standard for a Software Quality Metrics Methodology. In IEEE Standard for a Software Quality Metrics Methodology (1998).

[38] V. R. Basili and H. D. Rombach. 1988. The TAME project: towards improvement-oriented software environments. *IEEE Trans. Softw. Eng.* 14, 6 (June 1988), 758-773.

[39] Rick Kazman, Len Bass, Mike Webb, and Gregory Abowd. 1994. SAAM: a method for analyzing the properties of software architectures. In *Proceedings of the 16th international conference on Software engineering* (ICSE '94). IEEE Computer Society Press, Los Alamitos, CA, USA, 81-90.

[40] Chung-Horng Lung, Sonia Bot, Kalai Kalaichelvan, and Rick Kazman. 1997. An approach to software architecture analysis for evolution and reusability. In *Proceedings of the 1997 conference of the Centre for Advanced Studies on Collaborative research* (CASCON '97), J. Howard Johnson (Ed.). IBM Press 15-.

[41] Katherine *Isbister*, Patrick Doyle. Design and Evaluation of Embodied Conversational Agents: A Proposed Taxonomy. In Proc. of the AA-MAS02 Workshop on Embodied Conversational Agents: Lets Specify and Compare Them!, July 15-19, 2002, Bologna, Italy.

[42] D. Vrajitoru, NPCs and Chatterbots with Personality and Emotional Response. In: *Proceedings of the 2006 IEEE Symposium on Computational Intelligence and Games (CIG06)* (2006), 142-147

[43] Russell Beale and Chris Creed. 2009. Affective interaction: How emotional agents affect users. *Int. J. Hum.-Comput. Stud.* 67, 9 (September 2009), 755-776.

[44] Christoph Bartneck. 2003. Interacting with an embodied emotional

character. In *Proceedings of the 2003 international conference on Designing pleasurable products and interfaces* (DPPI '03). ACM, New York, NY, USA, 55-60.

[45] Michael Wooldridge and Michael J. Wooldridge. 2001. *Introduction to Multiagent Systems*. John Wiley & Sons, Inc., New York, NY, USA.

[46] Jonathan Gratch and Stacy Marsella. 2004. A domain-independent framework for modeling emotion. *Cogn. Syst. Res.* 5, 4 (December 2004), 269-306.

[47] Timothy W. Bickmore and Rosalind W. Picard. 2005. Establishing and maintaining long-term human-computer relationships. *ACM Trans. Comput.-Hum. Interact.* 12, 2 (June 2005), 293-327.

[48] Heidy Maldonado, Antoine Picard, and Barbara Hayes-Roth. 1998. Tigrito: a high-affect virtual toy. In *CHI 98 Cconference Summary on Human Factors in Computing Systems* (CHI '98). ACM, New York, NY, USA, 367-368.

[49] I. Mazzotta and F. de Rosis, Artifices for Persuading to Improve Eating Habits, AAAI Spring Symp. Argumentation for Consumers of Health Care, tech. report SS-06-01,AAAI Press, Mar. 2006, pp. 76–85.

[50] Winslow Burleson and Rosalind W. Picard. 2007. Gender-Specific Approaches to Developing Emotionally Intelligent Learning Companions. *IEEE Intelligent Systems* 22, 4 (July 2007), 62-69.

[51] Jorne Grolleman, Betsy van Dijk, Anton Nijholt, and Andrée van Emst. 2006. Break the habit! designing an e-therapy intervention using a virtual coach in aid of smoking cessation. In *Proceedings of the First international conference on Persuasive technology for human well-being* (PERSUASIVE'06), Wijnand IJsselsteijn, Yvonne de Kort, Cees Midden, Berry Eggen, and Elise van den Hoven (Eds.). Springer-Verlag, Berlin, Heidelberg, 133-141.

[52] Katherine Isbister. 2006. *Better Game Characters by Design: A Psychological Approach (The Morgan Kaufmann Series in Interactive 3D Technology)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[53] Li Gong. 2007. Is happy better than sad even if they are both non-adaptive? Effects of emotional expressions of talking-head interface agents. *Int. J. Hum.-Comput. Stud.* 65, 3 (March 2007), 183-191.

[54] Lin Padgham and Michael Winikoff. 2002. Prometheus: a methodology for developing intelligent agents. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1* (AAMAS '02). ACM, New York, NY, USA, 37-38.

[55] A. Pokahr, L. Braubach, and W. Lamersdorf Jadex Implementing a BDI Infrastructure for JADE Agents. Multiagent Systems, Artificial Societies, and Simulated Organizations, vol. 15, pp. 149-174, 2005.

ISBN: 978-988-19253-4-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)