Virtual Agents and Multi-modality of Interaction in Multimedia Applications for Cultural Heritage
A Case Study

João Neto[1,2], Claudia Ribeiro[1,2], João Pereira[1,2] and Maria João Neto[3]
[1]*INESC-ID, Lisbon, Portugal*
[2]*Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisbon, Portugal*
[3]*Faculdade de Letras da Universidade de Lisboa, Lisbon, Portugal*
{*joao.neto, claudia.sofia.ribeiro, joao.madeiras.pereira*}*@ist.utl.pt, mjneto@fl.ul.pt*

Keywords: Cultural Heritage, Multimedia Applications, Multi-modal, Embodied Conversational Agents, Multidisciplinary Approach, Immersive and Interactive Experience.

Abstract: Cultural Heritage encompasses a set of traditions and commodities inherited from our ancestors, and it is vital to convey and preserve them for the next generation. Cultural Multimedia applications and Serious Games form an important pedagogical and didactic medium, mainly for a young demography, which can learn while playing. In this paper is presented an innovative platform, named EI^{2}VA (Engine for Immersive Interaction with Virtual Agents), in which a virtual character, endowed with facial, corporal and behavioural animation can be integrated in multimedia applications. These multimedia applications, derived from the Fala Comigo project, which set the example for a possible use of this technology, preserving the historical and cultural contents indispensable for the users, based on a multi-modality of evident interaction. The resulting multimedia applications were used in a case study conduct with three different user groups (experts and non-experts) in different settings including classroom, museum and scientific conference.

1 INTRODUCTION

Nowadays, Cultural Heritage has been a vastly discussed subject (Petrelli et al., 2013; Antonaci et al., 2013). Like no other area, Cultural Heritage demands a multidisciplinary organization and effort in order to achieve its crucial goal - user learning (Anderson et al., 2010).

Without a solid foundation of contents, Cultural Heritage applications are just attractive pieces of software with no real educational use for the visitor. On the other hand, without a coherent and structured scientific approach, the technological systems will not meet the demanding requirements for this type of media. Moreover, this area demands a unique visual presentation, that engages visitors. The success of Cultural Heritage systems deeply relies on balancing these three core areas.

To achieve this, there must be a commitment produce renewed content. Is necessary to develop new cultural discourses, aimed at different age groups and that create awareness of how important it is to preserve artistic heritage. These studies must be presented according to the modern lines of dissemination, based on emerging technologies that capture the attention of these audiences, establishing a framework that allows both the dissemination and the enjoyment of heritage.

Exploring these new potentials, the Fala Comigo project [1] (Rego, 2004) explored different multimedia applications which, combined several technologies with a well-structured historical narrative, taking as a common factor the usage of interactive virtual agents. The main aim was to create something distinctive that entices and fascinates the public.

Consequently, the work described in this paper focused on the implementation of and innovative animation framework, named EI^{2}VA (Engine for Immersive Interaction with Virtual Agents), designed to optimize the performance of embodied conversational agents and their integration in multimedia systems. The scope and potential success of the EI^{2}VA framework was the object of a careful analysis using the environment of the Palace of Monserrate, in Sintra Portugal which served as a case study for the project. We evaluated how multimedia applications can influ-

[1]http://www.falacomigo.pt/en/

Neto J., Ribeiro C., Pereira J. and Neto M.
Virtual Agents and Multi-modality of Interaction in Multimedia Applications for Cultural Heritage - A Case Study.
In *Proceedings of the 10th International Conference on Computer Graphics Theory and Applications (VISIGRAPP 2015)*, pages 446-453
ISBN: 978-989-758-087-1

ence the assimilation of cultural contents and make an impact of the visitors' experience and immersion. In this sense, tests with different types of users, such as interviews, questionnaires and in-situ demonstrations were made.

Section II describes the case study conducted as well as user characterization where as the results and discussion presented in section III. We finalize with conclusions and future work (Section IV).

2 EI^{2}VA SYSTEM

The aim of this work was to create an innovative framework for virtual agents, completely integrated in a spoken dialogue system, where the purpose was to take the user on an immersive and multi-modal journey, where realism and interactivity are crucial factors.

2.1 Architecture

The organization of the EI^{2}VA framework is divided into three main blocks: Application, Animation Engine and Spoken Dialogue System. The Application Block can be seen as a standard multimedia support, with its features, functionality, and isolated objectives. In this manner it was intended to add a new interactive and communicative paradigm. This change forced the introduction of an innovative tool, where the inclusion of Embodied Conversational Agents, realistically animated and communicatively supported by an SDF, presents itself as a differentiating quality factor. Thus, the Animation Engine is responsible for conceding the realism by creating credible facial, emotional and bodily animations. The Spoken Dialogue System block is responsible for managing the interaction and knowledge domains of the virtual agent.

2.1.1 Application Block

From the high-level architectural introduction made previously, it is possible to identify the existence of three main functional blocks. Each of these blocks has a set of internal modules that contribute to the integration of the EI^{2}VA framework into any multimedia application.

Beginning with the Application Block, which has two essential management components: Client Application Manager and Recognition Manager. The Client Application Manager is the entry point for the system's execution. As primary responsibility, it makes the connection between the application, the Animation Engine and Spoken Dialogue System, controlling all activity of this environment. In addition to the features of communication management and enforcement, this module has other important responsibilities. Specifically, it serves as a trigger to the interactive process. As such, after creating the user's session structure, it initiates all the remaining components of the Application Layer. First it signals if the application is properly initiated and ready to start the recognition process, transferring the execution flow to the Recognition Manager.

The Recognition Manager is responsible of coordinating the whole recognition process. Specifically, when the Client Application Manager completes the initialization process, changes its state and informs the Recognition Manager that the recognition process can begin. The recognition process is continuous and cross-framework, always active, and only stopped when the agent is speaking, being restarted immediately when it finishes. This module is therefore separated from the animation engine, since their operation is independent of the activity of the later.

2.1.2 Animation Engine

Regarding the Animation Engine block, it consists of a set of unique modules, responsible for the entire animation aspects of the EIVA framework. In particular, this engine is formed by the following functional components: Visemes Manager; Emotions Manager, Facial Expressions Manager, Human Gestures Manager, Body Animations Manager, Non-verbal Behaviour Manager; Data Persistence Manager and Animation Module. The Visemes Manager prime responsibility of mapping the phonemes of any supported language into visemes. Then, it takes the resulting viseme set and converts it to visual animations belonging to the virtual agent. As input data, this component receives a data structure composed by the list of phonemes and their duration times, sent directly by the Spoken Dialogue System.

The Emotions Manager is responsible of mapping a set of human emotions into visual animations of the virtual agent. As input data, this module receives a data structure containing three separate lists: a list of emotions to generate; a list with the respective duration of each of these emotions and a list with particular intensities of each of the above poses.

The Facial Expressions Manager primary function is to map a set of facial expressions into visual animations that can be used by the virtual agent. As input data, this module receives a data structure containing three separate lists: a list of the facial expressions to generate; a list with the respective duration of each of

these expressions and a list with particular intensities of each of the above poses.

The Human Gestures Manager, in its turn, maps a set of human gestures in visual animations that can be used by the virtual agent. As input data, this module receives a data structure containing three separate lists: a list of the human gestures to generate; a list with the respective duration of each of these gestures and a list with particular intensities of each of the above poses.

The Body Animations Manager has the responsibility of defining a set of uncharacterised animations, which can be used at alternate times of the interactive process. Here we find oratory animations or animations of complex human movements. Currently there is no concrete and generally accepted specification to indicate which set of essential bodily animations a virtual agent should have. This led us to separate this module functionally and structurally. In this manner, its control is independent from any other animation manager.

The Non-verbal Behaviour Manager takes the responsibility of recreating non-verbal behaviours and associates them with text responses to be reproduced by the virtual agent. These behaviours can be created to encompass any of the animations related to emotions, gestures or facial expressions.

Any previously described animation can be defined as an interactive event which occurs over a period of time, being connected to a given sequence of facial or body animations and persisted in the Data Persistent Manager. As such, it must be possible to reference these animations easily through a standard specification. This abstract description is the solution to simplify the structural mapping between what is sent by the Spoken Dialogue System and what is featured in each of the presented managers. Thus, to better reference the different data structures received in each behavioural animation module, the descriptive language for virtual human animations VHML - Virtual Human Markup Language [2] was used.

2.1.3 Spoken Dialogue System

The Spoken Dialogue System consists of a set of components that enhance the multi-modal and interactive facets of the system. Through the integration of semantic processing tools, speech recognition and synthesis engines, it is possible to transform a virtual agent in an educational and communicative vehicle. To achieve an interactive form of communication it was necessary to innovate in different aspects, mainly in the creation and structuring of the Relational Knowledge Base. Specifically, it was necessary to develop something more prominent than a simple database, associating questions with respective answers. This layer is generally coordinated by the management module - Dialogue Manager, which takes the responsibility of structuring all communication and processing flow, liaising with the Animation Engine and the Client Application. The Speech Recognition Engine integrates the AUDIMUS which is a hybrid speech recogniser that combines the temporal modelling capabilities of Hidden Markov Models, with the discriminative pattern classification capabilities of multilayer perceptrons. This automatic recogniser can be used for distinctive tasks, based on a common structure, with different components. The acoustic models were adapted for a microphone input device. The system is using a different language model for each specific application scene in order to have a limited number of words in the vocabulary. With this approach the recognition percentage is much higher. However, the language model can have an increased complexity for each knowledge base driven interactions.

The Language Interpretation Module is responsible for extracting the intentions of the user's utterances. The visitor can interact with the applications by touching the screen or by speech commands. When real-time interaction is intendant, a suggestion table of questions is shown to the user to help her/him steer in the right direction. In a cultural exhibition case study is normal for a virtual agent interaction framework to receive many questions not stored in the current knowledge base. To avoid quick user detachment, we suggest possible questions to boost the interaction sequence. The visitor can choose a specific question using the touch screen interface or by reading the question. When this module receives input on the touch screen, it sends the chosen question to the Relational Knowledge Base module.

The Relational Knowledge Base maps questions to the appropriated answers. When a question arrives to the module, an answer is chosen and then sent to the Dialogue Manager.

The Dialogue Manager consists of several main modules, including the Behavioural Agent (BA) which has the responsibility of managing all the dialogue process. Frames are used to represent both the domain and the information collected during the interaction with the users. The Speech Synthesizer Engine integrates a TTS module (DIXI) which is a concatenated based synthesiser. This framework supports several voices and two different types of unit: fixed length units (such as diaphones), and variable length units. This latter data-driven approach can be

[2] http://www.vhml.org/

fine-tuned to a limited domain of applications scenes, by altering the design of the corpus.

2.2 Information Flow

Given the previous structural division, we can summarize the interactive flow organization through a state machine (see Fig 1) that mirrors the different presented phases.

At the beginning of the execution the system is in INIT state, where the different modules managers are initiated, beginning with the Application Block soon followed by the spoken dialogue layer. After this phase, the system is ready to start the recognition process, reaching the LOAD state. When this loading process is complete, the client application is ready to receive requests, going to the ACTIVE state. When the user interacts with the system in the form of a question, this is done transferring the execution flow to the Spoken Dialog System. The system enters a processing phase in which the current state - PROCESSING - reflects that. When this process ends, the result is sent to the Animation Engine that will generate an animated - ANIMATING state. The Application layer is responsible for initiating the display of the prior sequence to the user. When the agent finishes its response, the system waits for a new interaction returning to the ACTIVE state. When the interaction ends, the system returns to its initial state- INIT. The following image shows this same state machine.

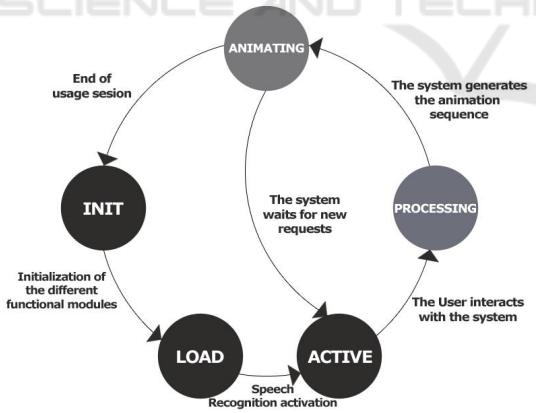

Figure 1: State Machine of the EI^{2}VA Framework.

3 FALA COMIGO PROJECT A CASE STUDY

It was decided to conduct an appraisal of this work, which consisted in evaluating in an integrated way, each of its components. This evaluation was based in a set of structural methodologies introduced by

Figure 2: At the time of the Cook Famiy Room example.

(Bernsen and Dybkjær, 2004) and (Fonseca et al., 2012). Collecting this information allowed doing a deeper reflection on fundamental aspects of the multimedia system produced for the Fala Comigo project [3] (Rego, 2004).

With the reference of Monserrate Palace in Sintra, the Fala Comigo project [4] (Rego, 2004) created different multimedia products, assuming as common denominator the virtual agents. These products obeyed, in terms of content, to a selection of information which is relevant to different types of audiences using different levels of language. Enhancing content accessibility increases the potential of one of the most fundamental objective of a Cultural Heritage application - user learning. By using complementary cultural contents, a simple user interface was designed, where the contents can be clearly conveyed and absorbed. These applications are group by three distinctive types: Multi-touch Informative, Multi-touch Interactive and Serious Games (Abt, 1970; Michael and Chen, 2006; Susi et al., 2007; Ribeiro et al., 2013). An example of one of the multimedia applications can be depicted in Figure 2.

The first important step of the evaluation process was to correctly identify the user groups that should test the system in its development phase. Following this premise, we formed an evaluative group for each of the functional areas presented in this work. This structure proved to be of utmost importance when drawing conclusions about the developed functional environment.

3.1 User Characterization

This step was conducted in a structured and methodological way, since it is crucial to select the right users to test the system. In particular, special steps were taken throughout this process, selecting heterogeneous assessment groups, skilled and unskilled, fo-

[3] http://www.falacomigo.pt/en/project/
[4] http://www.falacomigo.pt/en/project/

cused on examining particular parts of the system (Taras, 2005).

Follow-up Groups

Since the start of the development phase, both the Animation Engine and the Spoken Dialogue System, needed a specialized follow-up group responsible for analysing each iteration of these two parts. Separately, another team was assigned to monitor the multimedia applications creation process, as mentioned previously. The group responsible for the structural and functional analyse of the evolution of the Animation Engine and Spoken Dialogue System had the following composition: (a) 1 specialist in animation; (b) 1 specialist in spoken dialogue systems; (c) 2 expert users; (d) and, 3 non-expert users.

In parallel, the group with the task of verifying that aesthetic and functional evolution of the multimedia applications was composed as follow: (a) 1 specialist in historical and artistic contents; (b) 1 specialist in spoken dialogue systems; (c) 2 expert users; (d) and, 3 non-expert users.

From the organizational structure presented, is possible to observe that the two groups were comprised of three categories of users. This approach promotes the possibility of receiving different types of opinions and getting a better perspective on the strengths and weaknesses of the system. On both groups there are specialists, focused on matters directly related to fundamental aspects of the exposed functional blocks. Additionally, two experts users were assigned to each team that had both an academic and professional experience in this context. Due to their background, this users were able to provide a distinct and constructive view of all the environment. To promote a new critical and impartial vision about the various functional components, we also selected three non-experts users with no previous knowledge of the areas in question.

During the development stage, the two groups collaborated with us through different stages of formative evaluation, analysing the main functional aspects of the created system. In all these analytical moments, technical and design flaws were found. This iterative discussion process resulted in the improvement of the components discussed above. At the end of this step, the history log build during the different analysis phase was reviewed. This resulted in a summative evaluation of the complete process.

System Users

After concluding the development phase, it was necessary to find new users, without any knowledge of the work in progress. The selection of users followed the same methodological principles of the development phase. In terms of analysis, the focus was on the quality and effectiveness of the transmission of educational and historical content as well as the sociological ability of the virtual agent as a teaching and guiding vehicle during a cultural visit.

Given the interactive facet of the Fala Comigo project [5] (Rego, 2004), it was decided to validate the capabilities of this system in different environments. Initially, the system was presented at the international conference: ECLAP2013, 2nd International Conference on Information Technologies for Performing Arts, Media Access and Entertainment, Porto, 8-10 April 2013. Here, the main focus of the evaluation was the Animation Engine and the Spoken Dialogue System, via the technical skills of the participants. In short, with this demonstration, we sought the opinion and acceptance of people with expertise in the identified areas.

In a second phase, the pedagogical capabilities of the interactive agent were analysed. The presentation and use of educational multimedia content alone, very often has a low receptivity in the target audience. Without a well-structured conductive line, where the virtual agent assumes the role of a guide, it is difficult to achieve a correct transmission and assimilation of historical contents. As such, it was necessary to assess the actual pedagogical relevance of the produced applications. Therefore, it was decided to test the different applications with two groups of students, one of basic education (8th grade) and other higher education (3rd year of the degree in Art History). The context, the mode and manner of this assessment will be explained in the next section.

To complete this important step, the applications were also evaluated in-situ. The Monserrate Palace is a monument with very particular characteristics and is visited by many foreign tourists. Almost empty today, the historical contents need to be the bridge between the past and the present. The virtual agents, as recreated historical figures, are responsible for guiding the visit, on one hand, and for being storytellers, on the other. This phase lasted a week and we were able to collect all the information necessary regarding the created applications.

4 RESULTS AND VALIDATION

In this section is summarized the results of the evaluation process. In order to maintain a descriptive con-

[5] http://www.falacomigo.pt/en/project/

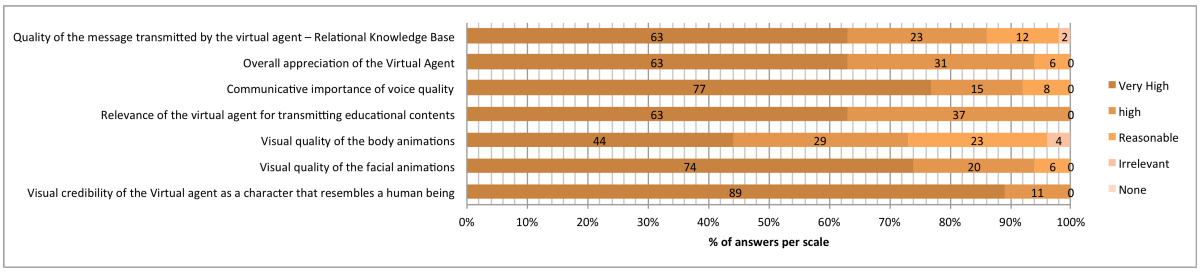

Figure 3: Technical Analysis Results.

sistency, we begin with the description and analysis of the results obtained in the international conference ECLAP, followed by the assessment of the student groups test phase and ending with the assessment of in-situ tests.

4.1 Technical Analysis

The international conference ECLAP was chosen to be the first demonstration attempt, disseminating the Fala Comigo project [6] (Rego, 2004) in a scientific environment, paying particular attention to the interaction with the Virtual Agent. The questionnaire used to collect information was administered to 15 users and was focused on the evaluation of the Animation Engine and the Spoken Dialogue System. Particularly, the objective was to obtain new opinions and also validate the chosen implementation options. The questionnaire was divided into three distinct and well defined sections: demographic analysis, analysis Animation Engine and analysis of Spoken Dialogue System.

The first phase of the evaluation comprehended a group of demographic questions for better characterization and contextualization of the respondents. The high level of education and technical skills of the participants provided a new vision in the two previous highlighted components. Moreover, the wide variety of participants nationalities was significant when assessing the quality of the communicative message of the virtual agent, with only three available languages, including Portuguese, English and Spanish. After answering the demographic questionnaire each user was submitted to guided interaction with each of the multimedia applications. At the end of interaction, each user answered a questionnaire composed of seven likert-scale questions focused on analysing the real impact of the virtual agent on their experience. Specifically, it was requested feedback in terms of visual credibility, both facial and emotional animation wise. The summary of the results obtained in this questionnaire are presented in Figure 3.

[6]http://www.falacomigo.pt/en/project/

93% of respondents consider that these animations reproduced faithfully and, with excellent quality, the expressions and emotions of a human being. The communicative importance of the virtual agent is another relevant aspect to take into consideration. The agent, many times, needs to be a communicator and a disseminator of cultural contents in a historical environment. The text-to-speech engine needs to have access to realistic voices. Each virtual agent should have a particular voice that needs to be convincing and feel natural. 80% of respondents qualified this qualitative aspect with Very High and 20% believe that the credibility conveyed by the virtual character to High. The used voice, together with a spoken discourse carefully outlined, is fundamental in this interactive process. Overall, the receptivity of respondents was quite acceptable, with 87% of respondents to rate the overall experience with the Virtual Agent with Very High. Finally, the last section of the analysis focused on issues related to Spoken Dialogue System. The ECLAP conference provided an environment suitable to test aspects concerning communicative quality of virtual agent dialogues, mainly the contents of the Relational Knowledge Base. Again, 87% of the inquired users were very satisfied with the assertiveness of the virtual agent's cultural message.

4.2 Sociological Analysis

After ECLAP, we looked for different points of view outside of the academic context. Specifically,the system was tested by two distinct groups of respondents. The main aim was to analyse the sociological, educational and communicative ability of the virtual agent. In this respect, the multimedia applications were tested with a group of students from the 8th grade (total of 20 students) and with a university class from the Art History course (total 28 students). This test was comprehended in two distinct phase: firstly the user listened to an oral presentation of cultural contents of the Palace of Monserrate and its history. After, the students answered to a quiz regarding educational questions about the presentation. On the

second phase, the students interacted with the multimedia applications for a period of 15 minutes. Finally, they answered to a quiz similar to the one they answer previously. These knowledge questionnaires were provided by experts and pre-test and post-test, although different, had the same level of difficulty. The summary of the results can be depicted in Figure 4 and Figure 5.

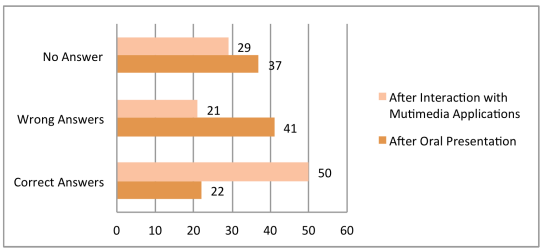

Figure 4: Art History Course Undergraduate Students Results.

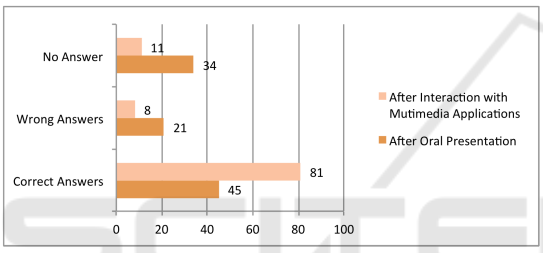

Figure 5: 8th Grade Results.

First, we have to point out an important aspect in relation to the environment in which the tests were conducted. Regarding the university students group, the underling process was conducted in the classroom. On the other hand, with the 8th grade students group the tests were conducted at the monument during a study visit. Even with specific knowledge in art history, the results of the undergraduate class show that the user location paradigm of is a factor to take into account. Only 22% correct and 41% wrong answers, adding up to 37% of unanswered questions.

The initial results of 8th grade class were a little more balanced with 45% correct answers.These results indicate that the presence of the visitor at the historical site has a measurable impact in the assimilation of the historical and educational contents. Particularly, this is reflected in the lower percentage of wrong answers and not answered, 21% and 34% respectively of the 8th grade students group. After the interaction phase, the second quiz showed a clear improvement of the respondents. In the 8th grade students we can see a 34% increment for the correct answers to 81%. In the undergraduate students, the correct answers percentage was maximized by 28 points to 50%. These are positive results that show some promise in terms of the sociological, educational and communicative capacities that the virtual agent offers.

4.3 In-Situ Analysis

To conclude the test phase, it was decided to conduct a week of field-testing at the Palace of Monserrate. Both visitors and monument's staff were invited to experience the multimedia applications. Here, with a sample of 47 elements, we aimed to test, not only the technical aspects of the Animation Engine and the Spoken Dialogue System, but also aspects related to the usability and design of the applications. In terms of the lastly referred aspects, the results were very interesting, which revealed the importance of the virtual characters being directly integrated in multimedia applications. The summary of the results can be depicted in Figure 6.

After the normal demographic analysis phase, feedback was requested regarding the Design and Usability facets of the multimedia applications. Specifically, the aim was to hear the user's opinion in terms of the presented cultural knowledge and the acquisition process of those historical contents. 77% and 68% of respondents scored as Excellent the structural aspects of Design and Usability, respectively. In terms of exposure and apprehension of knowledge 72% and 57% indicated as Excellent how the historical and educational contents are presented and acquired, in that order.

Regarding the evaluation of the Animation Engine, the respondent's views support the results obtained during the evaluation process carried out at the ECLAP conference. Specific technical aspects, such as the visual impact of the facial animations (Very high - 74%), communicative importance (Very High - 63%) or voices quality (Very high - 77%) were quite interesting in a non-specialized public. The results are less positive in the visual evaluation of the body animations, with 44% of respondents to evaluate these as Very High and 29% as High. 89% of respondents indicated with Very High the visual credibility of the virtual characters, higher value than 80% registered at the ECLAP conference. Finally, 63% of respondents scored as Very High the overall quality of the animation engine.

The final evaluation phase included the quality of the verbal message conveyed by the virtual characters, structured in the Relational Knowledge Base. 63% and 23% of users consider this as Very High and High respectively.

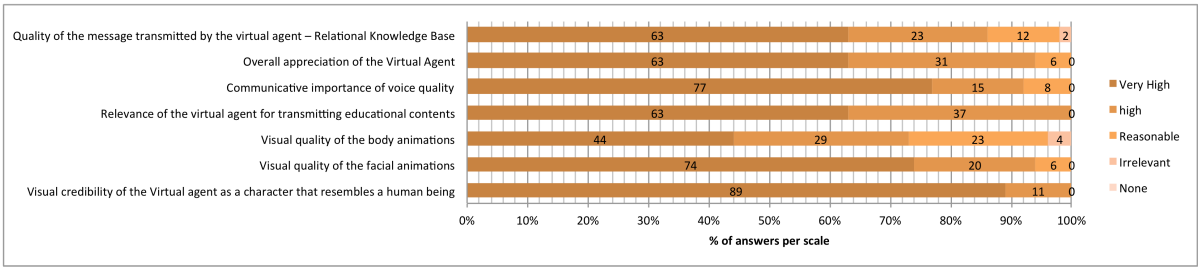

Figure 6: In-Situ Analysis Results.

5 CONCLUSIONS

In conclusion, it is necessary to highlight the current high demand for promoting Cultural Heritage. There is a firm belief that for a proper and attractive transfer of cultural information, the development of multimedia applications that heavily rely on realistic Embodied Conversational Agents is mandatory. This is an on-going research project, since in the past the delivery of the cultural message was compromised due to poor design and simplistic agents.

In general users believe that the agent animations and dialogue interaction was very credible. In particular, both experts as well as non-experts expressed the importance of the virtual agent to convey cultural heritage. Moreover, the level of learning was higher both among 8th grade students as well as undergraduate students after interacting with the multimedia applications.

Finally, as a future direction for this work we believe that developing a centralized library of virtual agents and respective animations that could be re-used by different application for different contexts could help lower the cost of developing this kind of applications.

ACKNOWLEDGEMENTS

This work was supported by FCT (INESC-ID multiannual funding) under the project PEst-OE/EEI/LA0021/2013. The authors also would like to acknowledge to the European funded Project Games and Learning Alliance (FP7 258169) the Network of Excellence (NoE) on Serious Games.

REFERENCES

Abt, C. (1970). *Serious Games*. University Press of America.

Anderson, E., McLoughlin, L., Liarokapis, F., Peters, C., Petridis, P., and de Freitas, S. (2010). Developing serious games for cultural heritage: a state-of-the-art review. *Virtual Reality*, 14(4):255–275.

Antonaci, A., Ott, M., and Pozzi, F. (2013). Virtual museums, cultural heritage education and 21st century skills. *Learning & Teaching with Media & Technology*, page 185.

Bernsen, N. O. and Dybkjær, L. (2004). Evaluation of spoken multimodal conversation. In *Proceedings of the 6th International Conference on Multimodal Interfaces*, ICMI '04, pages 38–45, New York, NY, USA. ACM.

Fonseca, M., Campos, P., and Gonçalves, D. (2012). *Introdução ao Design de Interfaces*. FCA - Editora de Informática.

Harlen, W. and James, M. (1997). Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in Education*, 4(3):365–379.

Michael, D. and Chen, S. (2006). *Serious games: games that educate, train and inform*. Thomson Course Technology.

Petrelli, D., Ciolfi, L., van Dijk, D., Hornecker, E., Not, E., and Schmidt, A. (2013). Integrating material and digital: a new way for cultural heritage. *interactions*, 20(4):58–63.

Rego, S. (2004). Fala comigo agente virtual conversacional e emocional. Master's thesis, Instituto Superior Técnico, Lisbon, Portugal.

Ribeiro, C., Antunes, T., Monteiro, M., and Pereira, J. (2013). Serious games in formal medical education: An experimental study. In *Games and Virtual Worlds for Serious Applications (VS-GAMES), 2013 5th International Conference on*, pages 1–8.

Susi, T., Johannesson, M., and Backlund, P. (2007). Serious games - an overview. Technical Report HS-IKI-TR-07-001, School of Humanities and Informatics, University of Skövde, Sweden.

Taras, M. (2005). Assessment–summative and formative–some theoretical reflections. *British Journal of Educational Studies*, 53(4):466–478.