Modelling Of More Realistic Intelligent Virtual Agent in Virtual and Mixed Reality

Dilyana Budakova[1, a)], Veselka Petrova-Dimitrova[2, a)], Velyo Vasilev[3, a)] and Lyudmil Dakovski[4, b)]

[1]*Technical University of Sofia, Branch Plovdiv, Plovdiv, Bulgaria*
[2]*Technical University of Sofia, Branch Plovdiv, Plovdiv, Bulgaria*
[3]*Technical University of Sofia, Branch Plovdiv, Plovdiv, Bulgaria*
[4]*European Polytechnic University, Pernik, Bulgaria*

[a)] *dilyana_budakova@yahoo.com*,
[b)] *vesi_s_petrova@yahoo.com*,
[c)] *velyo.vasilev@tu-plovdiv.bg*,
[d)] *l.dakovski@gmail.com*

Abstract. This article presents different options for modelling and realization of realistic intelligent virtual agents (IVA) in virtual and mixed reality. In relation to that a natural language dialogue with the users has been implemented; information about the preferences of the users is obtained, as well as their emotions, goals and the way, in which they prefer to achieve their goals. Emphasis is placed on the method for modelling the visualization of IVA; the facial expressions and emotions, which can be conveyed; the gaze of the eyes; the mouth and lip-sync with the spoken words; the gait and gestures. Some results of conducted user dialogues when solving a particular problem; the selection of a goal and a way of achieving it are discussed. Keywords: Intelligent Virtual Agent, Modelling, Agent Architectures, Reinforcement Learning, 3D Visualization, Virtual Reality, Mixed Reality, Virtual Rapport

INTRODUCTION

The development of society, services and the industry places ever higher demands on technology to find solutions to increasingly complex problems. It is often necessary for different technologies to be used together in order to solve a particular problem.

For example the agent-based modelling and simulation paradigm allows intelligent agents with different implementations to interact between each other, as well as with their environment. Every scenario is viewed as a complex adaptive system, which can be studied and analyzed. Modelling and simulation based on "Virtual Reality" build upon this paradigm by giving a person the option to interact with visually represented three-dimensional, real or imaginary systems [1] [2]. Users must wear a virtual reality headset and use gloves or motion controllers to interact with the simulation in real time. The user can be "immersed" in the virtual computer-generated environment, but loses touch with the real world. On the other hand "Augmented Reality" technology allows for a new visualization method, in which virtual content is added to the real world [1] [2]. With this method, however, a link between the virtual and real content is not created. The newest and most promising technology is "Mixed Reality" [1] [2] [3]. This technology combines the real and the virtual worlds and allows a real-world object to interact with a virtual object when a given scenario is being executed. Every Mixed Reality system is characterized by: an ability to combine between a real world object and a virtual one; real-time interaction; finding correspondence (mapping) between the virtual and the real object; execution of user scenarios. "Virtual Reality", "Augmented reality" and "Mixed Reality" technologies are considered strategic [4].

The Artificial Intelligence scientific community encourages research work with a cross-point between virtual agents, robotics and psychology. The synergy between agents with different implementations - from virtual agents, to social robots is being researched [5]. The new algorithms and models need to be able to be applied to agents with different implementations, so that they can interact when solving a given problem and as a result solve it together. These requirements are met by a wide range of problems (scientific, social, and industrial).

Research is also focused on studying and analyzing knowledge representation methods, decision-making algorithms, planning algorithms, control algorithms, methods for training and machine learning [6] [7] [8] (Reinforcement learning, Learning from Demonstration paradigm or Imitation learning and their improvements), probabilistic algorithms and models (Belief networks), Probabilistic causal networks, work with uncertain knowledge, connectionist methods and in particular training of Deep neural networks and Recurrent neural networks RNN [9] [10] [11] [12]. The goal is also to improve them and offer new algorithms for managing and planning the behavior of intelligent agents as well as to study them when applied in virtual, mixed and physical reality.

The joint use of these methods and technologies allows for the realization of more realistic IVAs, it also allows the intelligent agents to solve problems in the industry, to provide assistance to people with disabilities or help people in their daily lives, to participate in interactive games, to perform actions, which require gripping objects and using tools, to prepare food and many other intelligent activities. They are the basis for the construction of "Smart Cities", "Smart Business and Industrial Centers", "Smart Infrastructure", "Smart Homes" and social welfare.

This article goes over the possibilities for modelling and realization of realistic intelligent virtual agents (IVA) in virtual and mixed reality. In order for IVAs to be realistic, they have to carry out a dialogue with the users in a natural language; they have to understand the preferences of the user, their emotions, their goals and the way, in which they prefer to achieve these goals. It is necessary for IVAs to have a unique appearance; to be visualized in 3D space; to express emotions; to control and direct their eyes towards their interlocutor; to have unique facial expressions, gestures and lip-sync. They must be able to choose an appropriate goal and the most acceptable way to achieve it. The realization of all the possibilities will be discussed in this article.

MODELLING OF A MORE REALISTIC INTELLIGENT VIRTUAL AGENT IN VIRTUAL AND MIXED REALITY

The design of intelligent agents involves formalizing theories and data on human behavior. The acquired models are integrated into IVA (Figure 1) and its behavior is evaluated. This requires the use of techniques derived from psychology. In turn each one of these steps can be fundamental for psychological research. That is to say the connection between computer science, informatics, ICT, mechatronics, robotics, psychology and 3D modelling is clear. This interdisciplinary gives research freedom and an opportunity to gain new knowledge and carry out new fundamental research.

It is necessary to have an understanding of the technologies for building the space of the Virtual reality and Mixed reality, such as the Optical see-through and video-see-through technologies; calibration models; object recognition techniques; techniques for tracking objects (Object Tracking- sensor-based, vision-based, and hybrid) which are needed to position the real and virtual objects so that they can interact between each other; Registration and Mapping of the virtual model in the real world; visualization and rendering; data management strategy etc. [1] [2] [3]. An understanding of their program packages and software, and their capabilities and features is necessary.

Interactive systems with intelligent conversational agents include the following technologies: speech recognition, natural language understanding, speech synthesis, technologies for modelling facial expressions and emotions; facial animation technologies [13] [14] [15].

One IVA architecture is proposed in reference [16]. It has the following basic blocks: memory block; a block for marking the states, best suited for the execution of the task and the connections between them; block for explanations and a training block, which can include different learning methods. Reinforcement learning is used for the experiments described in this article. This architecture has been further developed with two additional blocks, which are a block for the implementation of a natural language dialogue with the user and a block for the visualization of both IVA and the environment. They are discussed in next sections.

Using Windows Desktop Speech Technology

The assets, used to implement the speech recognition ability are Windows Desktop Speech technology and the namespace System.Speech.Recognition [17]. This software offers a speech recognition infrastructure; digitalizes

acoustic signals and recovers words and speech elements from an audio input. Algorithms are developed to identify specific phrases and speech patterns and to control the behavior of the provided infrastructure during operation. Some of the useful classes are Class Choices, Class Grammar, Class GrammarBuilder, and Class SpeechRecognitionEngine. In order for a program with a speech recognition ability to be developed the following is done: grammars are created; the output, which the speech recognition engine produces is initialized, controlled and interpreted; the events which generate a reaction are used. An example grammar is defined in an XML file and the classes GrammarBuilder and Choices are used. In this way grammars with low to medium complexity are programmatically created.

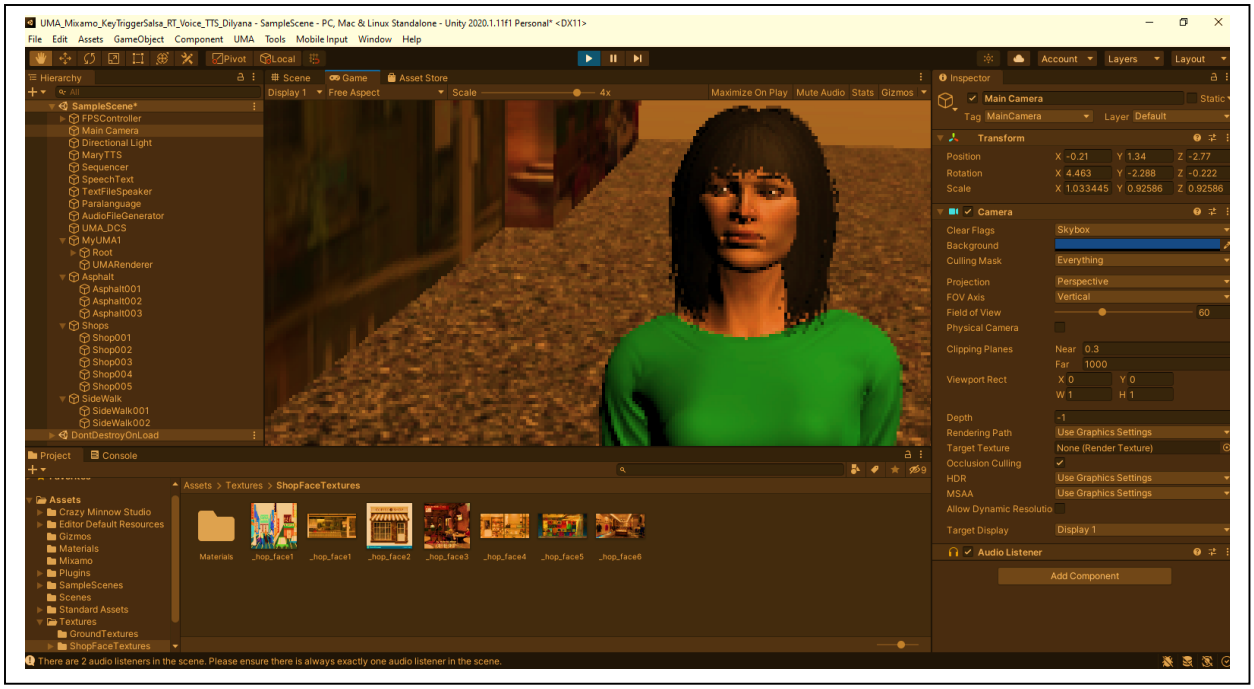


FIGURE 1. Modelling of a lifelike virtual conversational agent and a 3D virtual environment. Visualization of the head of the agent; eye direction control; synchronizing the movement of the mouth and lips with the spoken words; control of the emotions expressed by the face. Visualization of the movement of the body of the agent – its gait and gestures.

Text-To-Speech Solution for Unity. Lip-sync and Face Emotions.

In a Unity project one of the best solutions for activating text-to-speech is the RT-Voice Pro package [18]. TTS-voices, are already integrated in the system are used to pronounce written text during program execution. The package allows for change of speed, pitch and volume of speech. SSML – speech synthesis markup language and EmotionML – emotional markup language are supported. There are more than 1000 voices, which can be used.

In order to model a unique facial expression and facial gestures, a good solution is using the SALSA Lip-sync V2 package [19]. It allows for a 3D synchronization of lip movement with the spoken words; it allows for control of head and eye movement and expression of emotions. The Audio-dialogue files are processed in real time by using look-ahead technology.

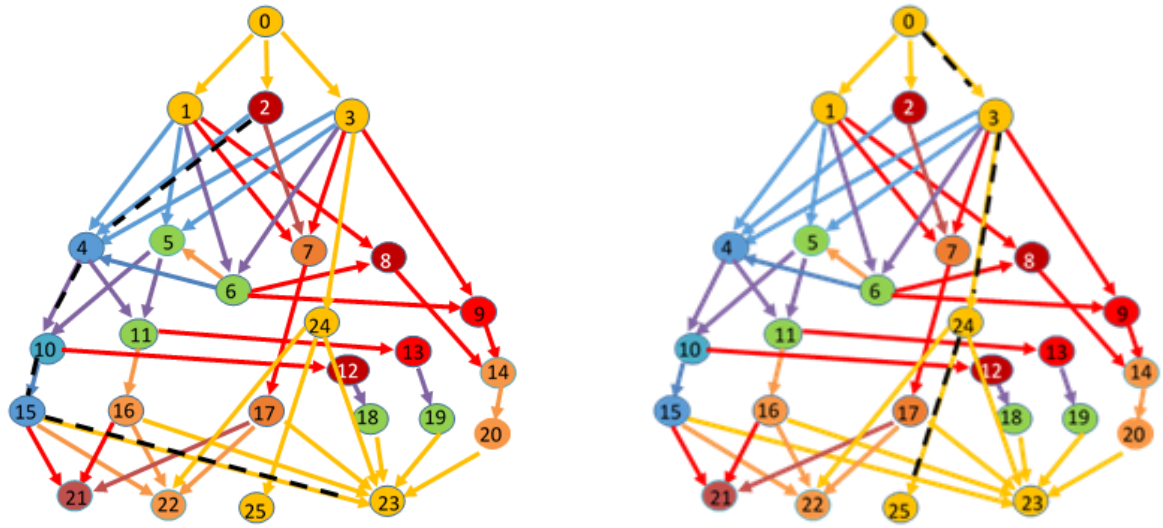
For the modelling and visualization of autonomous non-player characters (NPC), with a unique appearance and an intelligent behavior, it is appropriate to utilize the UMA2 – Unity Multipurpose Avatar system [20]. NPCs modelled with UMA2 can express emotions using their face, use gestures and lip-sync with the words they pronounce. With the help of Mixamo they can perform unique movements, such as walking, the ability to sit, look around, run, jump, etc. Mixamo is the first character animation service developed and provided by Stanford University [21] [22] [23]. Mixamo offers packages with thousands of animations, the most fitting of which can be

chosen for any project, game or application. These motion animations are ready to be imported into the developed application and contribute to an even more realistic presentation of the modelled NPC.

Strategy for carrying out the dialogue

A software system with a virtual agent model has a dialogue strategy and rules for shortening questions. The knowledge extraction can be automated through an algorithm with the following steps [8]: 1) all attributes, their possible values and the goals, which can be achieved are entered; 2) the software system combines all attributes and their values, and the user indicates to what extent the goals are achieved for each combination; 3) all contradictions are removed. However this algorithm is very complex and time consuming. For example obtaining the opinion and intentions of the user for purchasing a given item can be simplified by asking a series of questions with a limited number of answers. For instance some available answers may be: yes, no, this is acceptable, this is not acceptable, this is certain, I would like it very much, I would not like it. Such a dialogue is similar to the dialogue, which takes place, for example, when shopping in a store between a customer and a shop assistant or when buying residential property and talking to a broker, etc.

.



(a) (b)

FIGURE 2. (a) Chosen by IVA path to buying a one-room apartment by raising 20% of the cost and 80% from a mortgage credit. (b) Chosen by IVA path to buying a big flat with own funds.

EXPERIMENT FOR UNDERSTANDING THE PREFERENCES AND INTENTIONS OF THE USER

For the purposes of the experiment two conversations with two clients, who wish to buy residential property, were held. The model of the environment is presented in Figure 2 (a) and (b) by a graph, consisting of nodes and edges. The model covers the properties available for purchase – nodes 21,22,23,25; The different ways of purchase - with a bank loan – nodes 10, 11, 15; by raising funds in other ways – nodes 4,5,6,7; by purchasing with available funds – node 24. Each possible action for achieving a given goal is represented as an edge in the graph. The actions evoke different emotions in the users. These emotions are represented by a variety of colours. For example the colour green shows that the action evokes a sense of security; actions, denoted by the colour red evoke a sense of Panic, anxiety and dissatisfaction; the colour yellow represents actions which evoke Joy and enthusiasm; actions, denoted by the colour blue, represent actions, which evoke calmness and hope. (Figure 2 (a) and (b)).

From the conducted dialogues it is discovered that one of the users wishes to buy property for business purposes. He has all the necessary funds to do it. He prefers a large three-room flat. The second user does not have the

necessary funds to purchase residential property. He is looking for a small apartment with 80% of the cost covered by a bank loan. He enjoys travelling and prefers to travel to a nearby city until he raises 20% of the price of the desired property. As a result of these conversations with the users, the IVAs choose the most fitting properties and most acceptable ways for consumers to purchase. Figure 2 (a) and (b) show with a dotted line the sequence of actions, which the users need to take in order to purchase their desired property. They are respectively a large flat for the first user – node 25 (Figure 2 (b)) and a small one-room apartment for the second user – node 23 (Figure 2 (a)).

The first user will travel for several years to a nearby city to go to work every day and will take care of the other properties he owns. After raising 20% of the price of the property he will take a consumer credit for the remaining 80% and will purchase the desired small apartment (Figure 2 (a)). The second user will simply pay the price of the property and will purchase a large flat (Figure 2 (b)). It is seen that the agent does not offer actions that are unacceptable to the users (the red ones). Further experiments and a detailed description of the IVA architecture are given in reference [16].

DISCUSSION AND CONCLUSION

This article present the process of modelling and realization of realistic and lifelike intelligent virtual agents (IVAs) in virtual and mixed reality. This IVAs are conversational, autonomous, learning agents, with visualized 3D head and body, with a unique face and physical characteristics, with an ability to control their eyes, to express emotions, to understand the preferences of users, to choose a suitable goal and a way to achieve it, which is acceptable to users. Discussed are some of the results user dialogues when solving a particular problem; the selection of a goal and the way to achieve it; the appearance of the modelled IVA is shown as well as the process of its learning and realization.

The benefit of this work is that several steps have been taken towards achieving a realistic lifelike IVA that can be engaging for consumers. However successful social interactions such as social engagement, collaboration, and smoothness include building Rapport. The phenomenon of rapport means a close and harmonious relationship in which the affected people or groups understand their feelings or ideas and lead to communicative efficiency, better learning results, successful negotiations [24] [25]. Applications and interfaces based on such techniques could have impact across a wide-ranging of social domains. As this research advances, our realistic lifelike IVA model can be used for further steps to understand a computational rapport process model and for designing effective computer mediated human-human interaction under a variety of constraints [24] [25].

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support provided within the Technical University of Sofia, Research and Development Sector, Project for PhD student helping N202PD0007-19 “Intelligent Cognitive Agent behavior modelling and researching”.

REFERENCES

1. C. Flavian, S. Ibanez-Sanchez, C. Orus, “The impact of virtual, augmented and mixed reality technologies on the customer experience”, *Journal of Business Research*, Elsevier, 2018, https://doi.org/10.1016/j.jbusres.2018.10.050.
2. E. Constanza, A. Kunz, M. Fjeld, “Mixed Reality: A Survey”, *Lecture Notes in Computer Science*, 2009, DOI: 10.1007/978-3-642-00437-7_3
3. S. Rokhasaritalemi, A. Sadeghi-Niaraki, S. Choi, “A review on Mixed Reality: Current Trends, Challenges and Prospects”, *journal of MDPI, Applied sciences*, 2020. Appl. Sci. 2020, 10, 636;
4. Концепция за цифрова трансформация на българската индустрия (индустрия 4.0); https://www.mi.government.bg/files/useruploads/files/ip/kontseptsia_industria_4.0.pdf
5. B. Knijnenburg, N. Hubig, Human-Centric Preference Modelling for Virtual Agent, IVA '20: *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents* October 2020 Article No.: 33 Pages 1–3https://doi.org/10.1145/3383652.3423909
6. R. S. Sutton and A. G. Barto, “Reinforcement Learning: An Introduction”, *the MIT Press*, Cambridge, London, England, 2014. [Online]. Available from: http://incompleteideas.net/book/ebook/the-book.html, [retrieved: 12, 2019].

7. A. Gosavi, “Reinforcement Learning: A Tutorial Survey and Recent Advances,” *INFORMS Journal on Computing*, Vol. 21 No.2, 2008, pp. 178-192.
8. W. H. Patric, “Artificial Intellligence”, Third Edition, *Addison Wesley*, 1992.
9. R. R. Torrado, P. Bontrager, J. Togelius, J. Liu, D. Perez-Liebana, “Deep Reinforcement Learning for General Video Game AI,” *IEEE Conference on Computational Intelligence and Games*, CIG. August, 2018, 10.1109/CIG.2018.8490422
10. B. Argall, “Learning Mobile Robot Motion Control from Demonstration and Corrective Feedback”, *Robotics Institute Carnegie Mellon University Pittsburgh*, PA 15213, March 2009.
11. H. B. Amor, D. Vogt, M. Ewerton, E. Berger, B. Jung, J. Peters, “Learning Responsive Robot Behavior by Imitation,” *IEEE/RSJ International Conference on Intelligent Robots and Systems* (IROS 2013) IEEE, Tokyo, Japan, November 3-7, 2013, pp. 3257-3264.
12. K. Takahashi, K. Kim, T. Ogata, S. Sugano, “Tool-body assimilation model considering grasping motion through deep learning,” *Robotics and Autonomous Systems*, Elsevier, Volume 91, 2017, pp. 115–127.
13. J. Cassell, “Embodied Conversational Agents: Representation and Intelligence in User Interface”, *MIT Media lab*, AI Magazine, 2021.
14. J. Allen, G. Ferguson, A. Stent, “An Architecture for more realistic conversational systems”, *ACM* 2001, ACM 1-58113-325-1/01/0001
15. C. Pelachaud, M. Bilvi, “Computational Model of Believable Conversational Agents”, January 2003, *Lecture Notes in Computer Science* 2650:300-317; DOI: 10.1007/978-3-540-44972-0_17
16. D. Budakova, V. Petrova-Dimitrova, L. Dakovski, Smart Broker Agent Learning How to Reach Appropriate Goal by Making Appropriate Compromises, ICAART February 2021, Portugal, online streaming, 2021.
17. Windows Desktop Speech technology https://docs.microsoft.com/en-us/windows/apps/speech
18. RT-Voice PRO Hearing is understanding
https://www.crosstales.com/media/data/assets/rtvoice/RTVoice-api.pdf
19. Expressive AI-Driven Conversational Characters in AR: Spirit Character Engine in Unity with ARCore + SALSA LipSync https://medium.com/spirit-ai/expressive-ai-driven-conversational-characters-in-ar-spirit-character-engine-in-unity-with-arcore-17c0348a568d
20. UMA 2 - Unity Multipurpose Avatar https://assetstore.unity.com/packages/3d/characters/uma-2-unity-multipurpose-avatar-35611?aid=1101lNJe&utm_source=aff
21. https://en.wikipedia.org/wiki/Mixamo;
22. https://www.mixamo.com/#/;
23. https://techcrunch.com/2013/07/23/mixamo/ Mixamo Is Building A Platform For Game Developers To Create And Animate 3D Characters.
24. Jonathan Gratch, Ning Wang, Anna Okhmatovskaia, et al., “Can virtual humans be more engaging than real ones?”, *Appears in the 12th International Conference on Human-Computer Interaction*, Beijing, China, 2007, DOI: 10.1007/978-3-540-73110-8_30 · Source: DBLP
25. Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., van der Werf, R., et al., “Virtual Rapport”, *Paper presented at the 6th International Conference on Intelligent Virtual Agents*, Marina del Rey, CA. 2006.