Teach-Me DNA : an Interactive Course Using Voice Output in an Augmented Reality System

Mouna Kenoui
Telecom Department. CDTA
Centre de Développement des Technologies Avancées
Algiers, Algeria
mkenoui@cdta.dz

Mohamed Ait Mehdi
Informatics Department. USTHB
Université des Sciences et de la Technologie
Algiers, Algeria
aitmehdi044@gmail.com

***Abstract*— In this paper, we present an interactive system partly based on voice output in an augmented reality environment. Using a chatbot technology, this system allows the user to engage bidirectional communication with a conversational agent. The present system builds upon our previous work in which the user interacts with 3D virtual objects via voice commands. We describe, in the current document, how we integrate speech output using the Text to Speech Service (TTS API) available on the IBM Watson platform, the goal being to obtain an even more interactive system. We also employ Speech Synthesis Markup Language (SSML) to control some features of the natural-sounding speech thus produced. Moreover, we developed Teach-Me DNA, an interactive application that gives the user (pupil, student) the opportunity to learn and/or revise the DNA molecule's basics as a part of a biological course. On one hand, and via voice inputs the user is simply able to perform 3D selections and 3D manipulations of the molecule and its several components in order to learn about their 3D features and properties, this first part dealing with the vocal entries was fully documented in a previous work. On another hand, Teach-Me DNA is here enriched by producing a natural speech when reacting to the user's requests. In fact, an agent's voice is used to respond in real time to student's questions. Definitions and/or explanations are thus given in a very natural way which might immerse more significantly the student in the proposed learning environment based on augmented reality technology.**

Keywords—HCI, Natural Interaction, Speech Recognition, Voice Output, Augmented Reality, Pedagogical Agent, Chatbot, IBM Watson, Cloud Application, Biomedical Course, Education.

I. INTRODUCTION

Augmented Reality (AR) [1] systems continue constantly to attract more interest and attention in both research and industry fields. By using AR technology, education and training are some of the most prominent areas that seek to bring innovative tools and applications for learners, students and for teachers, educators as well [2, 3, 4]. The main aim is to enhance learning/teaching processes [5, 6] and to allow new experiences, the Merge Cube product [7] is only one example among others we could cite here, several applications using this original concept are available for STEM (Science, Technology, Engineering and Mathematics) learning for instance.

In this wide domain, 3D natural interaction [1, 8, 9] is seen as the part trying to use the different human body's senses in order to let users interact more efficiently and much easily with the augmented scene. If the hand gesture is by far, the most used modality to interact naturally with the 3D objects in augmented reality environments, other means as gaze, voice, body motion, facial expressions, emotions etc... are more and more sought to meet some specific needs in many applications [1, 9, 10]. Recently, voice interaction has been interestingly investigated and used in some systems related to AR environments [11, 12, 13], several works refer to the easiness and effortless character of this modality. The voice is also often employed as a second modality in multimodal AR interfaces [14, 15, 16], which benefits the user by letting them performing other hand tasks while interacting simply and easily by speech commands.

Another point tied to 3D interaction in general and to voice interfaces in particular, is that researchers and developers are increasingly interested to propose systems that are even more interactive than what it has been to date, in a way that users could interact intuitively with the system and could also get reactions from it as natural as possible for the human perception. The frontiers, between the human and the machine, are thus redefined and revisited to bring another generation of applications that might immensely change some domains as education, healthcare, environment, to cite just few examples. When we consider voice interaction more closely in literature, we notice that a lot of efforts have been made in speech inputs, the objective is that users can trigger different events within a system by using voice entries. The first type of those systems allows few voice commands by using simple words as in [13, 17, 18] but also, a more natural speech has been integrated in some other applications in which the user is able to interact with the system by speaking in a more natural manner and using longer phrases [9, 19]. However, a very few works tackle speech as real time outputs, especially in AR environments. These outputs represent a set of natural reactions produced by the system towards the user, helping sometimes to establish an intuitive communication between the user and the system. This certainly contributes to diminish the gap between the human and the machine/computer, and lead to groundbreaking applications for the benefit of the socio-economic needs.

The present work addresses, in particular, the topic of voice output in an augmented reality environment. More generally, the user is able to engage a verbal communication with the system, not only by introducing speech entries to select and manipulate 3D objects, but also by getting responses to some inquiries. The application especially focuses on voice interaction based on a pedagogical agent that can be considered as a part of new ways of communication between human and machines for education purposes.

Pedagogical agents have been used as tools to support learning and teaching processes [20]. They are also seen as useful assistants capable of increasing the user's motivation and of strengthening their engagement and immersion. We

978-1-7281-5835-8/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: Universidad Politecnica de Cartagena. Downloaded on October 03,2020 at 10:38:58 UTC from IEEE Xplore. Restrictions apply.

think that for learners, using AR technologies and having a conversation with a such agent, could allow them to have valuable personalized learning experiences while evolving in both real and virtual worlds seamlessly.

In this paper, we especially highlight the voice output module, even though the whole interaction sub-system is described as well to give the reader a more general insight of the interaction we proposed. In fact, as mentioned above the current work is based on an earlier prototype designed and developed to allow voice control in an augmented reality system [9]. Finally, the present document is organized as follows. Section II describes related work and focuses on pedagogical and/or conversational agents used in educational AR environments. Section III presents our prior work's background. Section IV gives details about the conversational agent and describes the voice output part. In Section V, we discuss the application, experimentation and results. Our conclusion is eventually given in Section VI.

II. Related Work

Voice-enabled AR applications with learning and teaching goals can allow an easy and an intuitive manner of interaction with 3D objects to make it easier for learners by letting them focus deeper on the learning activities. This kind of systems contribute to bring another generation of products inside classrooms and beyond for different levels of learners and for different subjects and/or topics including languages, STEM and architecture but not only. In [11] for example, the authors developed a desktop AR application that aims to teach English as a second language, the application is dedicated to non-native children and helps them learning about colors, shapes and spatial relationships, a speech recognition was done by using Kinect sensor and Microsoft Speech API (SAPI) which allow children to interact with the different objects within the augmented scene by using different voice commands. Another work [21] proposed to combine AR and speech recognition to enhance productive vocabulary learning and to improve pronunciation skills. This system was also developed for children and involved the implementation of speech recognition into the application through Unity3D tool and Speech Recognizer API. In the field of mathematics, a mobile AR application [22] was designed and developed based on Doman Method for early development, children between 4-7 years old were able to interact via gesture and voice with 3D objects that occur in real time depending on their specific physical environment. The children were able to learn new concepts or to reinforce other covered ones by mainly resolving different operations. Unlike the two works cited above, this application allowed voice input and also voice output for the interaction and the children were thus able to hear out loud speaking parts for some learning activities, they were also meant to respond to some questions in order to check their understanding.

In virtual reality/augmented reality, voice interaction is also related to new concepts as for instance animated characters, conversational assistants, and more specifically to learning/teaching purposes, pedagogical agents. These different entities can engage a closer exchange with the final user and accompany them to accomplish a task or to acquire a new learning, and this not only by having visual feedbacks but also by getting speech output in real time. In his work [23], J.T. Doswell used for example the term "Virtual instructor" to refer to a pedagogical agent which can be either an embodied (3D animated) or a non embodied character. In this work, a complete architecture was described aiming to provide a mobile augmented reality system based, in particular on speech interfaces and pedagogical services. A non-embodied virtual instructor was thus implemented allowing conversational communication; Java, C++ and OpenCV were used for developing the proposed software. ARGarden [24] is another application that introduced "Learning Companion" concept to describe an agent that took a bluebird's form which helped providing interactive edutainment experiences with users and allowing flower gardening teaching in real space. Among other interactive capabilities, the learning companion was capable of expressing peer support to the users' interaction by offering pedagogical comments to assist the users to solve problems. Besides, in [25] an AR platform called ARTutor was set up for allowing interactive distance learning in which educators are enabled to easily create AR content for existing textbooks, an interactive part make it possible for learners to naturally communicate with the learning material by using speech recognition and vocal feedbacks. In this work, text to speech functions have been implemented for a mobile AR environment. Moreover, the outstanding contribution of Artificial Intelligence (AI) has been tremendously improving natural interaction and especially natural language processing (NLP), partly by bringing more speech functions and services that are both efficient and easy to use in order to build useful speech-driven applications. For instance we can cite chatbots, which benefit from the latest AI capabilities and are employed in many educational environments and even with augmented reality as in MondlyAR used for teaching foreign languages [26], in AREDAI which combines AR, and chatbot module for building an educational tool [27], also the new concept of "Teacherbot" is lately becoming more discussed by bringing up the idea of having intelligent robots/chatbots for teaching students of different levels and especially in higher education as highlighted in [28, 29, 30].

III. Background and Motivation

In our prior work, we have designed and developed an AR system using a vision-based technique for the tracking part, "Hiro" marker has been chosen and employed with ARToolKit Library during the implementation [31]. This has allowed us to mix the 3D contents with the real objects within the user's environment which is normally captured by an input sensor, more specifically we have used a camera in a desktop setting for the application.

Only visual outputs were then available for the user who was able to get feedbacks to their own actions/commands via screen. Also, the user has been able to use the proposed voice interface allowing them to naturally and easily interact with the augmented scene by selecting, manipulating or zooming several 3D virtual objects which represent, for the experimentation part, the whole DNA molecule or its different components. The effort was especially made to design and implement the voice interaction sub-system, which was broken down, for the first version [9], into four different parts. The first part called "Entry/Command Manager" (E/C-M) devoted to handle each voice command pronounced by the user and to treat it in order to finally

Authorized licensed use limited to: Universidad Politecnica de Cartagena. Downloaded on October 03,2020 at 10:38:58 UTC from IEEE Xplore. Restrictions apply.

distinguish and apply the right action desired by the user. The E/C-M relies on the three other parts which are invoked for specific needs as follows:

A. Keywords Recognition Part

Words/Sentences that are firstly given in a dictionary are able to be recognized by using KeywordRecognizer from the SAPI of Windows. The sentences' length is, in this case, up to three words. Otherwise, E/C-M invokes the other two parts.

B. Speech to Text (STT) Part

This part is responsible of being an interface between the E/C-M and the IBM Watson STT service on the cloud platform [32]. Mainly the speech is sent to be treated and converted into text. When the text is returned, E/C-M will trigger the Assistant part for seeking the right action to perform.

C. Assistant (Conversation) Part

Similarly, this part is responsible of being an interface between the E/C-M and the IBM Watson Conversation service on the cloud platform. After getting the text related to the vocal entry of the user, the conversation service is then invoked and the final decision is communicated to the E/C-M capable of performing the action.

The modular design and implementation allow us to be able to integrate future extensions to the vocal system, by keeping a certain simplicity to some extent, while treating possible challenging functionalities. Hereinafter, “Fig. 1” giving more insights in the architecture set up for our previously developed system. Thorough explanations are available for the overall system in [9], especially for getting more technical specifics and a view on the used commands.

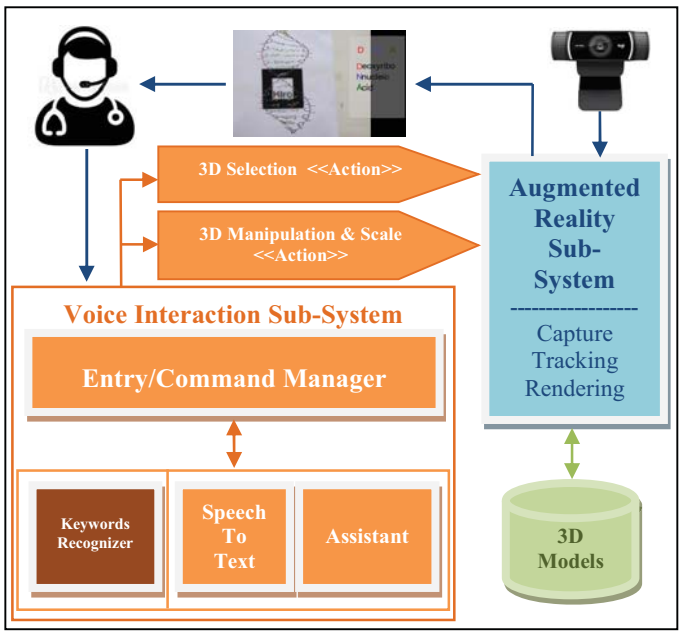

Fig. 1. A glimpse of our previous work - System Architecture.

Our motivation is clearly to evolve the current system to a more interactive system by allowing the user to have voice output, in addition to vocal commands that they are already able to initiate. Thus, the goal is to add more natural features into the system. As mentioned above in the introduction, we think that getting visual and audio feedbacks in educational environments can enhance both the acquisition process and the engagement, the motivation and the immersion of the learner for better personalized interactive experiences.

IV. Vocal Feedback

The whole voice interaction system is orchestrated by the E/C-M module, in which we have here integrated new functionalities in order to allow the production of real-time vocal feedback in response to part of voice commands of the user. For this purpose, the conversational agent already designed and implemented (see the bottom of “Fig. 2”), was enhanced and extended by specifying new elements and then by converting text responses to voice outputs that the user can instantly hear, this latter was especially done by integrating a new service called "Text to Speech" (TTS) from the IBM Watson platform [32], data circulation is ensured by the E/C-M module. “Fig. 2” below also shows all the details of our system and gives a quick look at the flow of data inside the proposed system, including communication with the entire local and distant environment.

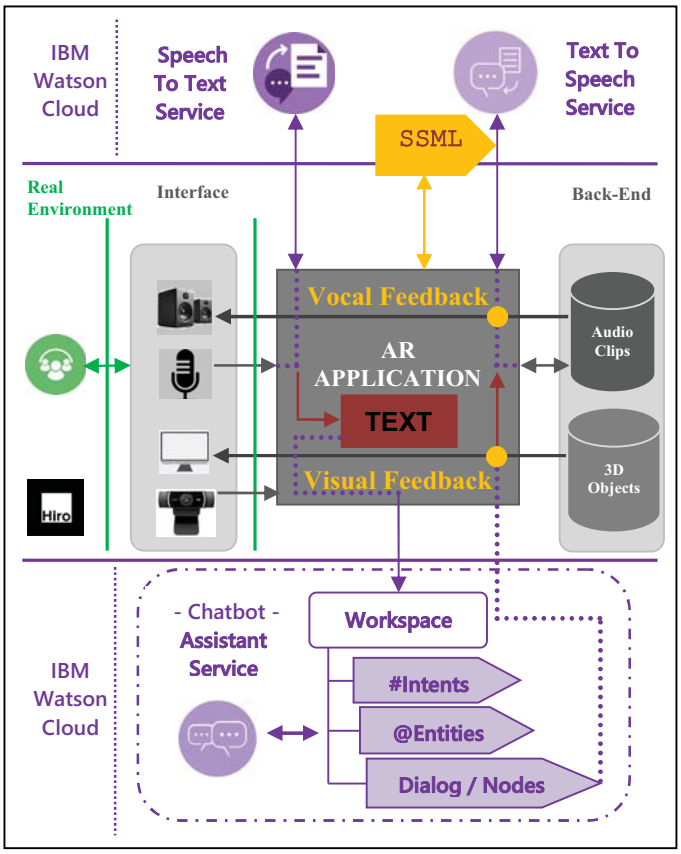

Fig. 2. Voice output made by Chatbot technology and TTS service.

The choice of IBM Watson platform was firstly motivated by the fact of investigating NLP services capabilities and how it is feasible to tailor them for allowing AR natural interaction in an input and an output setting. Secondly, one of the future possibilities, is to bring the different treated services that are on the cloud of the company (IBM Watson) to even a private cloud, which can be very promising in particular for us, as we work for an organization that disposes of its own cloud platform. Lastly, many other services and tools can be incorporated as well, including for instance visual recognition services, AI tools,

Authorized licensed use limited to: Universidad Politecnica de Cartagena. Downloaded on October 03,2020 at 10:38:58 UTC from IEEE Xplore. Restrictions apply.

machine learning and/or deep learning capabilities and much more, which is a real opportunity for designing and developing useful systems and applications for augmented, virtual and/or mixed realities, in both easier and richer manner than in the past.

In a more specific view related to our current contribution here, a chatbot technology based on Watson Assistant service has been successfully integrated and extended in order to vocally command 3D objects in AR environment and more particularly for managing the possible vocal outputs according to users requests, here we note that the IBM Watson chatbot is actually an only text version. In fact, we have built a vocal layer based on combining other services and features as here Watson TTS service/SSML so as to produce the desired vocal outputs in response to different students' inquiries in a biological course and in particular for teaching DNA molecule, the students are not interacting by introducing text via keyboard but only by voice as if they were asking a teacher about their questions and interests.

A. Conversational Agent

As explained earlier, this part has been designed and implemented by using the powerful chatbot technology of IBM Waston. We have created a workspace for the system, we have then tailored all the needed parts to fit our application needs. Hence, in order to allow any response from the system, the user's entry is analyzed as described in the previous section, when their request is about a visual action, a 3D effect is in return produced, but when the user whishes a vocal response, the conversational agent is able to provide a text format of the required response, this one is then communicated to the E-C/M module which is responsible of performing the voice version of the feedback to the user. To do that, we enriched the number of recognized *Intents* and *Entities* which permits to provide the user with more flexibility to ask many questions/inquiries in different ways and then to get the right answer to each of them, this has been managed in the *Dialogue Nodes Part* where decisions are given in a textual format as we already mentioned. For a better understanding of the reader, we find it necessary to highlight the three concepts related to the chatbot we used here, below follows a quick description of each of them:

- Intent: is a category that represents and defines the user's goal/purpose.
- Entity: is a significant part of the user's entry/input that gives a specific context and can be used to alter/personalized the response to the user's intent.
- Dialog/Node: using the application's context and the recognized intents and entities, this part represents the conversation flow that ultimately gives the user the required response to a given intent.

B. Voice Output

Once the decision is given by the Dialog/Node part, the text obtained is communicated to the E/C-M which treats it whether it induces a visual or a vocal effect, and even sometimes both effects simultaneously, depending on the pronounced phrase. In case a voice output is needed, E/C-M retrieves the audio file that matches with the current response and performs it towards the user hearing. We prepared a database of all known responses, which optimizes the time required to return the response. Otherwise, E/C-M has to invokes the TTS service on the cloud in real time and for the same response over each time it is needed. Hereinafter, more explanation of the two elements incorporated in our application helping to produce the audio clips:

- TTS service (Text to Speech): is another IBM Watson service (as mentioned earlier) which uses speech-synthesis capabilities in order to convert a text input to a natural speech. This can be done in different languages, accents, voices and much more. To fit specific requirements of different applications, the gender of the voice-sounding can be chosen as well. All these features contribute to give more naturalness to the output interaction.
- SSML (Speech Synthesis Markup Language): is the language we integrated as well in this version in order to add and control all the features cited above, as at first hand, TTS produces a neutral declarative style speech synthesis, which is often inconvenient for learning applications. SSML is an XML-based markup language that enhances the expressiveness of the produced speech, we thus combined SSML and TTS to beforehand create a set of audio clips and then play those audio files at the right time.

V. Experimentation and Results

For the development of the proposed system, we have chosen to implement it under Unity3D Engine [33], the very known 3D game tool, we have also added two specific SDKs allowing to work on the AR sub-system and the interaction sub-system. Hence, we have used the 17.2 Personal version of Unity app. For the AR part, ARToolkit SDK for Unity (version 5.3.2) has been used and tailored to provide more interesting characteristics within the application as for instance to give the possibility for the user to choose a given camera to use among all the available cameras connected to the computer, all these features have been gathered in the setting part that the user can set up before starting the learning experience. “Fig. 3” shows here the first user interface of the application. We can also see in “Fig. 4” The real environment captured by the camera where the DNA molecule is visualized and where a Hiro marker appears in the scene.

Fig. 3. First user interface of the application.

Some annotations can be also displayed in order to enhance the comprehension of the material related to the course as shown here below.

Authorized licensed use limited to: Universidad Politecnica de Cartagena. Downloaded on October 03,2020 at 10:38:58 UTC from IEEE Xplore. Restrictions apply.

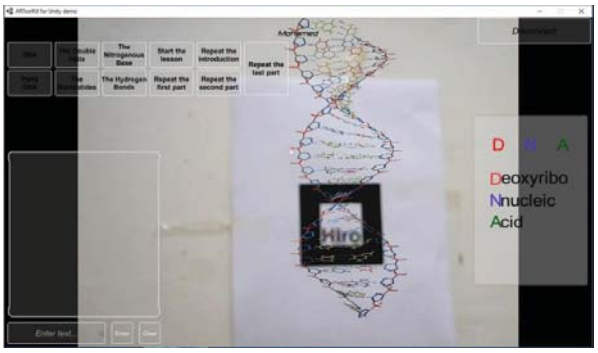

Fig. 4. Student "Mohamed" visualizes the DNA molecule.

For the interaction part, the IBM Watson SDK for Unity has been added, with all the services mentioned in the previous section. Several C# scripts have been created in Unity to implement the Entry/Command Manager and the other parts as interfaces between this Manager and the remote services on the cloud platform. The user is then able to perform a series of actions by simply using voice inputs, they can select a specific component of the DNA, translate it, rotate it and/or zoom it. For this aim, they could use more than a way for the same action. For example, to select the Adenine, a user can say:

1. Show me the nitrogenous base Adenine (see "Fig. 5")
2. Please, I want to see the nitrogenous base Adenine
3. Can I see the Adenine molecule, please?

And so on…

Here, the user's intent being to select/to visualize the "nitrogenous base Adenine".

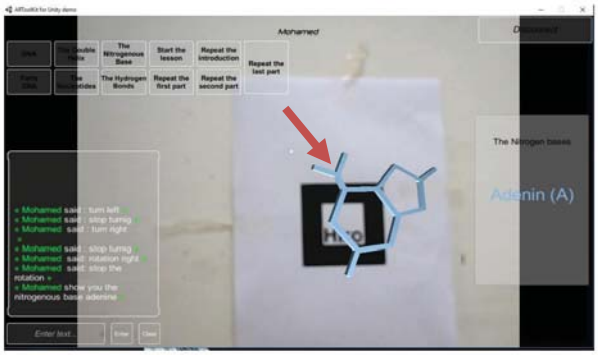

Fig. 5. Student "Mohamed" selects/visualizes the Adenine (A)

The user is also able to perform rotation of the selected molecule, as following:

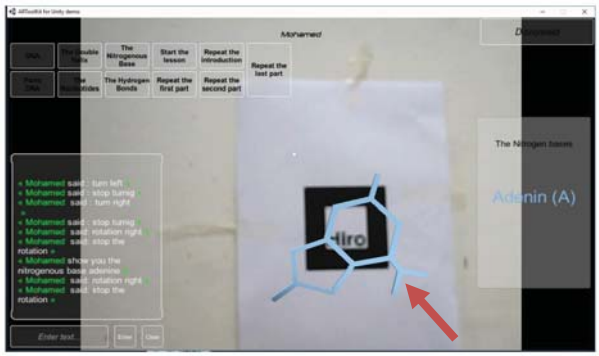

Fig. 6. Student "Mohamed" rotates the Adenine (A)

Finally, added to these visual 3D effects, the user can trigger vocal responses related to the course, "Fig. 7" shows student "Mohamed" saying:

"Start the lesson"

Following this command, definitions related to the course are played, here Mohamed is able to hear definitions related to the DNA molecule, and this is because he was already visualizing the entire molecule and not a part of it, so the different states/contexts of the system/application are also taken into account when a command is pronounced.

Mohamed is then able to see different visual parts as the definition is given, he could stop the lesson and ask another question or repeat (a part of) the lesson another time.

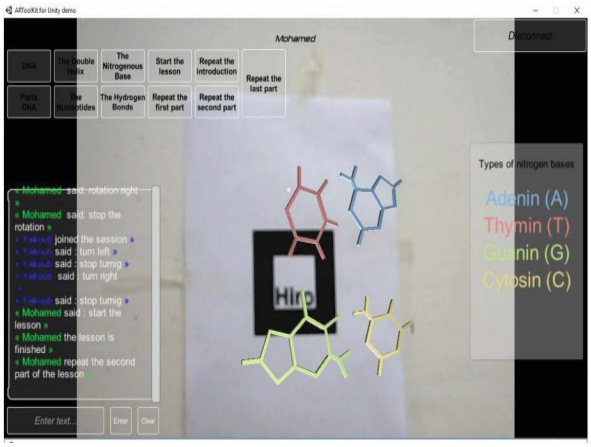

Fig. 7. User asks for an explanation and gets vocal and visual responses.

Although our system can be extended to other contents of this course for teaching biology and even to other subjects/courses such as anatomy or geometry, at this stage our contribution was limited due to some issues, like for instance, costs related to acquiring significant and richer 3D objects related to DNA and/or other biological molecules/contents, in our case we could use only few free objects easily downloadable on the Internet. For the interaction itself, the results we have had here are interesting, but some points are to notice as well: (i) Natural and long phrases are recognized allowing the user to use an intuitive speech to command 3D objects, but this part relies on services reachable on the cloud and thus requires an internet connection, we have also added command buttons to use if any connection loss happens, (ii) The entire course is based on voice interaction, but some studies have shown that for some tasks, combining this modality with hand gesture could present useful outcomes, which we consider as an element to add in future works.

VI. Conclusion

Throughout this work we have seen that natural interaction applications are increasingly on demand these last decades, especially in education area. When augmented reality technology is integrated to educational environments where learners and/or teachers are seamlessly using real and virtual objects at the same time, using natural modalities becomes a pivotal way to ultimately provide educational products that are more likely to be adopted in classrooms for common use and for a wider expansion in education community. In this paper, we have tried not only to allow voice commands from the user (pupil/student) but also to give back responses to different inquiries of the learner

Authorized licensed use limited to: Universidad Politecnica de Cartagena. Downloaded on October 03,2020 at 10:38:58 UTC from IEEE Xplore. Restrictions apply.

within an AR biological course. Our system has shown that augmented reality can take advantage of different new concepts and tools to build interactive applications for learners/teachers to allow new experiences and innovative ways to learn and to teach. We are currently working on improving the proposed application, one of the future functions is to add a collaborative session where multiple users can share a same augmented scene and interact together for a better and a richer learning experience.

REFERENCES

[1] M. Billinghurst, A. Clark, and G. Lee, "A survey of augmented reality," Foundations and Trends in Human-Computer Interaction. vol. 8, no. 2–3, pp. 73-272, 2015.

[2] M. Meng, P. Fallavollita, T. Blum, U. Eck, C. Sandor, S. Weidert, J. Waschke, and N. Navab. "Kinect for Interactive AR Anatomy Learning," in Proceeding of IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Adelaide, Australia, pp. 277–278, October 2013.

[3] S. Boonbrahm, C. Kaewrat, and P. Boonbrahm. "Using Augmented Reality Technology in Assisting English Learning for Primary School Students," Lecture Notes in Computer Science, Springer, Cham, vol. 9192, pp. 24-32, 2015.

[4] J. Barrow, C. Forker, A. Sands, D. O'Hare, and W. Hurst. "Augmented Reality for Enhancing Life Science Education," in Proceedings of the Fourth International Conference on Applications and Systems of Visual Paradigms. Italy, June 2019.

[5] W. Guo, "Improving Engineering Education Using Augmented Reality Environment," In: Zaphiris P., Ioannou A. (eds) Learning and Collaboration Technologies. Design, Development and Technological Innovation. LCT 2018. Lecture Notes in Computer Science, Springer, Cham, vol. 10924. pp 233-242.

[6] E. İbili, M. Çat, D. Resnyansky, S. Şahin, and M. Billinghurst, "An assessment of geometry teaching supported with augmented reality teaching materials to enhance students' 3D geometry thinking skills," International Journal of Mathematical Education in Science and Technology. pp 1-23 , 2019.

[7] JP. Van Arnhem, "Merge Cube's Handiness with Holograms Makes it a Good Place to Start with Augmented Reality," Mobile Apps and Gear for Libraries. Advisor Reports from the Field, The Charleston Advisor, July 2018.

[8] K. O'Hara, R. Harper, H. Mentis, A. Sellen, and A. Taylor, "On the naturalness of touchless: putting the interaction back into NUI," ACM Transactions on Computer-Human Interaction Journal (TOCHI), Special issue on the theory and practice of embodied interaction in HCI and interaction design. USA, vol. 20, issue 1, article 5, March 2013.

[9] M. Kenoui, and M. Ait Mehdi, "3D natural interaction for an augmented reality system," IEEE International Conference on Advanced Electrical Engineering (ICAEE), Algiers, Algeria, November 2019. "in press"

[10] J. Aliprantis, M. Konstantakis, R. Nikopoulou, P. Mylonas, and G. Caridakis, "Natural Interaction in Augmented Reality Context," Proceedings of 1st International Workshop on Visual Pattern Extraction and Recognition for Cultural Heritage Understanding co-located with 15th Italian Research Conference on Digital Libraries (IRCDL 2019), CNR Area in Pisa, pp. 50-61, Italy, January 2019.

[11] C.S.C Dalim, A. Dey, T. Piumsomboon, M. Billinghurst, and S. Sunar, "TeachAR: An Interactive Augmented Reality Tool for Teaching Basic English to Non-Native Children," In Proceedings IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 82-86, 2016.

[12] M. Whitlock, E. Hanner, J.R. Brubaker, S. Kane, and D.A. Szafir "Interacting with Distant Objects in Augmented Reality," in proceedings of IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Reutlingen, Germany, March2018.

[13] J. Liu, S.J. Al'Aref, G. Singh, A. Caprio, A.A.A. Moghadam, S.J. Jang, S.C. Wong, J.K. Min, S. Dunham, and B. Mosadegh, "An augmented reality system for image guidance of transcatheter procedures for structural heart disease," PLoS ONE, vol. 14(7), 2019.

[14] S. Irawati, S. Green, M. Billinghurst, A. Duenser, and H. Ko, "An Evaluation of an Augmented Reality Multimodal Interface Using Speech and Paddle Gestures," Advances in Artificial Reality and Tele-Existence (ICAT), Lecture Notes in Computer Science, vol. 4282. Springer, Berlin, Heidelberg, 2006.

[15] M. Lee, M. Billinghurst, W. Baek, R. Green, and W. Woo, "A usability study of multimodal input in an augmented reality environment," Virtual Reality journal, Springer, vol. 17(4), pp. 293-305, November 2013.

[16] T. Piumsomboon, D. Altimira, H. Kim, A. Clark, G. Lee, and M. Billinghurst. "Grasp-Shell vs Gesture-Speech: A Comparison of Direct and Indirect Natural Interaction Techniques in Augmented Reality," IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Munich, Germany, pp. 73-82, November 2014.

[17] Z. Chen, J. Li, Y. Hua, R. Shen, and A. Basu, "Multimodal Interaction in Augmented Reality," In Proceedings of IEEE International Conference on Systems, Man, and Cybernetics (SMC) Banff, AB, Canada, pp. 206–209, 2017.

[18] A. Sheldon, T. Dobbs, A. Fabbri, N. Gardner, M.H. Haeusler, C. Ramos, and Y. Zavoleas, "Putting the AR in (AR)chitecture : Integrating voice recognition and gesture control for Augmented Reality interaction to enhance design practice," Intelligent & Informed, Proceedings of the 24th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Vol. 1, pp. 475-484, 2019.

[19] MR. Mirzaei, S. Ghorshi, and M. Mortazavi, "Audio-visual speech recognition techniques in augmented reality environments," The visual comuputer, Springer, vol. 30, issue 3, pp 245–257, March 2014.

[20] A.S.D. Martha, and H.B. Santoso, "The Design and impact of the pedagogical agent : a systematic literature review," Journal of educators Online, vol. 16(1), 2019.

[21] N. Che Hashim, N.A. Abd Majid, H. Arshad, and W.K. Obeidy, "User Satisfaction for an Augmented Reality Application to Support Productive Vocabulary Using Speech Recognition," Advances in Multimedia, Vol. 2018-9753979, June 2018.

[22] A. Solano, F. Ugalde, J. Gómez, and L. Sánchez, "An Augmented Reality Application to Enhance the Children's Engagement in an Early Development Method for Mathematics Literacy," In (Eds) Advances in Usability and User Experience, AHFE 2017, Advances in Intelligent Systems and Computing, vol. 607, Springer, Cham.

[23] J.T. Doswell, "Context-Aware Mobile Augmented Reality Architecture for Lifelong Learning," In Proceedings of the Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06), Kerkrade, Netherlands, July 2006.

[24] S. Oh, and W. Woo, "ARGarden: Augmented Edutainment System with a Learning Companion," In (Eds) Transactions on Edutainment I, Lecture Notes in Computer Science, vol. 5080, pp. 40–50, Springer, Berlin, Heidelberg, 2008.

[25] C. Lytridis, A. Tsinakos, and I. Kazanidis, "ARTutor—An Augmented Reality Platform for Interactive Distance Learning," Education Sciences Journal, vol. 8, article. 6, Basel, Switzerland, 2018.

[26] MondlyAR. https://www.mondly.com/ar

[27] P.S. Rajendran, I. Sam Christian, and M. Sanjay Shedge, "AREDAI Augmented Reality Based Educational Artificial Intelligence System," International Journal of Recent Technology and Engineering (IJRTE), Vol. 8-1, May 2019.

[28] S. Bayne, "Teacherbot: interventions in automated teaching," Teaching in Higher Education: Critical Perspectives Journal, vol. 20(4), pp. 455-467, 2015.

[29] S.A.D. Popenici, and Sharon Kerr, "Exploring the impact of artificial intelligence on teaching and learning in higher education," Research and Practice in Technology Enhanced Learning, vol 12:22, Springer Open, 2017.

[30] M. Fahimirad, and S.S. Kotamjani, "A Review on Application of Artificial Intelligence in Teaching and Learning in Educational Contexts," International Journal of Learning and Development, vol. 8, no. 4, Dec 2018.

[31] H. Kato, and M. Billinghurst, "Marker tracking and HMD calibration for a video-based augmented reality conferencing system," In IWAR '99 Proceedings of 2nd IEEE/Acm International Workshop on Augmented Reality, pp. 85-94, 1999.

[32] www.ibm.com/watson (official site)

[33] www.unity.com (official site)

Authorized licensed use limited to: Universidad Politecnica de Cartagena. Downloaded on October 03,2020 at 10:38:58 UTC from IEEE Xplore. Restrictions apply.