POLITECNICO DI TORINO

Master in Computer Engineering

Master Degree Thesis

Design and Implementation of a Conversational Agent to Stimulate Teacher-Students Interactions in Synchronous On-line Lecture Environments

Supervisor
Prof. Fabrizio Lamberti

Candidate
Javad Alizadeh Shabkhoslati

April 2021

Summary

Recently, many universities and schools have switched to online teaching due to the COVID-19 pandemic. Distance learning could limit social interaction between teachers and peer-learners. As this may demotivate learners in the long term, better social engagement providing solutions such as Virtual Reality (VR) can be used for teaching and learning. In the field of technology-enhanced learning, research has indicated that using Conversational Agents (CAs) to engage learners in one-to-one (student-agent) tutorial dialogues can improve students' comprehension and foster students' engagement and motivation. Such agents try to simulate the behavior of a human instructor or tutor and engage in a discussion with a learner on a series of predefined topics.

Research in the field of Computer-Supported Collaborative Learning (CSCL) has revealed that unsolicited CA interventions can intensify the knowledge exchange among learning partners and increase students' explicit reasoning and participation levels. In this thesis, it was designed a virtual conversational agent (VCA) intended to act as a classmate to stimulate teacher-student interactions in an on-line learning environment during a synchronous lecture. The VCA was experimented in a controlled lecturing scenario. It was implemented to be plugged into Mozilla Hubs, which provides the virtual classroom in which a real teacher teaches to real students and their virtual and active classmate (a Solar System lecture was used). The VCA exploits DialogFlow (NLP platform to design conversational user interface) to interpret the meaning of conversations and reacts appropriately to increase the level of attention and interaction during the lecture.

Experiments were designed to evaluate the potential of CA and co-learning with a virtual classmate (agent) to increase engagement, commitment and enjoyment of learning in VR.

Acknowledgements

First of all, I want to thank a lot my supervisor, Prof. Lamberti Fabrizio for the constant support, guidance and availability he provided through the whole development of the thesis.

My completion of this project could not have been accomplished without the support of Filippo Gabriele Pratticò and Davide Calandra. I offer my sincere appreciation for the support provided by them

Finally, to my caring, loving, and supportive wife, Farzaneh: my deepest gratitude. Your encouragement when the times got rough are much appreciated and duly noted. My heartfelt thanks.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Active learning can be generated by using technology in learning environments. Active learning happens when the students try to contribute and improve their roles as the key actor. The implementation of such environments that provides active learning requires wide researches in this domain and identifying the variables.

Making decisions needs a lot of data generated by each student. To achieve this objective Artificial Intelligence (AI) is considered. As same as the other parts of society, educational places would like to have improvement in their services and experience of its users. For this reason step by step the universities are going to be intelligent campuses. The most important issue is how to improve the students' performance in a sustainable way. Due to this, data generation and data analysis plays a significant role. While teaching can be based in or out of the classrooms, the use of computers and the Internet forms the major component of E-learning. So internet of things and cloud computing processing is important.

Moreover data acquisition and implementing this kind of technologies need a lot of efforts and has its own difficulties. But, at the end the benefits of this system worth, because the students experience better environment with deep understanding of the specific subject and many other advantages that will be covered later. In this level of AI, everything depends on the data which is evaluated to be exploited for decision-making.

When it comes to the classroom, it always evokes a space with the presence of a teacher and a number of students with a unidirectional communication between them. In the conventional teaching methods, it has been proven that passive roles of students lead to inefficient methods in the teaching and learning process. Using of new technologies, provide new opportunities in universities by generating active learning, where students are more interested to form a bi-directional activity and their interests make them the main actor in their education. However developing such an environment to let the students experience active learning, requires a lot of effort and knowledge. The number of variables and parameters involved in academic environment is high and needs precise attention for handling them. It is necessary to analyze the behavioral pattern of students and their generated data in search of the possible patterns that allows to classify them according to their needs. After the process of identification of the students' needs, it is possible to make decisions that contribute to learning of each student [23].

Universities and in general, educational environments, seek for constant improvement of their services and the given experiences to their members. Transforming traditional universities to intelligent campuses is gradually in progress. These smart campuses are targeting creating an ecosystem of ICT and their members' interaction where all plans and resources are focused on reaching the members' goals. Different players in this ecosystem are connected and the objective is to keep them united in planned path. One of these players are the students that improving their education in a sustainable way is the most important goal in smart campuses [23].

At the beginning of 2020, because of COVID-19 pandemic situation most of the schools and universities have used online methods. E-learning is going to be widespread also in future. Many students and teachers are going to switch to online courses [19]. However, the lack of real interactions between teacher and students and also between classmates is a big problem. Peer-learners motivation can be affected. VR and its immersive attribute can be a good solution in order to prevent the demotivation of learners in the long time.

In recent years, due to the technological progresses and fast pace improvement of Internet quality and speed, many of in-person lectures (traditional classes) have been considered to be changed to electronic versions. Specially after the wide spread of COVID-19 virus in the world and the strict limitations of not allowing the gathering too many students together in a classroom, many schools and universities have decided to focus and invest more in new solutions. Switching classes to remote ones can also have other side benefits such as lowering the carbon dioxide production by reducing travels. As well, eliminating the necessity of travel and presence in a common location, can save time and enhance geographic flexibility. In addition, researches have shown that remote learning can reduce the social stress among specific students, who specially are afraid of being among too many people and tend not to be seen physically [11].

There are many software that have been used for years in remote instructing and learning. Each application satisfies a subset of users, because these applications are designed to be used for general purposes and to have the most fundamental features needed for a remote learning. Most of well-known and trusted software such as Skype, Twitch or Zoom, are used as video conferencing tools. With these software, teachers can establish a private or public room for their students or learners. Usually the interaction is mono-directional and the teacher is the sole speaker. Tutor follows a context and narrates or explains the material to the students. In recent versions of these software, there features that students can use for better interaction. Such as chat, raising hand, sharing screen and more. These features are designed to facilitate the interaction among teacher and student. Although these tools and their features are generally useful and accepted for remote instructing still they may lack interactivity as we expect of in-person lectures.

Many studies have focused on the effects of using these software. For an effective remote instructing and learning, some factors such as: students' distractions, viewing-related discomforts and technical problems must be considered. Previous researches have revealed common technical problems while utilization of technologies like video-conferencing [11].

1.1 Computer-Supported Collaborative Learning

In a non-experimental study [22], 106 learners were asked to answer a questionnaire. The purpose of presenting this questionnaire was to identify the components involved in interactions in the academic environment that lead to the creation of a framework that can be used to achieve academic goals. Although extensive research has been conducted to establish a framework for implementing and modeling interactions in distance learning and teaching, and CSCL studies focus on this issue, information and research are still lacking. The students answered these questions after participating in five different topics that were implemented using the CSCL method. Factorial analysis shows that in the process of cooperation and convergence to achieve knowledge, three types of interactions were more important and bold for students: the design, implementation and assessment phases of collaborative learning.

The results of this study show that educators and institutions that seek to promote and implement CSCL should use the solutions appropriate for learners. Because achieving great learning and educational goals are intertwined with social and organizational aspects design, implementation and measurement of collaborative learning.

1.1.1 Design Phase

Designing and implementing an appropriate scenario can increase the efficiency and effectiveness of the CSCL method. Proper and fruitful interactions need proper planning and structuring because technological tools and elements are also involved in the learning process and play an important role as the main medium. The use of these tools should be aligned with the main goal, which is to increase productivity and learning, and not to hinder them [22]. This research shows that when the educational topic is a complex or project-oriented issue, learners have more opportunities to combine general and interactive skills.

The choice of tools or technologies should be in line with these goals so that in a dynamic space and environment, social activities and interactions are continuously formed and flowing. This continuous flow of interaction, leads to stability in the formation of teacher-student and student-student relationships and interactions, which leads to stability in solving problems and tasks presented in the classroom.

Understanding the philosophy of cooperation and interaction is the key to its implementation and expansion. Students should realize that more interactions to solve the problem bring them closer to the goal and make the learning process easier and deeper. They need to know why and how they interact and how these interactions lead them to different outcomes. In collaborative learning, what kind and level of cooperation is expected, and how the group continues to do the task.

1.1.2 Implementation Phase

In collaborative processes, each individual or member uses their prior knowledge and restructures it to the form of the group's needs for problem solving and communicating with other members of the group through social cognitive. These interactions are formed by each member with the other members of the group, which are formed at three levels, the social, cognitive and organizational level that transfer and circulates the information [22].

Learners have some needs in order to be able to form a proper social interaction with each other. Learners need to understand each other by recognizing behavior, exchanging feelings and social interactions with each other, so a better relationship is formed between them, which leads to more motivation to continue communication and achieve the common goal.

Research is conclusive on the need for social interaction, to promote emotional intragroup support and to recognize individuals at a personal level. In fact, the lack of communication and social interactions indirectly leads to the separation of members. If there is no communication between the members, there is a feeling of leaving and cognitive exchanges are reduced. Each member tries to achieve it regardless of the main goal or the group's goal. Here, the supportive and managerial role of the instructor is important. The instructor, by timely understanding of what is happening among the members, by providing appropriate feedback leads them to the desired direction to prevent the creation of gaps and isolation of the members, and increase the interaction between the members.

Cognitive interaction occurs through the negotiation of a shared meaning or knowledge convergence. Where each member strives for the common goal to share their knowledge through social interactions with others and bring the group closer to the result. Where the group makes an effort to integrate every individual contribution into a common construct and group members are exposed to knowledge convergence and divergence.

1.1.3 Evaluation

In educational activities, assessment and evaluation should be in accordance with the teaching methods and what has been taught, and inform the learner to what extent the desired goal has been achieved. This assessment should be transparent, applicable to any situation and each student. Because having a general method, and applicable to each student, makes a correct understanding of the performance of each student without the involvement of other factors [22]. It is crucial to have a metric applicable for all the involved students in a unique activity.

In CSCL method, the learning process takes place through collaboration between students and the social connections that are formed between them to solve common problems. For this reason, in this method, two issues are measured and evaluated. First, the relationships between them and secondly, the students' ability to solve problems.

1.2 Co-Learning

CSCL refers to learning situations mediated by technologies where small groups of three to five students are exposed to interaction in order to solve a complex unstructured problem or are required to design a project [22]. To solve the problem, the group must participate in an intensive process of collaboration and negotiation, which includes the interrelation of teaching presence, cognitive presence, and social presence, as defined by the Community of Inquiry (CoI) system. Students participate in processes that cause cognitive presence and information convergence through the creation of common understanding if communication is effectively planned and encouraged through teaching presence. The procedure must be based on social presence, i.e., personal recognition and intra-group emotional support.

In order to be more effective and to meet the needs of students for a deep learning, educators must design the teaching and learning process in a way that encourages greater collaboration and creativity. Structured and planned collaboration enhances individual learning and has a direct and significant impact on students' satisfaction with learning.

The main challenge in CSCL is to design and implement a planned and purposeful process to increase the ability of cooperation and interaction between students. One of the pillars of CSCL is teamwork, and by default, the goal of teamwork is to achieve a common goal. In a group work, people use their personal, knowledge and professional abilities to have a dynamic and effective interaction. But this issue should not cause the main challenge of CSCL to be forgotten, and these activities must be planned and purposeful.

In 1989, Moore proposed three different types of interaction frameworks: learner-to-instructor interaction, learner-to-learner interaction and learner-to-content interaction. In 1990 and 1994 two other types of interactions, intra-instructors and learner-to-interface interactions were introduces, respectively.

The researches in CSCL have proposed new types and models of interactions which connect different aspects of it while identifying positive and important influence between them:

- interaction among teacher and students and the students among their work group;
- interaction among the students and their work group and their emotional support in group;
- interaction of the student in his/her group and the collaborative learning;
- online tools and the interaction of the students among their work group.

Collaboration happens when several people work together to achieve a goal, and learning through collaboration also occurs when interactions between learners are directed toward learning. This is why in this type of learning, goals, behavioral patterns and topics find meaning according to the keyword participation.

In order for interactions to take place effectively and purposefully, in order to increase productivity in students' learning and academic activities, it is necessary to design the educational process with maximum accuracy. At this stage, the instructor plays a key role and, based on experience and educational content, outlines the elements and topics to be taught and, based on them, determines the main parameters. In the implementation phase, students work towards achieving the academic goal according to the schema of the previous phase, which are prepared by the instructor and under the supervision of the instructor (who plays the role of facilitator of activities). They use tools such as interaction, communication and knowledge exchange to achieve their goal. In the last stage, the evaluations are done. The instructor analyzes the results by reviewing the feedback and results obtained by the students during the learning process.

Collaborative learning is based on communication among peer-learners. Presenting and grasping information through different communication channel rather than one single channel is more effective. In online learning, learners communicate via possible mediums (depends on the learning environment), such as auditory voice messages which are becoming one of the basic features of any platform, as it is the easiest and the accustomed medium for in daily conversations. On the other side, text-based messages are still the oldest and the most widely used method of communication.

1.2.1 Co-Presence

In order to form social interactions, it is necessary for the two sides of the relationship to provide a situation for further interactions by their presence and communication with each other [11]. In educational environment, these connections and co-presence occur primarily through the communication that the teacher forms with the students.

Co-presence is the factor which can be interpreted from the observer's side. The degree that the observer believes he or she is not left alone in the class, the degree of focal awareness of another mate, and also the level and degree to which, other mates are focally aware of the observer.

Students are likely not to be aware of other classmates or person in the classroom. Unless they are prompted to interact with others and that is the time they are aware of their mates. This may never happen in a traditional classroom where teacher and all students are well aware of their classmates. But in a remote class handled by universities, lack of the experience of co-presence may lead the students feel they do not belong to that class or university.

1.2.2 Attentional Allocation

When it comes to an interaction, at least two person are involved. The amount of the attention one peer allocates to the other peer, demonstrates the level of attentional allocation. This parameters let us know when in a class, one person talks, others can focus on him/her and at the same time, how the person think others have focused on him/her. The research [11], shows a prominent relationship among this parameters and the previous parameter, co-presence. The results echo the amount of attentional allocation has a direct relation with the level of co-presence the person has among the class. So that if the person is aware of other peers and their presence he/she is more occupied with the level of the attention giving or receiving from other peers. During a lecture, it is more important for the student to have the ability to focus on others than knowing how others are focusing on him/her. Specially the student needs to focus on the teacher during the class, who is presenting and teaching to multiple students.

1.2.3 Perceived Message Understanding

In a traditional class, teacher and students have direct (not blinded) interaction. They feel each other presence not only by their voice, but also by their body movements and figure. Students have to ability to adapt themselves according to emotional and attitudinal states fo others. As interaction is not just held by voice, but also by the level of interest or importance added to what is being told. This parameter points to the audience understanding and perception of the presenter's attitude For example teacher can show the importance of a subject by changing his/her voice or taking on the appearance of a serious person.

1.2.4 Usability

Tools are meant to be easy to use, ease the process of reaching a goal, not making it harder. It is not an exception in the field of academic. As technology is moving forward, many new tools are needed to be developed and be used by students and teachers, who need to utilize these tools for a better learning experience. Learning from a distance has it's own challenges and problems and should not become more complex, but easier and fruitful for end-users. The users must not be involved in the complex processes of setting up the application environment. Everything must be handled and prepared by the tool, so enhances the user's experience.

1.3 Virtual Reality in Education

VR technology and its usages in the educational domains has a long history and with the current vast availability of consumer-grade VR hardware (Head-Mounted Displays, or HMDs), allows creating large-scale experience with reasonable and affordable costs. Therefore, it is possible to employ immersive personalized and unlimited VR experiences in virtual classrooms in the near future. However, transformation from traditional classes to digital ones, reflect a critical step when implementing VR environments for educational purposes which requires further study and research [19].

VR brings the sense of immersion in a 3D space by allowing to experience and interact with simulated objects. This technology has removed the boundaries and the only limit is the imagination. There is no limitation in the design of the environment (in which the user is hosted) and the user can be transported into any desired situation. The user can interact directly with computer generated objects. These countless features of VR have important reason for early adoption of this technology as an important experimental training tool [1]. VR has been employed in different number of domains such as aerospace and military, where training with real life equipment and in real situations is very costly and dangerous. In fact, the necessity to answer such these needs and developing of early VR systems, with features including interactivity and immersiveness, was the major reason of different studies and fast paced progress of VR.

A study of using VR in a business meeting, shows that compared to video-based meeting, qualities such as presence, closeness and arousal has improved. Female participants have declared their preference of avatars over real-life imagery. In another study of a guided VR field trip, the same factors, co-presence, social presence and engagement have been found with high ratings. VR in education has been suggested in respect of its benefits in increase of presence, motivation and learners' engagement [12] .

VR has been progressed significantly in recent decades and it is no surprise that many VR-based studies, simulations and trainings have been done in the last 30 years. Including studies and professional training simulations on emergency and safety, medical, surgical, aero space, educational, mental disease and many more. The major part of these studies, focus on the practical potentials of VR while superior learning outcomes with VR technology has been remained scarce. Simulation sickness is of the reasons that can be stated, specially in the early versions of VR hardware and HMDs. This side-effect is disappearing with the development of new devices and technologies which is observed in recent VR training studies in which stronger cognitive effects are shown. For example an earthquake safety training was stated to be drastically better with VR-based training course than watching pre-recorded videos.

The other part of this new technology that needs further research is the effects of training with partners. Most of the current VR-base training focus on one participant in the virtual environment while in many scenarios, being and co-operating with other participants is necessary. One reason of not have been done many researches on multiple participants can be on heavy computational load and costs. Having multiple VR hardware that are synchronized and attached to one base computation unit is not easy to afford for and is very complex to be handled. On the other side, developing the software and the virtual environment hosting the participants and managing their behavioral state is not simple task for small groups of researches, specially when deeper programming knowledge is required.

Emerging technologies have provided new doors and opportunities for creative use and implementation of new ideas. VR technology is one of these creative technologies that has been created for decades but its capabilities and potentials are still being explored. In recent years, due to the production of higher quality and cheaper products related to VR tools, the creation, development and use of VR-based products have also expanded and can be used in various and more diverse fields.

Exploring the possibility and usages of VR in the STEM (Science, Technology, Engineering, Mathematics) fields, has been interesting and full of positive results. Studies on celestial bodies and treating Alzheimer's are some good examples [21]. Other supportive technologies have been developed and progressed beside VR technologies that together have brought a wide range of new and easy-to-use tools to facilitate the process of development for researchers. Apart from STEM-related fields, researches on linguistic science and language acquisition have become more frequent in recent years [29] thanks to these technologies that let the researchers and learners to focus on their objectives, not on the infrastructure or technological barriers. Specially for learning languages other than English, there are a breadth of unexplored ideas, possibilities and applications to be implemented and studied.

Beyond the potential practical benefits and solutions implementation of VR, unique affordance associated with effective learning is also offered. Employing constructive learning theory and emphasizing on dynamic learning, invites direct and non-symbolic interactions and experiments with the environments and elements which provides active exploration and shifts in perspective.

Responsiveness of the VR environment and the intrinsic motivation and engagement matter as external stimuli are blocked out. Heptic feedbacks from touching in VR environment address to multi-modal interaction. In virtual settings, like virtual classroom, these multi-modal interactions lead to immersion. In addition, VR hardware are integrated with different sensors and micro-chips that can track any kind of movement (both in real and virtual world). Since all the user's behavior can be tracked in real-time, there a possibility to process this data to build a more precise model of learner and how to adapt the learning process to address individual characteristics and behavior. For example by showing personalized messages and feedbacks.

Compared to other scientific studies, training based on VR, in the major part of subjects are new and still more studies are needed to get closer to benefit the most out of its potential. Respect to on-screen learning, VR-based learning has shown dramatic improvements in the learning experience. Unique features of VR and its unlimited given opportunities (presenting in three-dimensional and multi-modal forms), are motivating to design and develop new ideas in educational domain.

VR provides better conditions for participation and evaluating situations. Although VR has a long history in education and training section, it needs more investigation in the digital transformations of classrooms.

Technical problems are one of inseparable parts of remote learnings. In real situations, problems such as setup difficulties, distractions, viewing-related discomforts and internet connections should be considered. Different studies of technologies like video-conferencing state that although it can be good for remote learnings, the existence of technical problems or distractions are inevitable and common. VR is not an exception and it is also has its own issues [12]. One of the common issues reported in VR hardware is motion sickness which happens when users put on a headset and enter a virtual world they feel dizzy or nauseous.

1.4 Social VR Platforms

In 2019 an online virtual poster session was held in a VR platform at ACM UIST [11]. This event was held in Mozilla Hubs[1] investigating the possibility of enhancing aspects like networking and social activities of a virtual conference. The participants used their own VR devices which lets them immerse in the VR room designed by event managers. The research of this event showed an increase in the sense of user presence. Participants stated that they felt more immersed in the VR space and co-watching the talks like they were watching the events with their friends.

Back to 2011, IBM hosted a business event in another VR platform named Second Life. This platform provides an online 3D world with avatars and objects and let the participants to interact with each other. This event was a successful one of its kind and some technical issues were reported. Two years before, in 2009, another event was held for a Program Committee meeting of IEEE VR. The results suggested that not many participants had good experience and many suffered from technical issues. Also users preferred face-to-face meetings than virtual meetings, likely due to the lack of feeling related to their presence in VR.

In studies about VR and virtual classrooms, presence and the related feelings, have been always the most notable feature that VR can add to remote learning [11].

VR platforms can be accessed with both VR headsets and desktop computers. It is not possible for every user to afford a VR hardware or even due to other issues. The VR platforms have given the option to the users to select the level of immersion. Some studies have been done on the differences between VR or desktop viewing and comparing which one is advantageous. Neither using VR headset nor desktop viewing approaches consistently been found better. In some cases, using desktop approach outperformed using VR headset in learning, environment navigation, or memory-based scenarios [12]. One of the reasons can be due to the cognitive load that usage of VR headset has. Another study show contrary results, in which they found for spatial learning in a high-fidelity space, using VR headset is much appropriate and useful.

The fact is that, the case and the domain defines which tools are more appropriate to be used. If there are more interactions in a 3D space, using VR headset can be more useful, which gives more level of immersion to the user.

In a study which was about a virtual field trip [18], it was asked for a guest lecturer to be present in the virtual classroom and present to the students. The opportunity to be connected to the virtual classroom from anywhere in the world, let the guest to be in on an equal footing with the main instructor. The guest has all the features and tools available for the teacher and form the time he/she enters, can handle and manage the class with his/her lecture. Previous knowledge and familiarity of this VR platform and virtual environment, was one of the key elements of the successful lecture of the guest. The guest knew well how he should control her avatar and behave in the classroom and could deploy nonverbal actions and behavior to support his teaching and ultimately be more effective. In the post-question of this study, one student stated that: “It was a strange feeling for me, having a stranger among and beside us for the first time and from a far distance without even knowing or seeing her before. I could sense her commandeering of the personality and the environment. I focused on her avatar all the time, how she spoke and her motions and how she fluently guided us in the environment.”

[1] https://hubs.mozilla.com/

Using tools is not always without challenges. Known as technical challenges have always been around the end-users whom are affected directly and mostly. Most of the studies have claimed positive feedback and the positive effects of implementing virtual classrooms. At the same time, they have mentioned technical difficulties that with wiping them out, the virtual experience can be improved. For example, due to the high technology and heavy data loads that these VR platforms needs to run, accessing them is dependent on high speed and good quality Internet connection. Unfortunately low speed Internet connection leads to laggy VR systems which makes it difficult to stick with the class.

1.5 Pedagogical Agents

In recent years, utilizing and developing intelligent pedagogical agent in learning environment in order to facilitate learning is increased. Researchers have developed new-fashioned learning environments such as educational realistic simulations.

Pedagogical agents (also known as virtual humans or synthetic humans) are visually-present in the virtual environment and regardless of their physical characteristics, they are meant to facilitate the learning process [13]. The investigation of their effectiveness for learning shows conflicting results. Positive effects were found in a meta-analysis while a systematic review states that major part of the studied did not show remarkable differences. This difference in found results can be to some extent related to the settings applied in the studies and in fact, how effective are the studies and used methods themselves, apart from the usage of pedagogical agents in the environment. The aforementioned meta-analysis suggests some specific conditions that these agents can be more effective in learning. The effectiveness of pedagogical agents presence in learning environments still needs to be investigated and researched more, as their physical characteristics and intelligence are understood by peer-learners.

Pedagogical agents have both internal and external characteristics. Internal characteristics imply how an agent thinks and makes decision to behave or say something. And external ones, focus on how the agent should look like to match the environment's needs and style. Researches have been done on both aspects, however most of them utilized information delivery agents, than a learning facilitator. This can be due to intrinsic feature of the agents. Designing and creating advanced and intelligent agents that do more than just speak or gesture or resembles the features of a real student is complex. Despite these challenges, researchers have investigated some internal features of the pedagogical agents, such as the impact of reacting to course content or other peer's verbally and visually.

In a virtual environment, before any interaction is made with the pedagogical agent, the learner's perception of the agents is limited to it's visual characteristics. External features of the agents are one of the interesting research area which is directly associated with social model and the first impact it has on other learners. Researches have revealed that the agent's visual appearance is the most critical design feature as it effects the learners' perception. The physical features of the agent such as its gender, age, style should be in relation to the context of the learning environment and materials. The agents look can influence how it is perceived and consequently the learning. Other studies mention that agent's appearance should follow the environment's characteristics and extraordinary features could likely increase the cognitive load which potentially hinders learning. The agents must be mindfully designed and the aim of their appearance must be to maximize the effectiveness of the agent, not just designing a complex or unrealistic agent. Particularly it should be noted that learning is not increased by simply adding an agent in the virtual environment. Rather this is efficient when the agent is purposefully designed to be effective and believable through instructional strategies.

In most of the studies that exploited pedagogical agents (both internal and external features), the learning outcome has been examined, not the effectiveness of the agents. One major limitation of the researches is that they focus on the content knowledge grasped by the learner not their effective state. This can be due to the low validity and reliability evidence of the instruments purposefully designed to be implement pedagogical agents.

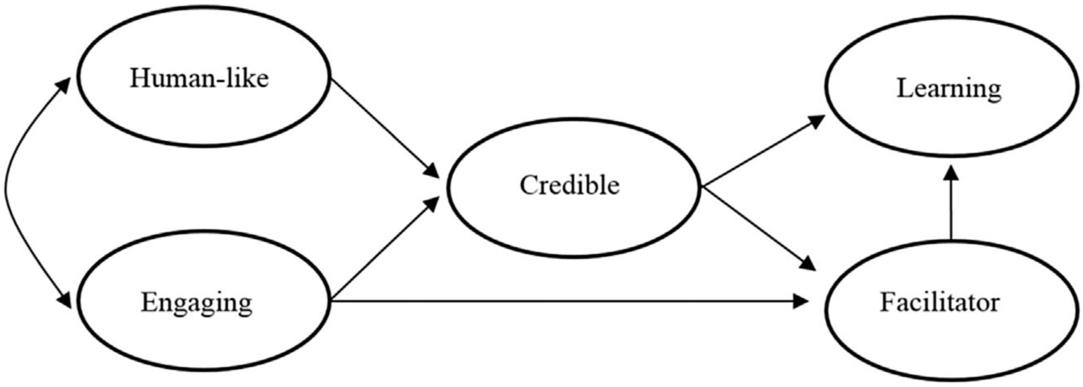


Figure 1.1. Hypothesized model of how learners' perceptions influence learning [13].

In 2010, a study was conducted to assess the educational effectiveness of likable, dislikable and neutral agents. The purpose of this study was how perceived the different types of agents and this learner's perception can affect the learning results. With two experiments, two factors, motivation and learning outcome, were studied. The findings show that appealing social cues are important and motivate the learners to responds and interact more. The students that worked interacted with agents with unappealing appearance and voice, showed lower transfer text scores than the students that were more interested in the visual appearance of the agent. Differently appealing cues might trigger social responses and it could affect the learner's motivation and learning.

The Figure 1.1 shows a formulated and causal hypothesis for how an agent may facilitate learning through virtual media designed in the research [13].

However, it worth to note that the interesting appearance of an agent or being interested in it does not necessarily leads the student to think deeper about the content to learn more. This fact is also valid for the conditions that the agent interacts enjoyably, which does not mean the agent's behavior enhances learner's attention on the content. The pedagogical agents' aim should be to add significant positive impacts on learning, not just fill the environment by their presence.

1.6 Conversational Agents

Human-computer interaction has been progressed and changed in past years. It is not anymore a simple User Interface (UI) dedicated for users to interact and command the software with pressing just a button. Nowadays many applications have applied new type of UI that let the users to talk directly to the application and command via their voice.

This technology has had opened many new doors of opportunities for the researchers and tech companies to facilitate and fill the gaps that was not possible to be solved before. With the help of AI, now it is possible to have a human-like conversation with a computer program where the user uses his/her voice as the command input. It means the interaction is upgraded to a higher level in which the user needs to use his/her mother tongue language. The main objective of these conversational interfaces are to simulate an intelligent human-like conversation in a way that the interlocutor feels as much as possible, like a conversation with another person.

AI is used to implement an intelligent behavior on the agent. The Conversational Agents (CAs) are the first consumer of this new technology that employ Natural Language Understanding (NLU), based on some defined conversational flow and structured interactions, can generate a human-like behavior.

A CA is a computer system intended to converse with a human and people are going to use it more than before. The advancement of the CA and the power of natural language can be used to communicate.

CAs also known as chat bots or computer assistants, can talk to humans and the users. These agents can be used in different ways, in smartphones, websites and home products such as speakers, in mobility and many other stuffs based on their designs. They can interact with user due to the recent advances in Natural Language Processing (NLP). CAs can presently give a modern helpful way of connection with clients in a specific way. In various applications the operators can be used to computerize an arrangement of assignments and forms. There are numerous victory stories encompassing the utilization of CAs. They can be seen in every aspects of life such as education, marketing, healthcare, customer benefit,finance [8]. These agents have great potential to use in real world.

In spite of the fact that conversational operator innovation has developed over time, still it requires investigation on how operators can fittingly include esteem to technological learning situations.

A CA or an artificially intelligent tool is an intelligent agent that that interacts with human via auditory or textual methods. Behind this intelligent interaction, a set of tools are used to identify the human words and meanings. Many agents across a wide range of domains such as customer service, e-commerce, healthcare, education and training have been developed and deployed. Most agents are accessed through virtual assistants such as, Apple Siri, Google Assistant or Amazon Alexa, or via messaging apps such as Facebook Messenger or WeChat.

The structure and functionality of CAs is derived from human conversation. Based on decision trees (handled by artificial intelligence), these bots process the human conversation obtained and through natural language processing the meanings are extracted. The NLP allows algorithms to understand, interpret and manipulate human language. Moreover, some advanced agents implements advanced technologies like Deep Learning to learn from conversations in real-time.

The structure of agents is based on two important components. The artificial intelligence is the core of their logic in which all the thinking and decision are done. In fact, it is a set of complex arithmetic computations and advance algorithms that let the bot to decide based on the previous training. The algorithm learns from what has been told to bot (model definition) and after that, can decide for new words that match the previous patterns. The main task is to extract what was the meaning of the words (told by the human), analyze the user entries and generate appropriate response (as it was trained).

Implementation of CAs into application is defined in two sections. The User Experience (UX) is responsible of making the conversation as natural as possible. The user should not feel strange or that he/she is talking to a bot. Eliminating such these feelings and experiences is on this intelligent and logical part.

On the other side, the UI, is the component to which the end-user directly interacts. These are the elements that the end-use can physically see and hear and based on the graphical changes or event, decides and follows the conversation with the bot.

For taking the most out of the agents and their hidden potentials, it is crucial to know how it is used. In a human-human conversation, in case of any events, any peer can adapt quickly not to lose the track of the conversation. But as the resources are limited or the technology is not ready yet, recovering from unwanted or unpredicted situations is hard for agents. Conversational Design is a set of rules or best practices for agents developers to both design a near-perfect conversational bot and provide the best experiences to the users. The Conversational Design is responsible for preparing and providing the human logic to it's artificial intelligence and algorithm.

One of the fields that the usage of agents have had promising results, is education. This area has showed great possibilities to deal with potential in terms of improving student's engagement, interaction and learning. Intelligent tutoring agents, efficient teaching assignments, course and lecture assessment and enhancing student engagement have been some of the challenges that are been faced and improved by the agents.

Although many years have been passed since the ideas of developing and using agents in the academic projects, but still they usually fail to deliver an accepted level of experience to the learners. It can be due to several reasons including technological problems, inadequate usability, not clear context and inappropriate responses. These problems lead to a low level of satisfaction which finally leads to a unrealistic conversation.

In rule-based method, the agent is designed to behave in a range of defined rules. In such systems, when a user asks or says something, the agent tries to find the appropriate answer by using some pre-defined rules. The other methods used in agent development is using a bank of possible questions and answeres, so that the bot is ready to behave in the defined and predicted conversation. Most of these agents retrive the appropriate answer using a large pre-stored question-answer database or from the previous discussion threads.

CAs have proven that they can enhance their activity and collaboration by interacting and purposefully guiding students to further interact and transfer knowledge among themselves [14]. As the level of learning deepens and students learn more during this joint activity. These facts from the results of research in the field CSCL prove that the use of intelligent agents are effective.

In fact, the opportunities created by the use of intelligent agents are endless. Their many features have made them able to be used and exploited in any field and practically open new doors for research and development. The followings can be mentioned as the features of intelligent agents.

- **Autonomy**
  The main feature of these intelligent agents is that they can operate automatically and without constant supervision, in the specified domain. It is enough to have a set of rules and regulations that they need to make decisions and show the right behaviors in accordance with the conditions and interactions of real users.

- **Perception**
  Intelligent agents are created to interact properly with their surroundings. They

can dominate the environment by understanding what is happening around them and paying attention to what is important.

- **Reasoning**
  In recent years, these agents have been able to appear and play a role smarter than ever. By considering more external parameters and factors, faster and better information processing, they can respond more intelligently than ever to interact with users that have more complex and different needs.

- **Cooperation**
  Each of these agents can perform their duties independently and based on the model they are designed for. They can also interact with each other to form more complex and complete collaborative groups. In this way each agent performs specific tasks and totally they move towards a single goal.

Chapter 2

State of the Art

2.1 Introduction

Thanks to technologies such as VR and low-cost tools (economical head-mounted display) a new experience of education has been provided to students. Each year, these technologies create new opportunities as they progress, and in the context of their presence, the atmosphere of distance education has also changed a lot. The experience of being present, listening and talking to others in a fully three-dimensional space (Immersive Educational Experience) offers great opportunities for research and development [21].

These intelligent agents play the role of a teacher according to the title and context, then interact with students by simulating human behaviors. Creating intelligent agents who can operate in educational environments is not limited to one-to-one interaction. Researchers have also investigated the need for intelligent agents who play a supporting role in educational activities. Research on supported learning shows that even the unplanned presence and involvement of an intelligent agent leads to greater student participation, which ultimately leads to deeper learning and the spread of knowledge among learning partners [14]. In another research [15] usage of Conversational Agent (CA) has had the similar results. Eventually enhanced the student's engagement in academic activities and leveraged the tendency to support that students often provide to each other. Also minimized the dropout rates among them.

Research on the impact of using intelligent agents is not limited to one-to-one situations. Extensive research has shown the effectiveness of using these types of bots. The use of intelligent agents in large educational settings such as universities and MOOCs has been effective in providing student ongoing support. Indeed, a CA may be able to compensate for the insufficient individual support of instructors, which constitutes one of the key factors negatively affecting retention rates [21]. All these studies show that bots play an effective, useful and direct role in academic applications and learning.

The use of intelligent agents in the field of education has been researched in various titles and goals. This research suggests that the use of intelligent agents with conversational ability and one-to-one interaction (student-bot) increases learning, understanding, interaction and motivation [8].

2.2 Background

In many training courses, facilities such as discussion forums and commenting have been used for learners. By categorizing the same and similar questions and answers, students can easily access the questions that are often asked and the answers given by the instructors or other students. [3] examined and analyzed the atmosphere and behaviors of learners in these chat rooms to understand how learners interact and cooperate in asking and answering questions. Although access to other people's questions, answers, and opinions allows for more learning and interaction with other learners, these connections are not instantaneous and do not interact instantly, which over time reduces the motivation and effort of knowledge. [4] provided an environment for direct communication that learners can ask questions and receive answers on a common topic. The results showed that this environment and type of interaction developed creativity of the learners. However, the problem of presence of a coach who can be present in this environment and intervene properly if necessary, has been identified in this experiment. On the other hand, the requirement for learners to take courses online simultaneously conflicts the flexibility of online learning.

Researches have been performed for non-VR platforms (without the use of VR hardware). Learning English along with Virtualized Co-learners is done by [5]. In this environment, students have to learn English to solve problems that occur in real situation in this three-dimensional space by having a connection with a VC. [6] also provided a virtual classroom for students to interact with their classmates using the Second Life platform. They were also asked to use this environment to solve the problems raised in the lab and the final exam. And [7] also used a virtual environment to teach software development topics. Attending this immersed virtual classroom made the students more collaborative and the believable interactions between them made the learning process more engaging.

Apart from research on the performance and role of bot in virtual education environment, some research has also addressed their personality and behavioral characteristics. In a study conducted features such as facial and body movements (which convey a lot of meaning to humans in this way) were included in the virtual agent. Addressing the emotions that agents show in reacting and interacting with users can increase the realism of what is felt inside virtual environment. In addition to emotions, some studies focused on the character of agents and the development of agents with high social interaction capability to accelerate the learning process. Some also tried to improve the learning process by using a virtual classmate. In this way, sentences had already been prepared for the agent, and the agent continued the conversation according to the circumstances and interactions with the student. They presented a model that, instead of using pre-prepared texts and answers, learned the appropriate answers from other (real) users and in future interactions with new users from conversations and answers used a new one.

Due to the rapid growth of technology, today one of the most important and common sources of learning are online MOOC platforms. Learners can choose a topic of interest from a variety of available topics and continue learning process at any speed. But this learning process takes place alone. This means that the learner continues learning process without having any immediate contact with the instructor or other students who are interested in the same topic that have chosen to learn. The main problem with this type of platform is creating a sense of loneliness and lack of companionship with other people who are pursuing the common goal in learning, and increase of learners' sense of isolation. Over time, learners lose interest in learning and stop learning [10].

Unidirectional education, lack of communication between two or more people, lack of reliable environment for the learner to ask questions, are all factors that many researchers have tried investigate and provide solutions to solve them. Numerous studies have introduced solutions such as chat rooms and forums to establish more communication between online learners. However, these methods do not solve the problem well. Because of the low participation of learners, these environments do not provide the expected enthusiasm and engagement.

Some studies have transformed these environment into 3D spaces, where learners can visually see and interact with their peers. When learners have an idea of the peer learners or those who accompany them in the learning process, it has an effective role in improving the deep motivation and perception of learning. Previous research has shown that seeing other students in a VR environment has had a positive effect.

With the help of new technologies, it is possible to be in the form of a three-dimensional avatar among other people. This way, each person connects to a virtual room using a computer (client) and hardware (VR devices like HMD) and can talk or interact with other people in that room. Using these tools, researchers have created virtual classrooms for students to feel being in these three-dimensional and virtual environments, such as Second Life, Mozilla Hubs and other social VR platforms. In VR environments, each user is transformed into virtualized character (avatar) with capability of having realistic interactions and sense of immersion.

Users can move, see and hear in 3D space by using tools. These lead to a more realistic understanding of this virtual world and give the user a closer sense of presence and immersion in this space. In addition, users are not alone in these environments. They can talk and communicate with other users who are seen as avatars. Surf the environment together, find other friends, and work in groups of multiplayer. This companionship and being with other (real) users, along with the many capabilities that platforms offer them, such as:

- ability to load 3D objects;
- ability to share videos and files;
- ability to live broadcast using the camera connected to the computer;
- ability to choose the desired appearance.

All these features provide the conditions to form a new type of online communication. New conditions that convey new experiences and feelings to users. Attending a virtual classroom with these features increases motivation and social interactions between students [19].

Easier access to software and hardware tools have made it easier for users to be present in 3D and virtual environments every day. Today, many HMDs are more affordable. The abundance and low price of the tools provide a great opportunity to use VR in the field of education, particularly given its provided immersion and potential for teaching.

The results of various studies show that the use of VR can be very effective both for teaching and learning. For example, in an human anatomy training experiment. The post-session knowledge tests results, explain that in a study that both VR and Augmented Reality (AR) tools are used, the students were more active in the learning process and had a closer sense of seeing and touching the content presented.

In another experiment in an engineering school and to measure the effectiveness of learning with the help of VR and learning without it was done. Investigating the post-quizzes results, shows that the use of VR, compared to when it was not used, was able to improve student performance during the final test. These results can be because of providing a higher and deeper understanding of what is presented in class. Regarding to the same results found in [12], it can be interpreted that using VR tools for learning spatial topics, can be advantageous, as it provides a 360 degree sense of the environment and objects. This also shows that VR has been able to play a successful role in teaching these topics and conveying concepts to students better.

In addition, the coaches have been able to increase their abilities and meet their needs with the help of VR. VR can be used as a tool to help educators develop specific skills that are effective in their training process [20]. VR gives the instructors to load 3D objects and demonstrate different aspects of the object. Letting the students utilize their visual senses.

In addition to the processes and methods used for teaching and learning, there are other important aspects that may not seem important at first glance, but studies have shown how useful they can be in conveying a positive feeling and motivation. The virtual environment that is designed and developed with the help of VR has no limits and is usually designed or even selected (using sample environments) depending on the subject of teaching. A virtual environment is a space that allows people to move around and experience. The more purposefully this environment is created, the more immersive the students will be in the environment and the more motivated they will be to interact with both the components inside the space and other students.

Studies have also been conducted on the impact of virtual environment and their comparison with traditional education methods among high-school students. They presented a history classroom, using VR tools, to answer the question, what is the difference between learning in this space compared to the normal classroom environment (face-to-face class)? The results show that students enjoyed learning history more in a VR environment, as they could immerse themselves in the characters, events and the history.

In another study, students and their teacher took part in a virtual field trip [18]. This research was done in order to assess the possibility of long-term use of cheap and affordable VR tools. Whether continuous access to these tools and maintaining a learning process in the VR can be effective in the long period or not. This research was also conducted with the help of school students. Research was conducted on the presence of students on a trip in two traditional ways (using whiteboards and slides) and using VR tools (HMD). In this way, each time by asking questions before and after the class, students' opinions and the effectiveness of both methods were measured. The results of these experiments show that students are more motivated when attending virtual classes.

2.3 Previous works

2.3.1 colMOOC

This study aims automating a conversational intervention from a computer agent that interacts with small-group of learners in chat-based environment. The proposed facilitation strategies are operated by a teacher-configurable CA, which adopts an event-driven approach and operates on the basis of specific patterns that serve as intervention opportunities. Without requiring a large development effort, this kind of agent-based facilitation can enable MOOCs to provide valuable context-responsive support during chat-based learning activities, scaffolding and improving the quality of peer discussions [8]. This project is designed to be used in MOOC platforms and the features of it can be named: natural conversation (chat-based), low cost and reusable.

This research project is designed to provide a low-cost and salable solution that can play a constructive role in collaborative learning in accordance with the environment for which it is prepared. This new model of interventions and interactions speech, operates on the basis of MOOCs. The main focus of this project is on the aspects of participating in the learning process, practicing as many students as possible in this process and problem solving.

The MOOC platform provides learners a dialogue environment and indirect and non-synchronized communication. But there is no way or principles to properly understand the activities and types of interactions between learners who use these environments. In the absence of such methods and strategies, the presence of an agent who can intervene to improve conversations and activities between learners is essential. This agent, by being in groups of learners who are in MOOCs environments, automatically monitors the activities and leads them to get better.

Due to the growth of technology and the diversity of training courses, there is a large amount of MOOC platforms that each responds to a subset of the learners' needs. Providing a applicable solution, method, and project that can be reused greatly reduces development costs. The development of the project colMOOC plays a role in different platforms by focusing on reusability and scalability.

The design and implementation of the project colMOOC is such that it can be easily controlled and adjusted by the instructor. In particular, the agents developed in this project are subjected to the rules and regulations set by the instructor. With a good understanding of the learners' behavior, the instructor can access the agent creation environment using a visual tool called *editor*. These smart agents are designed as smart tools that make the strategy desired by the instructor more effective. The instructor determines the path and activities of the agent by making a set of logical connections between the behaviors and interventions that the bot should do.

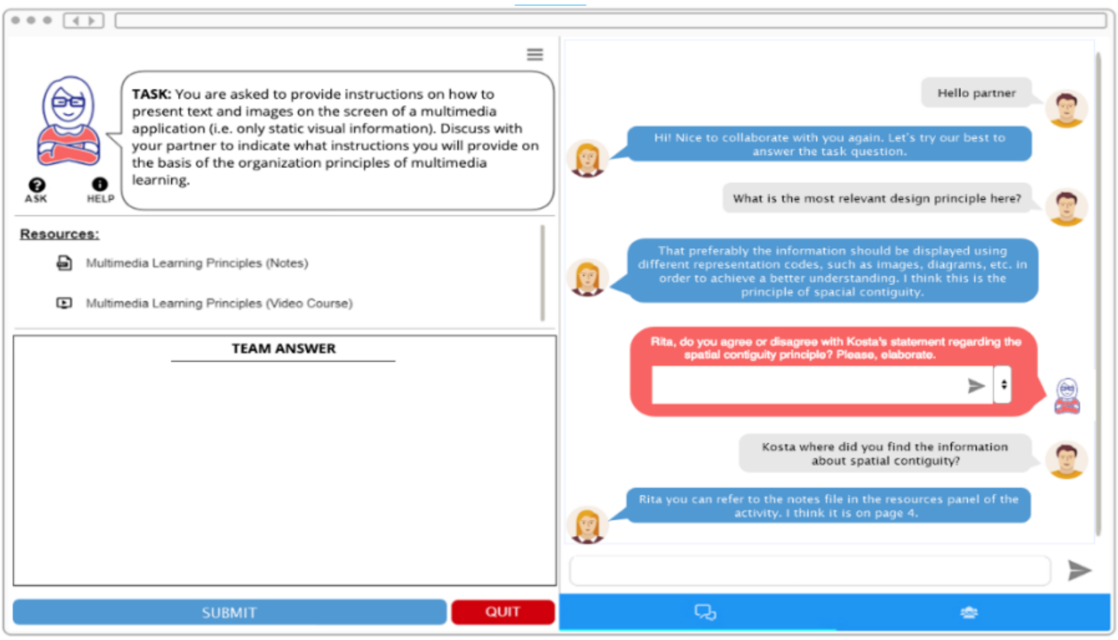
According to this framework, these agents deliver a series of interventions (or moves) as a means to trigger productive forms of peer dialogue and scaffold students' learning (Figure 2.1).

Finally, the set of commands and rules that are added to the agent by the instructor determines the scope of the activity and the main model of the task. Learners in the learning process need to do certain things to measure their learning progress. These tasks are handled automatically by using the agent made by the instructor. In a chat-based environment the agent conducts a conversation with the learner and tries to convey the question that the instructor wants to ask the learner, then answers the question with his active participation.

2.3.2 Bazzar

Collaborative learning, which is based on the interaction between several students, has reached a new level of effectiveness with the presence of intelligent agents. One of the major problems with collaborative learning in most MOOCs is the impossibility of synchronous interaction between learners, and on the other hand, there is no possibility for learners to participate with their chosen friends. Practically, the lack of this feature has made collaborative learning not implemented properly in the context of MOOCs and they can not use its benefits.

The project, Bazaar [9], has been developed to make up for this shortfall. Bazaar provides a special collaborative environment for the learners. In this space, learners can



TASK: You are asked to provide instructions on how to present text and images on the screen of a multimedia application (i.e. only static visual information). Discuss with your partner to indicate what instructions you will provide on the basis of the organization principles of multimedia learning.

Resources:

- Multimedia Learning Principles (Notes)
- Multimedia Learning Principles (Video Course)

TEAM ANSWER

Hello partner

Hi! Nice to collaborate with you again. Let's try our best to answer the task question.

What is the most relevant design principle here?

That preferably the information should be displayed using different representation codes, such as images, diagrams, etc. in order to achieve a better understanding. I think this is the principle of spacial contiguity.

Rita, do you agree or disagree with Kosta's statement regarding the spatial contiguity principle? Please, elaborate.

Kosta where did you find the information about spatial contiguity?

Rita you can refer to the notes file in the resources panel of the activity. I think it is on page 4.

Figure 2.1. An example of an agent intervention (red bubble) displayed in an online chat activity [8]

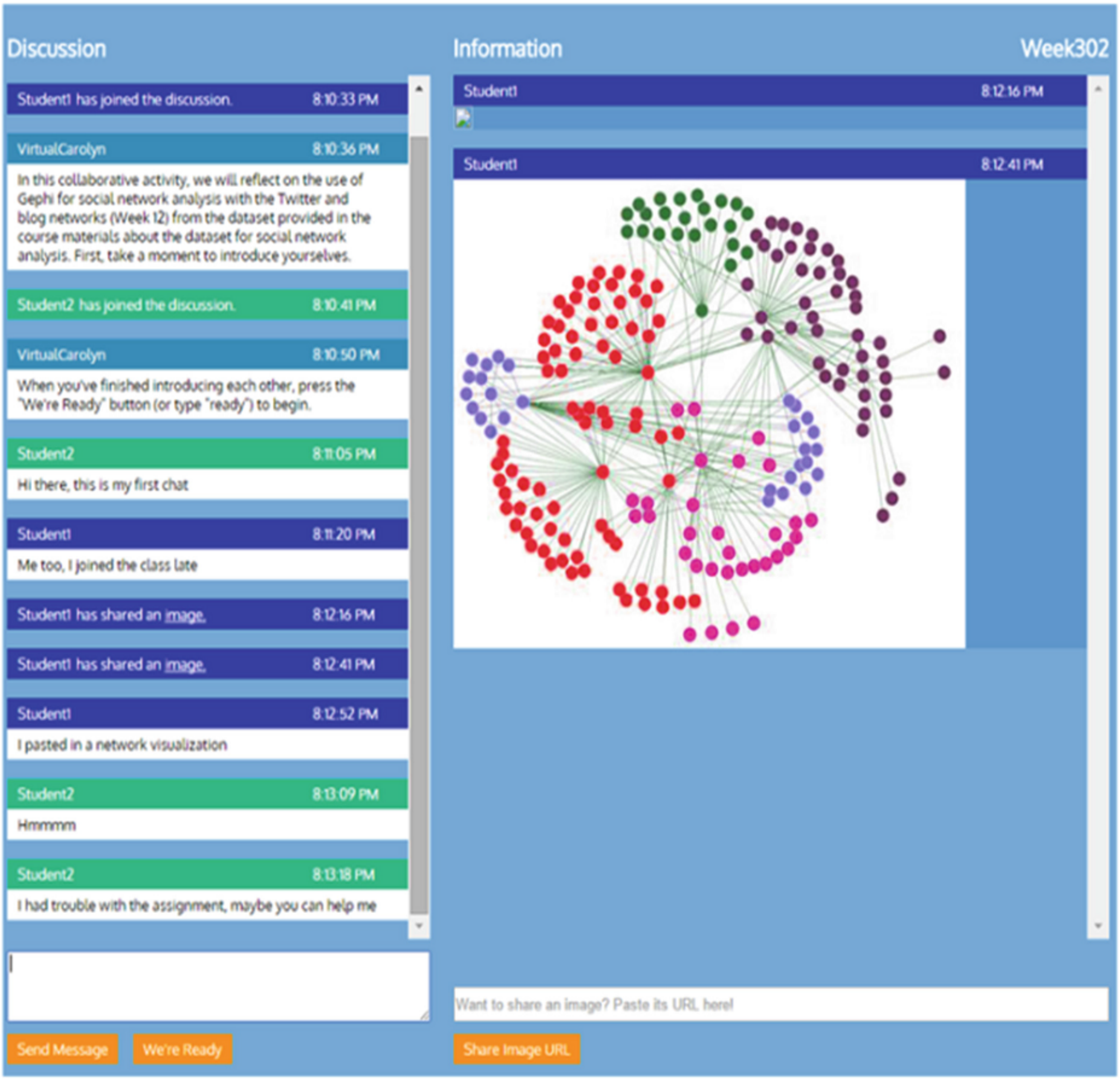
interact with each other in the form of written dialogue. Learners can form their own groups and join their favorite groups and exchange knowledge. In each group, with the presence of an intelligent agent who can talk and understand conversations like other learners (humans), learners are encouraged to be more active, and this intelligent agent monitors these activities and leads them to the positive side. This project is based on two other researches which have already been established and implemented on the role and impact of intelligent agents in text dialog environments.

Implementing this project in the context of MOOC exposes it to a large number of users. There were over 20,000 users on this MOOC platform. On the other hand, because of the feature of this project, which gives learners the ability to create a group and publish it. This feature is a positive point for many users, but finding the right peer to join the group (which is done automatically by the system) takes long time and this was a negative point according to users' point of view.

2.3.3 Embodying Historical Learners' Messages as Learning Companion

This project [10] that is designed and implemented in a VR environment, allows students to observe and learn lessons with their virtual classmates in a virtual classroom. The purpose of this project is to investigate a technique for generating appropriate answers and the effect of changing environmental parameters on students' learning process. Research shows that students in a virtual classroom with fewer students are more comfortable and they can focus on the subjects taught better. The high activity of classmates (if they are many virtual agents) is more distracting and confusing than being motivator or helpful. Also, being present and immersed in the three-dimensional space of a virtual classroom increases students' concentration and accuracy.

This project is designed so that a real student is present in a virtual classroom along



Student1 has joined the discussion. 8:10:33 PM

VirtualCarolyn 8:10:36 PM
In this collaborative activity, we will reflect on the use of Gephi for social network analysis with the Twitter and blog networks (Week 12) from the dataset provided in the course materials about the dataset for social network analysis. First, take a moment to introduce yourselves.

Student2 has joined the discussion. 8:10:41 PM

VirtualCarolyn 8:10:50 PM
When you've finished introducing each other, press the "We're Ready" button (or type "ready") to begin.

Student2 8:11:05 PM
Hi there, this is my first chat

Student1 8:11:20 PM
Me too, I joined the class late

Student1 has shared an image. 8:12:16 PM

Student1 has shared an image. 8:12:41 PM

Student1 8:12:52 PM
I pasted in a network visualization

Student2 8:13:09 PM
Hmmmm

Student2 8:13:18 PM
I had trouble with the assignment, maybe you can help me

Student1 8:12:16 PM

Student1 8:12:41 PM

Figure 2.2. Bazaar collaborative reflection chat [17].

with the number of virtual students (virtualized classmates). Each of these virtual classmates tries to behave naturally and performance by using the answers previously given by other real students analyzed and recorded in the system. The activity and behavioral features of these intelligent agents can be controlled.

To make the speech and training activity of intelligent agents more natural and logical, they have used a technique called *Comment Mapping*. In this method, the previous opinions and answers of real students are collected and provided to the agent.

In this research, two issues have been addressed. The effect of using and not using the Comment Mapping technique and the effect of the number of virtual classmates on the student's performance and focus. The results of various experiments (Figure 2.3) show that when the number of virtual classmates is less than 5, the Comment Mapping technique can be useful, and on the other hand, when the number of these classmates is high (20 virtual classmates), students are more social and successful. The researchers of this project believe that the Comment Mapping technique is due to the answers obtained from real students and also they are not produced and selected by system logic, they create a more natural atmosphere when chatting, which can be further explored in the future with more precision and experiments in online environments and other situations.

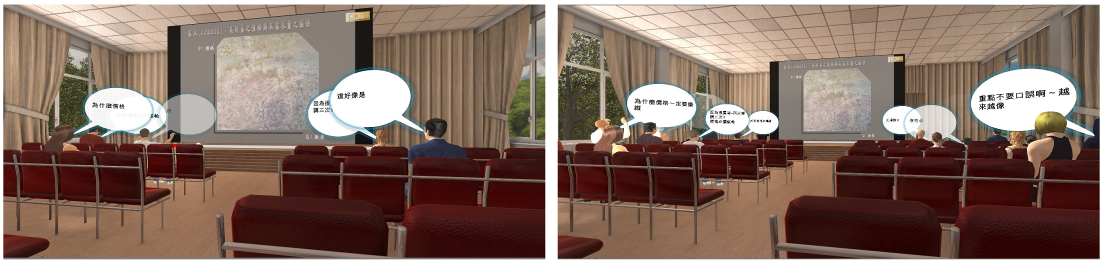


Figure 2.3. Left: The scene of the experimental condition with 5 virtual classmates as learning companions. Right: The scene of the experimental condition with 20 virtual classmates as learning companions.

Virtualized Learner denotes In the virtual classroom, the online avatar's behavior that is handled and controlled by the real student. On the other side, there are other virtual classmates. These avatars are completely controlled by computer which are called "virtual learner". During the different experiments done by real students, they have provided different comments or responses to same questions. For each specific time points of the lecture video, these comments are gathered together, grouped in similar comments, so that in the future experiments, the top appropriate (and chosen) comment is reread by the virtual leaner (avatar).

In this research two aspects of a learning with and among other classmates is evaluated. How it is effective using Comment Mapping technique with low number of virtual classmates (5 person) and with high number of them (20 person). The results implied that although using the Comment Mapping technique is useful in giving valuable and helpful comments during the lecture, but it is very important to provide appropriate condition for the real student to read and have control over what is told by other classmates. Giving too much good information can be much distracting (in case of large amount of virtual classmate) than being effective. It can be considered that when the users are among many other virtual classmates that are commenting on the lecture points, they tend to focus more on the comments and can not perform social behaviors. They try to keep their learning performance up.

In terms of interaction among real user and virtual classmates, the foundings result into a contrary fact to what is effective on the previous aspect. Social interactivity tests imply learning with high number of virtual classmates leads to a significant higher level of social interactions. This is true when there is no Comment Mapping applied during learning process and the real student, has a complete chance of interacting with the elements in the virtual classroom and enjoy his/her presence in an immersive space.

The virtual classmates are regenerated based on the verbal responses from other real students. So the creators and developers can manipulate these virtualized students and their behavior, their functionality and therefore their level of effects on the real student can be controlled.

One of the problems that may arise in the Comment Mapping technique is receiving repetitive or similar answers. In order to prevent the agent from repeating or saying duplicate answers, the obtained answers are monitored and the duplicate answers are removed. Also, the behavior and logic that each virtual classmate adopts during the class should be consistent, so, the similar answers are assigned to each unique virtualized student.

This virtual classroom is designed and developed in Unity3D[1]. The classroom is simulated like a traditional classroom with all the elements inside. For visualizing the characters (virtual classmates), MakeHuman[2] is utilized. Lecture is presented as a video playing content to which the virtual students react and give comments. These comments are synchronized with specific time points of the video. The experiment is held in VR and as the video is playing in front of the class, virtual classmates use the Comment Mapping technique to express their comments and thoughts in a dialogue boxes with body motions. Eventually, the real user can observe the behavior and read what his/her virtual classmates are saying about the topic presented in the class. This gives the opportunity to the real student to be among other active students and learn from them.

Virtual classmates express different responses and comments. Each comment has a meaning and a set of senses that that comment spreads. To be closer to a realistic presentation of virtual classmates, they matches their behavior according the emotions of that comment. The analysis approach, is used to classify what the virtual classmate intent to say and select the appropriate emotion category, such as: joy, relief, sadness and anger. Then the appropriate animation is picked up from a MOCAP database. This MOCAP[3] file contains all the necessary information that is needed to animate the character and express the emotion with body movements.

The experiments were conducted in the laboratory settings. There were 100 participants in total (54 female and 46 male aged from 18 to 35) with prior experiences on distance learning or using online learning web sites. The selected course topic is about "Introduction to Economics", which took long for about 15 minutes (each session). During the lecture a set of comments were used (as virtual classmates' responses) collected from previous experiments. These set contained more than 400 time-anchored comments collected from more than 50 online students.

For assessing the learning outcomes, a set of pre-test and post-test question based on the lecture video and content were prepared. And a survey (14 questions of 5-Linkert scale) was designed to understand the learners' experience of using this virtual classroom and to measure the perceived social engagement and interactivity and perceived focus and attention.

2.3.4 Learning Chinese

This research [21] was conducted at Sino-British university with the aim of examining the students' willingness to use VR tools to learn Chinese. In this study, international students who had sufficient knowledge of English, participated in groups in a three-dimensional virtual environment. This virtual classroom was hosted in Mozilla Hubs, provided a virtual classroom for students to participate and learn Chinese. By reviewing the results of the questions asked after each session in the virtual classroom, the acceptance and satisfaction of technology users to learn a new language is analyzed.

In this research that has been done specifically in the field of learning foreign languages in online and virtual environments, it has always shown the learners' satisfaction and significant growth in self-efficacy. In this study, the experience of learning Chinese in a virtual world has led to desirable and positive results. These desired benefits, were

[1] https://unity.com/

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

[3] Motion capture (sometimes referred as mo-cap or mocap, for short)

the results of the pedagogical nature of the VR and virtual world (as opposed to its entertaining nature).

Other positive results show signs of improving the ability to speak and pronounce Chinese words after using VR tools. This can be due to the safe space that the virtual environment gives to the student and the student tries to express his abilities and use them with more calmness and concentration.

2.3.5 7-Week VR Class

During the spread of COVID-19 virus many universities have searched and planned for alternative solutions for keeping up the learning process for all students. Against the previous researches that have been done in a controlled and laboratory conditions, this seven-week study [11] was conducted to provide remote access to a virtual class for all the students. The teacher and all the students, met each other entirely in the online virtual class that was hosted in Mozilla Hubs (Figure 2.4).

The students' comments during and after the class about attending such a class and learning in a different way gave them a pleasant feeling and experience. The results also show that on average, figures related to the features studied such as: attending a virtual class, the experience of being with friends, a good understanding of what is taught in the class, the efficiency and use of the software, and the final experience in general, all lead to a positive points.

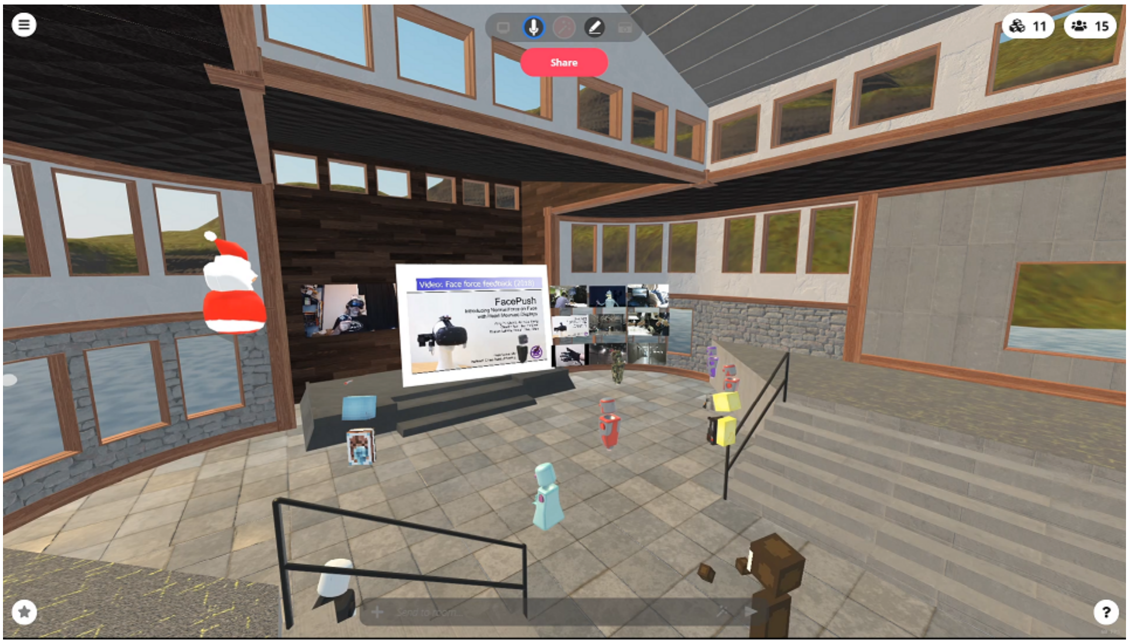


Figure 2.4. A lecture in Mozilla Hubs [11].

The answers to the questionnaire (Figure 2.5) suggests that although the use of the Mozilla Hubs does not depend only on the VR environment and it can be used without any special or expensive tools, in this experiment, all students were equipped with VR hardware so that they all had the same experience of being in the VR environment and out of it (using desktop viewing). One of the most important negative factors that students expressed in post-session tests, was the simulator sickness, which was an unpleasant experience for them. And students who did not report or experienced this, received a positive feedback on the use of VR tool.

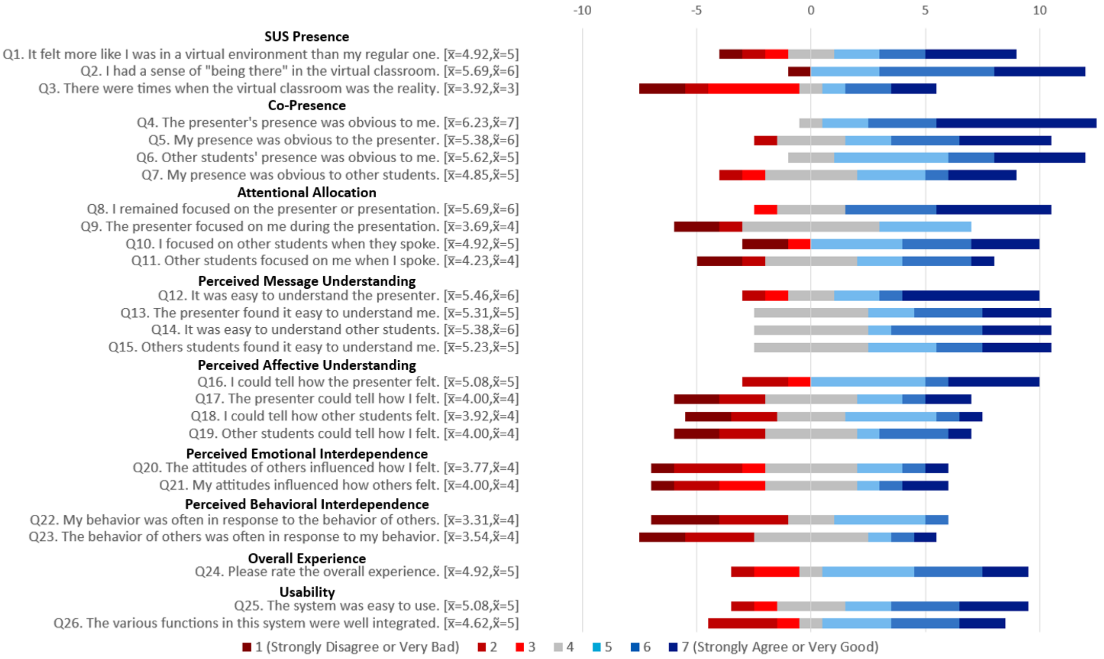


Figure 2.5. Main questionnaire items and counts of the responses [11].

Launching online classes definitely can not be without problems and users should always be prepared for different technical issues. These problems include poor internet connection and inadequate hardware. During the seven weeks that the class was completely held online, similar problems occurred. In the final evaluation, students were asked how they would rate the use of VR tools for such classes if all these technical problems were resolved. 84% (11 out of 13 students) gave a completely positive opinion, showing that although there may still be technical and infrastructural problems, the positive impact that users get despite all these problems is noticeable.

The responses of the students about their tendency to two types of classes are shown in the Figure 2.6. The students were supposed to state how they felt about the in-person class versus the virtual class held in Hubs. The figure indicates that the students seem to be neutral about which class type is better, however they are a bit tended to like in-person classes. Also, students seem to be slightly toward virtual class because of the feeling of confidence they have in Hubs, as they do not need to be seen physically, rather represented as virtualized avatars.

Students were also asked about the positive and negative points of VR lecture versus the in-person class. The responses are could be used as useful guidelines for future developments.

- **Positive Points**
  More engaging and interactive than real-life class (4 responses).
  No need to leave home (2 responses).
  Ease of use (2 responses).
  The possibility of embedding other electronic documents such as lecture slide or videos (1 response).
  The ability to re-access the lecture as it can be recorded (1 response).

- **Negative Points**
  Technical problems and difficulties, such as audio or video glitches or lag (7 responses, about half of the students).
  VR sickness (1 response).
  Feeling of isolation (1 response).
  Being distracted (1 response).
  Seeing own avatar (1 response).

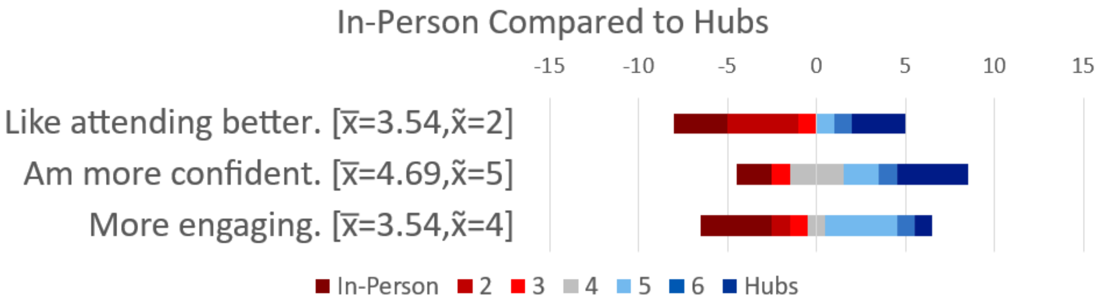


Figure 2.6. Responses comparing in-person lectures to hubs [11].

The students were also questioned to express how they compare learning in Hubs and a video-conferencing class. The Figure 2.7 clearly demonstrates a one-sided results between these two types of classes. They mentioned positive and negative aspects of VR lecture versus video-conferencing class.

- **Positive Points**
  They did not need to use webcam and everything was seen from the HMD (3 responses).
  VR seems more engaging and interactive (3 responses).
  Less distracting as they were immersed in the virtual environment (2 responses).
  The ability to load and see lecture slides (1 response).
  Being able to gesture (1 response).
  Fully immersive like being there in the classroom (1 response).
  Ease of use (1 response).

- **Negative Points**
  Technical problems and difficulties (5 responses).
  Discomfort from using the HMD (2 responses).
  Environmental distractions (1 response).
  VR fatigue and sickness (1 response).

Possibly, one reasons that the students felt more focused on the interaction and engagement, is that they did not to be present physically and they could be themselves, without worrying about critics of their look. And again, the technical problems were mentioned as the first issue to be solved.

2.3.6 VR Field Trips

This paper [18] describes a one-semester long study on a high-level educational experiment in virtual worlds in a large university in United States. Due to the spread of COVID-19, an instantaneous action was needed in order to fill the gap of not being present in

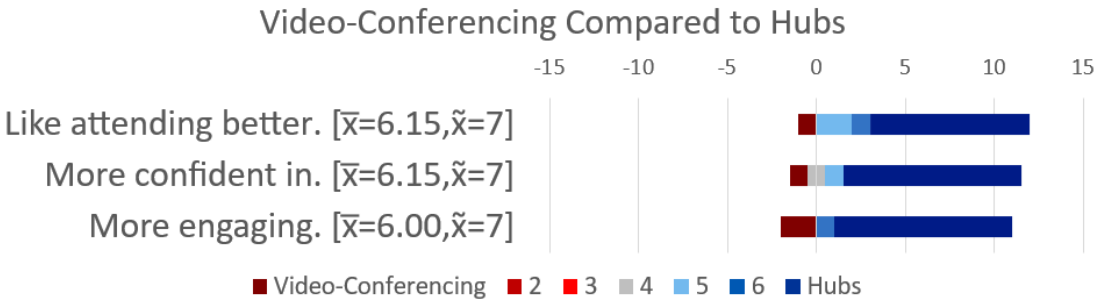


Figure 2.7. Responses comparing video-conferencing lectures to Hubs [11].

traditional classes. Already distance learning has been examined, but this forced situation became an opportunity for both teachers and students to connect remotely via desktop and VR headsets and experience their "Communication in Virtual Worlds" course, in a virtual environment rather than using video-conferencing tools.

During the separation of students from the university, the instructions became virtual and all members were connected remotely. One of the challenges was accommodating the students who were connecting via desktop equally with those who used VR hardware.

The objective of conducting this study was to collect the data of students' experiences in virtual environments when they are not co-located. The virtual environment brought the student closer to each other. Findings indicate that the students had more feelings of togetherness with their peer-learners and teachers. Seeing each other as embodied avatars, made them feel more sense of reality which led to less sense of isolation and separation in virtual classrooms or groups. The overall results of the survey shows a positive statements and they are enthusiastic about this virtual experience.

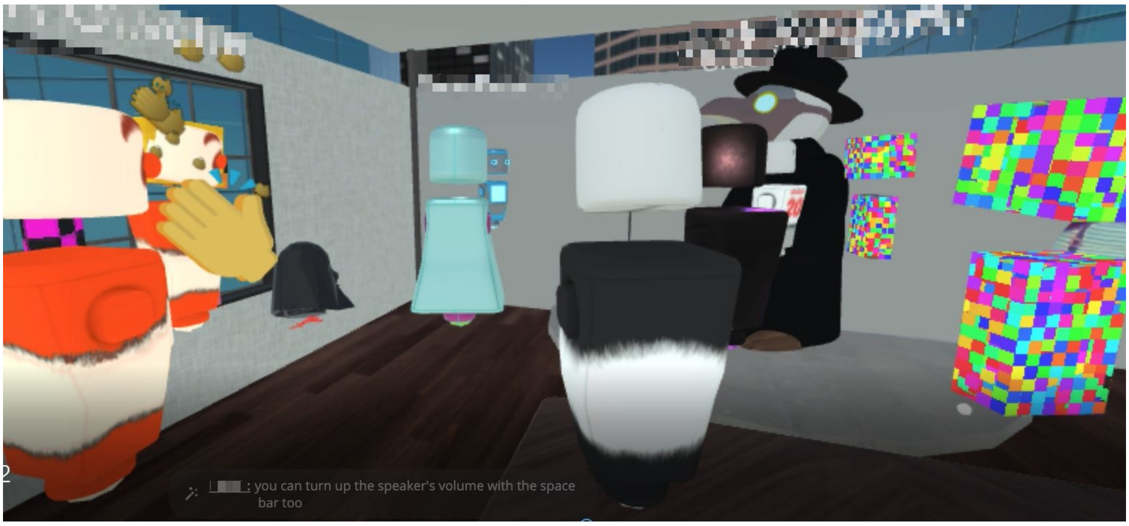
Each week, students participated in one class session on Zoom, and visited one virtual world or environment as a "field trip". During the scientific trips, they were guided and tutored by their teacher or guest lecturer and they discussed the material of the class. Totally, they visited three different virtual worlds in three different VR platform (Mozilla Hubs, Second Life, and Rumii). These were chosen because they are easily accessible to students using a desktop computer (Mac or PC) and a headset. Six of the twenty-four students opted to borrow a headset, while the others entered class sessions via laptop.

In this field trip, students visited a private room in Mozilla Hubs (Figure 2.8). Students selected stock avatars but were required to use names, so they were not anonymous. Students preferred stock avatars but had to use names, so they weren't totally anonymous. They interacted via chat and speech, but due to audio issues caused by the large number of attendees, some students had to log in several times. They were divided into small groups to explore the interface after listening to a brief lecture on social presence and nonverbal behavior.

After each session, a survey was answered by the instructors and the students. The findings are categorized in 7 sections, as reported below:

1. **To Be Heard**
   An indication of being noticed and heard by other peer-learners. Particularly students stated they felt "most heard" in either Rumii or Zoom.

2. **Closeness to Tutor**
   It was important for the students to see the tutor's face and it gave them feel more



[redacted]: you can turn up the speaker's volume with the space bar too

Figure 2.8. The tutor is the blue robot in the back corner and the students are listening to the tutor [18].

close to her. This situation was possible in Zoom, as the tutor was always facing the camera and every student could see her, but in virtual classroom, where she was embodied as an avatar, it was difficult for the student. Students mention the teacher was more active in Zoom while the guest speaker was more active in the virtual environment.

3. **Closeness to Peer-Learners**
   This statement can be considered as the *social interaction* and effects of the virtual environment. Regarding the students' statements, many of them felt more close and socially engaged while they were in virtual environment, rather than Zoom. "With zoom, most of my classmates did not speak or have their videos on, so I did not feel connected to them in any way, I saw their avatars moving and people were more willing to talk or chat, so it felt more authentic than Zoom."

4. **Verbal & Non-verbal Communication**
   In Zoom, the students showed more interest in using the chat-based communication instead of audio. This habit and behavior was ended with the direct intervention of the teacher. However in virtual classroom, students showed more interest on communication non-verbally. In the virtual classroom, the students formed a semi-circle shape and stood beside each other facing the teacher. They showed active listening by head noding and they also exploited using the emojis available in both platforms Mozilla Hubs and Rumii. Where a student stated "interesting, social rules still apply."

5. **Technical Issues**
   Most of the students were struggling with the technical problems and the low power of the VR platforms that can not handle too many students simultaneously. Although with these problems, the students showed steep learning curve in resolving their problems. In the first sessions, the teacher had to spend more than 20 minutes to explain the tools and features, so students could get what is their avatar capable of, but in subsequent classes, it was reduced to about 10 minutes.

6. **Class Benefits**
   Increased involvement has been identified as a benefit of extensive work on learning

in XR. “Overall, though, I found it to be extremely fascinating and engaging,” one student said. Even though it felt like a video game at times, I was shocked at how well I was able to stay engaged.” Several students, on the other hand, described their Zoom experience as "more easily distracted. Easier to zone out if cameras were off.” As a result, immersive systems might be able to reduce the risk of student disengagement while learning remotely.

7. **Headset and Desktop Users** The possible discrepancy between students who had access to a headset and those who did not was one question. Three of the six students who agreed to their data being included in this analysis had also borrowed headphones, so their answers were over-represented. Some students were apprehensive about borrowing a headset in the first place. “I’m not sure whether it would have made it harder or easier for me to immerse in a new world of distractions and adaption,” one student said when asked if they wished they had used a headset. However, I assume it would have deterred me more if we had used the channels for a longer period of time and at a higher level.” However, some students said that if they had been able to pick up headsets before the campus closed, they would have considered them useful. “I believe I would have been more interested in the virtual worlds and therefore found it easier to pay attention in class and avoid distractions,” for example.

8. **Classmates Benefits**
   Finally, although not extensively studied in VR, peer-learner involvement expressed by hand-raising actions can affect learners’ attention and visual behaviors in VR classrooms, which could be investigated further. [19].

2.4 Common Issues

Future research is required to fine-tune the design of agents and build mediation strategies that are both pedagogically useful and versatile enough to be used in a number of discussion contexts without a lot of setup time. [8].

One downside of such VR and online learning tools is that lack of social interaction, peer accompaniment, or immersion can affect learners’ motivation and performance [19]. In addition, realism in immersive settings may have a variety of effects on learning and interaction.

Several studies have been conducted to solve these problems, with the aim of creating more practical and interactive environments. For example, explores the design of virtual reality environments for classrooms by simulating real-world learning conditions and improving learning through real-time interaction between students and teachers. Furthermore, by synthesizing previous learners’ time-anchored comments, researchers find that when students are supported by a small number of virtual peer-learners created from prior learners’ comments, their learning outcomes increase. The existence of virtual teachers, in addition to virtual peer-learners, may have an effect on learning in VR.

When a virtual teacher was presented, learners interacted more with the environment and advanced more with the interaction prompts, according to study. These studies and findings suggest that the styles and forms of virtual agents in virtual environments may have a range of effects on students’ attention and interpretation during immersion, and that they should be considered. The measurement of real-time visual attention against similar configurations, which could be done with sensors like eye trackers, could not only help to explain learning processes, but also provide analytical information about experiences during virtual classes for interactive classroom transformations in VR.

They [9] showed some general problems in their project:

- the debate is focused on a long list of messages that students find difficult to understand and tutors and moderators find boring to track;
- the conversation that took place during the collaborative activity is no longer accessible, and the collaborative information that was created is lost;
- since text-formatted posts are excluded from real-life conversations and physical interaction, there are no possibilities for social gains from actual collaboration.

All of these shortcomings result in rudimentary collaborative learning activities, which are unappealing and lack interest, reducing learners' self-motivation and participation in their learning process.

2.5 Thesis Objectives

The basis of all the previous mentioned researches has been measuring the impact of changing the classroom space from a traditional environment and transferring it to virtual environment. Students interact with their instructors in a three-dimensional environment using VR software and hardware.

In this study, in addition to continuing the above research on the context of a virtual environment, another aspect of events within a virtual classroom is addressed. In fact, the objective of this study is to analyze the potential of CAs to considerably increase the engagement and co-learning with a virtual classmate (agent) in order to increase student's engagement, commitment and enjoyment of learning with classmates in VR. This CA behaves as a virtual classmate that understands what is being taught by the teacher and interacts synchronously with him/her.

The purpose of having this intelligent and active agent in the class is to show a good example of an active student in a class can stimulate the real student and spark new ideas to his/her mind. Additionally, the agent with its natural behavior, tries to keep the attention level high for the real student.

We analyze how an active classmate (CA) can affect the level of real student's engagement during the lecture by providing a synchronized collaborative learning space. We also assess the effect of CA's activeness on the levels of knowledge grasping by the real student.

Chapter 3

Methodologies

3.1 Introduction

This chapter deals with various aspects of the project presented in this thesis. At first, the reasons for choosing the subject taught in the virtual classroom are discussed. In detail, the tools, methods of the project and how a CA in the virtual classroom plays the role of a stimulus in the presence of the teacher and its classmates, are discussed. Also, which processes does this CA go through in order to be able to correctly identify the behaviors of others and analyze them in accordance with the events in the classroom and to select the appropriate behavior and speech.

The Figure 3.1 demonstrates the project's schema that is developed for this research.

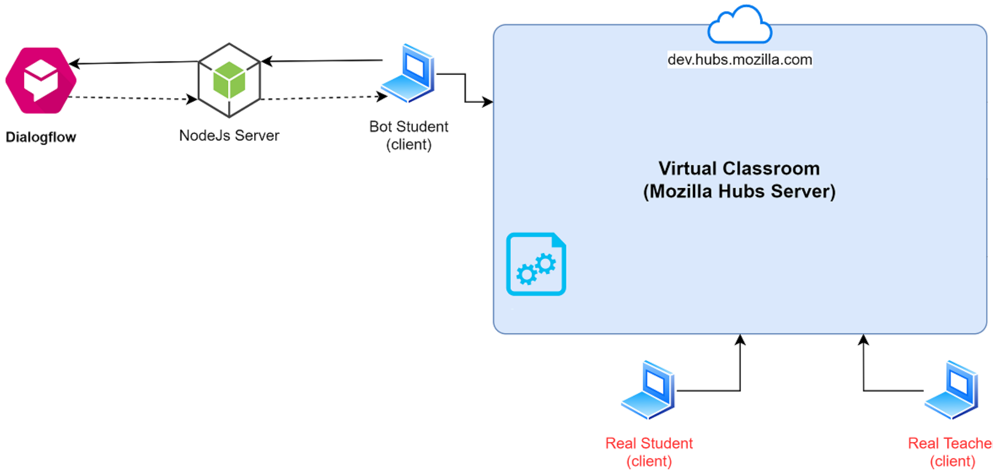


Figure 3.1. Project's Schema

3.2 The Topic: Solar System

Each classroom is centered around a topic in which the teacher tries to utilize the best methods or tools to teach students. Depending on the subject being taught, the teacher may use a variety of tools.

One of the interesting topics that has lots of potential for presentation in a virtual classroom is the topic of solar system. In the context of this subject, it deals with topics such as knowing space and planets, which, if properly presented using rich media such as images, video or animation, will enhance the appeal of this topic for the students.

The nature of this topic includes numbers, names and scientific facts that engage the listener's mind and can evoke questions. Therefore, this topic can be considered as one of the interesting scientific topics that is a good candidate for a virtual class. In the following, we will discuss the important reasons that led to the choice of this issue.

- This topic is familiar for everyone that has finished elementary schools and sure they have studied about it in science books during the school. Even if a person doesn't have any interest on it, learning or reading about our planet is not very new or strange thing.
- There are very good and verified sources for this topic. Like the websites of NASA[1], Wikipedia[2] and Encyclopedia Britannica[3] that provide rich content (text, image and video).
- There are good sources of forums that enthusiasts of astronomy ask questions and answer them. Some of popular questions related to the content provided in this class, has been derived from the most popular website Astronomy Stack Exchange[4]
- Agents and mainly the Machine Learning (ML) methods used in their systems, still needs many efforts of scientists to learn and understand like humans. They are very good at understanding the facts not abstracts. Like specific topics, names or numbers, which are discreet and each item has meaning without depending on previous or later words. Solar System, is a topic of this kind, which is a very good candidate to be learned by AI.

3.2.1 Course Content Sections

After doing a research on what are the features of a science class in reality, a script was prepared to be used by the teacher in the virtual classroom, which covers from the start till the end of the lesson (the so called first session of this topic).

This script is made of six parts. Dividing the script into different sections gives more power to the teacher to control the flow of teaching and organizing his/her thoughts during the lesson. It also helps the audience to keep up with the teaching flow.

1. Warm-up
   Teacher starts with greeting the students and knowing them by asking their names.
2. Introduction
   Teacher talks about what is this topic and how the class will be held during the semester and this first session.

[1] https://solarsystem.nasa.gov/

[2] https://en.wikipedia.org/wiki/Solar_System

[3] https://www.britannica.com/science/solar-system

[4] https://astronomy.stackexchange.com/

3. Solar System
4. Sun
5. Planets
   Mercury, Venus and Earth
6. Finishing
   Teacher says goodbye to students

During the sections 3, 4 and 5, two things can happen.

1. Teacher asks questions
   Teacher tries to stimulate students to think more about what they just learned in indirect way by asking about their opinion. This makes the class more interactive and keeps up the attention needed from the students.
2. Agent asks questions
   Sometimes the CA tries to interact with the teacher on what just he/she said. It can be a question or an engagement on the subject.

The second situation, gives two great opportunity to increase the level of interaction in the class. When the agent asks a question, this question covers a part that is not meant to be taught by the teacher. It leads to more explanation on the subject from the teacher and this repetition or further given details, gives the student to grasp more information on that subject.

Also this is a great point that teacher can ask indirectly the student to answer the question that the agent asked. This type of unplanned interaction that the teacher asks for help or support from another student, makes the student to speculate and increase his/her attention and interaction in the class, instead of just listening to the teacher's response. Additionally, student learns to be more cautions about aspects below.

- What was the subject?
- What was the agent's (classmate) or teacher's question?
- Using what he/she learned during the class or his/her background knowledge to speculate the answer

The goal is not to get the best or right answer from the real student, but increase his/her level of interaction in a way that he/she enjoys the learning process and be more interested to participate the conversations. This is how the true understanding is reached.

This division has also technical benefits. For the agent, it is crucial to distinguish the current section of the lesson presented by the teacher, so it can organize what to say or when to interrupt the teacher. Technically, the agent is sensitive to the teacher's words and based on what he/she says, with the help of NLP, agent gets the teacher's intention.

The agent is already trained in such a way to be sensitive to some patterns of sentences told by the teacher. These patterns are defined in the used NLP service, called Google DialogFlow[5].

[5] https://cloud.google.com/dialogflow

3.3 Challenges

Existing social VR platforms are designed to be used by real users. These platforms are served on servers that can be accessed on the Internet. The user has to use a browser to open the website and sign in into the platform with a username. After selecting the desired room or using the provided links, the user can join a room and have access to other available features provided by that platform.

For this project, it is needed to plug in a agent. This agent should visually look like any other participant in a virtual room. Also this agent should have access to hear what other participants in the virtual room are saying and in return, communicate with others via voice. After a research among available social VR platforms, no successful result was reached. In fact no solution is provided or designed for this need. All these platforms are designed real user-friendly and there is no endpoint via which lets human-machine communication.

One solution was designing a new platform that supports the needs of this project. This new platform should provide minimum requirements of a virtual space, like:

- a virtual 3D space of a science classroom hosting the participants;
- transferring audio or text messages among users;
- allowing sharing documents or videos in the classroom.

Developing such this system is time consuming and needs many efforts and it would never reach the level of tools and usability that current available VR platforms provide.

Finally it was decided to use an open-source platform that is available publicly. Available platforms, were prioritized based on their features and specially, how easy it is possible for end-user to access and use them. At the same time, they were technically analyzed to fit our needs and the other tools intended to be used in the project.

Mozilla Hubs[6] was selected. The main interesting features of this platform are listed below:

- It was already used for a virtual. exhibition[7] during COVID-19 and it was satisfactory.
- It is an open-source project founded and supported by Mozilla[8]. The source code of the Hubs Client and back-end server are available on Github[9].
- The platform is OS-free and only needs a browser to access the desired virtual room. No installation is needed and any room can be created with one click.
- It uses web standards (WebVR and WebXR) to deliver the contents and supports every single Mixed Reality headset. Users can enjoy the experience with advanced hardwares such as HTC Vive or Oculus Rift or even use desktop or mobile phones if the user can not access VR tools [24].

[6] https://hubs.mozilla.com/

[7] http://www.phd-dauin.polito.it/phddayforall2020.html

[8] https://mozilla.com

[9] https://github.com/mozilla/hubs

- It has a supportive community and developers that frequently publish new updates and compatibility with new technologies.
- Spoke[10] gives unlimited possibilities to design and create 3D social scenes for Hubs easily.
- It can host up to 24 participants in a room which can be planned for further researches and developments of this project.

3.4 Types of Interactions

In section 1.2 different types of interactions and influence of different players of a classroom were discussed. These interactions are also valid in a virtual classroom, specially when there exist a third player, the agent. The interaction among teacher and student, is obvious and can be fit in some specific models.

- Teacher-Student
  When a teacher asks a question or with any purpose, directly points to the student.
- Student-Teacher
  When students asks a question or responds to the teacher.

With the presence of the agent, as a virtual agent, the previous models are also valid and possibly new models can be observed. The agent breaks one-to-one interaction among the teacher and the student and this can be done, by interruptions or some other forms.

- The agent can interrupt the teacher while he/she is teaching by asking for more clarification on the course content or ask a question.
- The agent can respond to teacher's question (in case the student does not respond for some seconds) to fill the silence and the gap. This is meant to cover missing required interaction.

These interactions can be illustrated as shown in the Figure 3.2. The agent's goal is not to force the student to be active, but indirectly lead him/her to be more expressive and engaged.

The process of making decision between which type of interaction the agent should make, is described in details in the subsection 3.10.1.

3.5 Mozilla Hubs

Mozilla Hubs was released in 2018 and is in the list of recently emerged MUVRLE[11] [21]. It is designed to be accessible for everyone from anywhere with any available tools. Hubs is a light-weight platform that is compatible with many devices such as VR hardware (to let an immersive experience) and desktops (without any special hardware). It can be

[10] https://hubs.mozilla.com/spoke

[11] Multi-User Virtual Reality Learning Environments

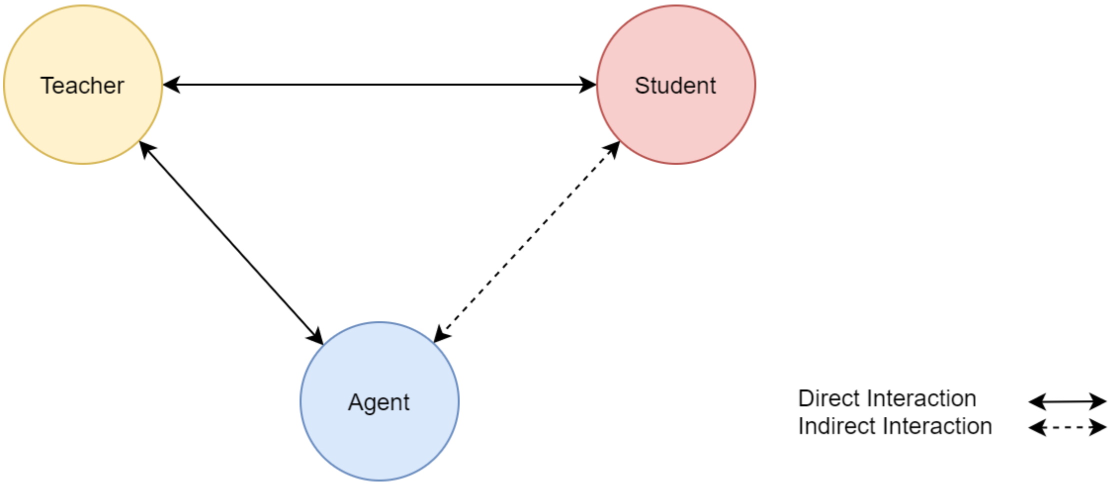


Figure 3.2. Types of interactions among teacher, student and agent

accessed via browser and doesn't need downloading or installing extra files or packages on running device.

Any user can open it and create a "room". Other users can participate in the room with using the shared link and enjoy being together in the virtual room. They can share content such as links, images, 2D or 3D objects from their local device or the internet. Having these shared objects in the room, let the users to interact with them and explore the objects freely from different aspects and share their ideas real-time.

Hubs also provides plenty of features for hosting virtual classes for small or big-size classes (maximum 24 participants in the current version). The 3D space of each room is called "scene". The scene can be easily changed by selecting from many already designed and free scenes from other users. Such as: classroom, halls, open spaces, conference rooms and many more. In Figure 3.3 some of them are shown.

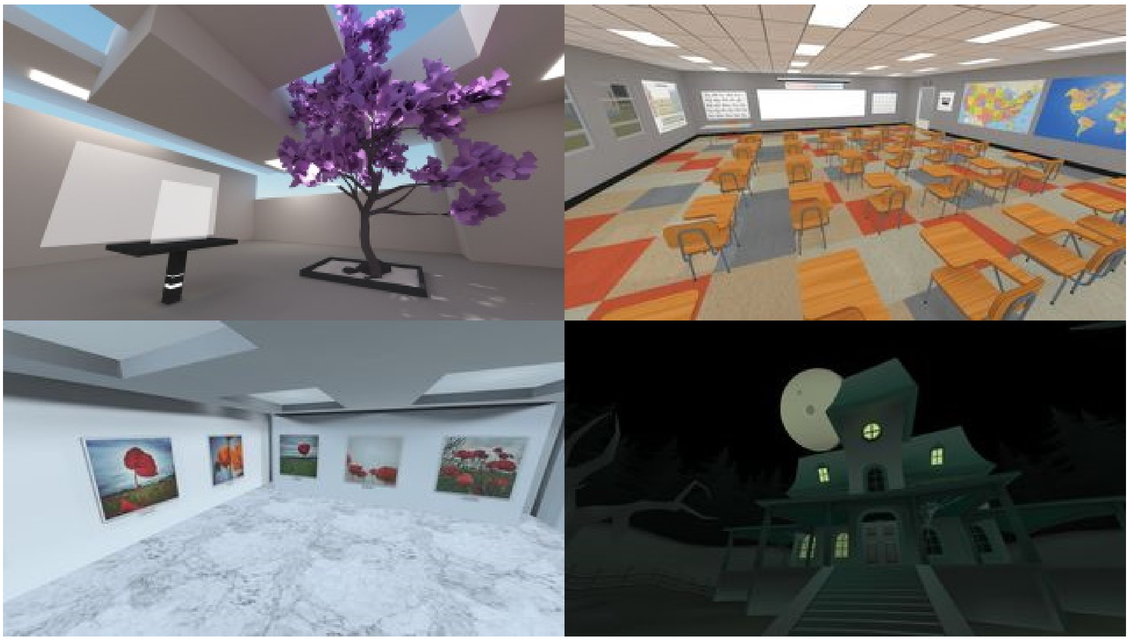


Figure 3.3. Different rooms (scenes) available

Hubs provides a set of internal tools to let the teacher to keep the level of teaching and interactive as high as possible and at the same time, the students have different ways of interaction and communication. Teacher can upload slides of the lecture on specified boards (on classroom's wall). Teacher can upload 3D objects from a rich source of free 3D objects (Figure 3.4) or even write with 3D pen. Not only students can hear the teacher, but they can view the streamed video of him/her. Students can show their emotions with some 3D emojis or type and send message to all participants in the class.

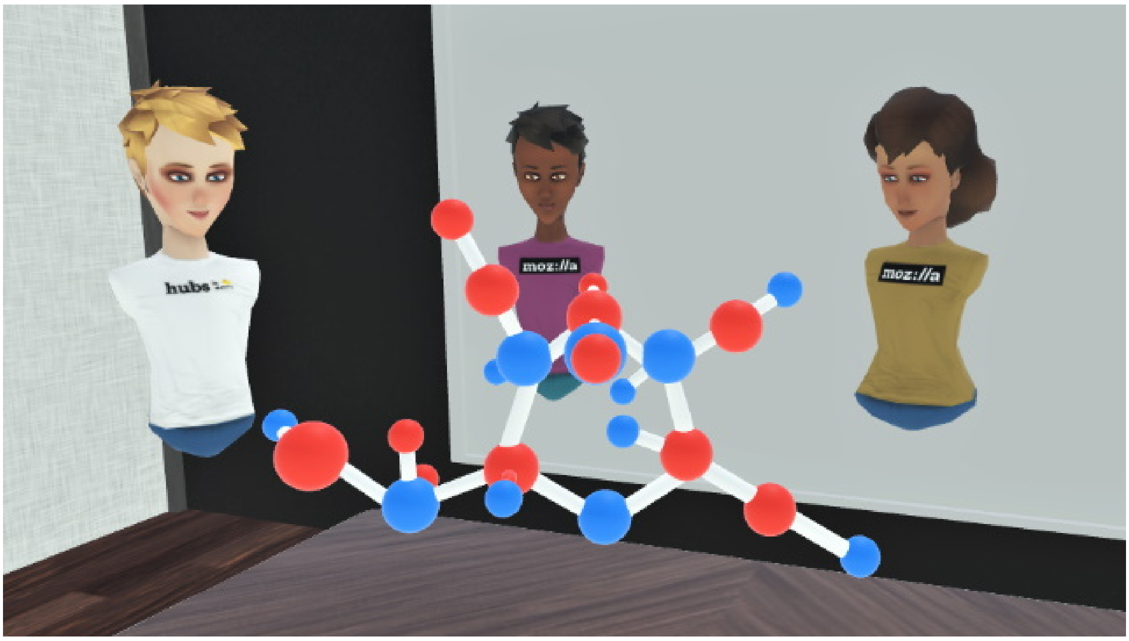


Figure 3.4. Avatars (users) in Hubs interacting with a 3D object

3.5.1 Virtual Classroom Design

The scene is designed for this project is a building of a school with three classrooms. When a user enters the Hubs room, he/she is inside the building and should walk forward to find the class number 3 and enters in it.

The virtual classroom is like a science class that is designed with auditorium style. It has three rows and seven columns of chairs dedicated for students. There are also some other scientific pictures on the wall to give more impression. Also there is a chair and table for the teacher beside the white board that is used for presenting lecture slides.

The agent, whose name is Aria, sits in the first row, so teacher and real student can see it easily (Figure 3.5). The real student is free to select a chair to sit, while is advised to pick one which is closer to the white board and can see it better. Figure 3.6 the whole classroom from a student's point of view, sitting at the top most seat.

For distinguishing between the different roles of the participants in the virtual classroom, two different styles are used for the teacher and the students (Figure 3.7). This helps the real student to find his/her classmates easily.

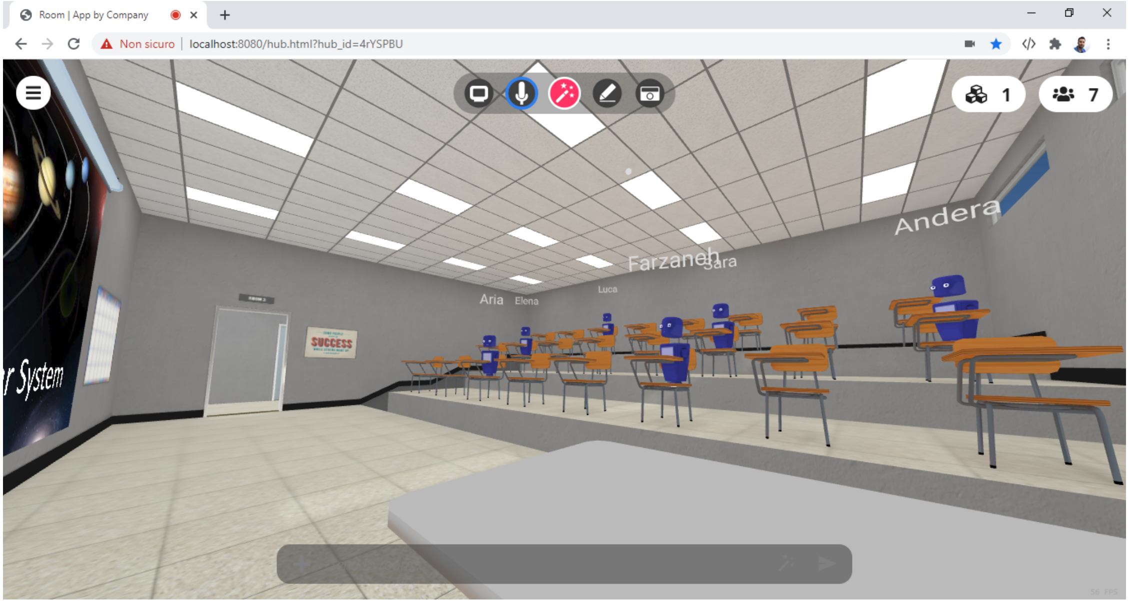


Figure 3.5. Virtual classroom from teacher's point of view

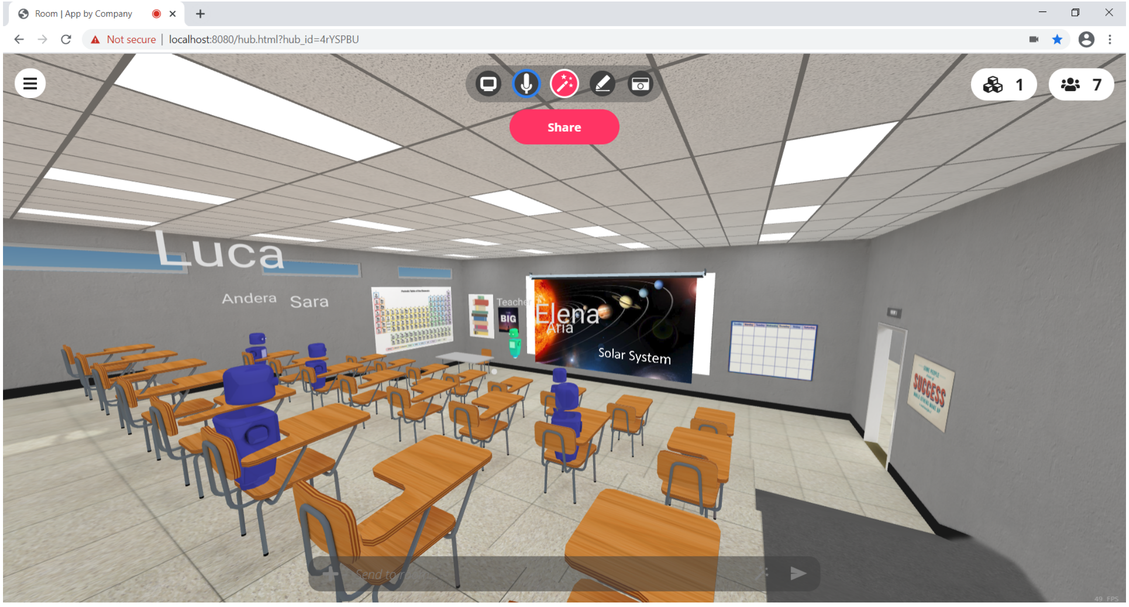


Figure 3.6. Virtual classroom from student's point of view

3.6 Turn-taking

We as human, during many years of social interactions, have learned how to inform the other pears of conversation, one's words are finished and it's their turn to talk. In human-human conversation, different signs are used for turn-taking or showing attention. This turn-taking is done naturally without indicating every time we want to speak. Techniques such as: talking fast or slowly, pausing, questioning, waiting for the response or in some cases directly noticing the other pears. The easiest way to change the turn is asking a question, that pushes others to respond (take their turn) [25].

Agent always waits for other participants to finish theirs words (turn). As soon as it

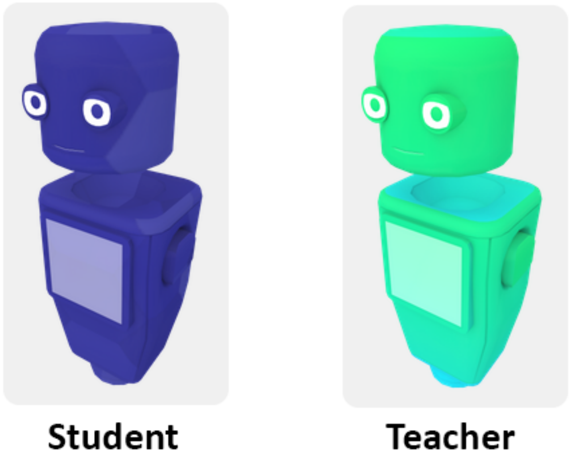


Figure 3.7. Two different avatar styles for teacher and students

happens, agent does a logical process and in case of possibility, it can seize the turn. As reported in a previous experiment [2] exploiting a configurable offset of 2 seconds, is a proper choice to let the participant to carry on talking again, as it happens normally during human-human conversation. Note that still speech recognition systems are not intelligent enough to recognize meaningless words such as "aha", "hmm" or even "intentional pauses" are used naturally for giving oneself a moment to think. It is a way of indication from the person that indirectly means "I am thinking what to say...". A simple and sure way of one's attention and existence. This simple sign is still a big problem for machines (computers) that need to learn more about subtle form of turn-taking.

3.7 Natural Language Processing

For us humans, conversation is natural. It's a part of our everyday life. We fundamentally understand it and all the nuances around it because honestly, it's part of who we are. This is why trying to teach a machine to have a conversation is so difficult. It's all about conversation experience.

> USER: OK, Google, what's the weather like tomorrow?
> GOOGLE: In Mountain View tomorrow, there will be scattered showers with
> a high of 59 and low of 46.

How we interact with machines right now seems pretty simple. We just ask something, and the machine responds. But it turns out that this is a really hard thing to get right for a machine since people ask for information in various different ways. For example, to do something as simple as get the weather you could say, "what's the forecast like today" or "what's the weather right now", or "what's the temperature like in San Francisco tomorrow?".

Now, if we were to code this traditionally, we'll most likely need a whole matrix of conditionals to figure out all the edge cases for all the ways the users can ask for this single information. And, of course, that's not maintainable. This is where NLU or Natural Language Understanding, comes in. NLU is a technology that helps translate human language into computer language and vice versa. It's very similar to Natural Language Processing, but it goes a step further to understand conversations that it hasn't been trained to understand, like errors, spelling mistakes, accents, sentiments, which makes NLU a great fit for agents. NLU is very similar to ML, but it's not the same thing. Rather, ML helps drive NLU.

Having a realistic and natural human-machine interaction is addressed in a sub field of computer science. NLP is the usage of arithmetic and computational methods and to learn, understand and generate a human language content [26]. This technique is designed to facilitate the process of conversational interactions among humans and agents (machines) which improves this experience and brings up new opportunities.

Chat-agents are the first users of this technique. NLP is the core of many new tools and technologies like Google Home, Alex, Siri which exploit chat-agents and users use their voice to exchange their ideas and needs. In recent years the usage of chat-agents in business is increased. More than 200000 and 10000 chat-agents have been developed for Facebook Messenger and Amazon Alexa, respectively [27]. These numbers are apart from the chat-agents that are developed for private businesses such as: online order takers, customer service and support, banks and many more field.

3.7.1 NLP Platforms

Previously it was necessary developers have good knowledge of math and ML for implementing their own algorithms. Meanwhile developing Graphical User Interfaces (GUI) and train the algorithm for the desired goals. All these actions and knowledge are not available for sole developers and even for big companies, developing such an algorithm are very costly. Because of the high level of demands from industries, different platforms were built. Such as: LUIS[12] (Microsoft), Wit.ai[13] (Facebook), Lex[14] (Amazon) and DialogFlow[15] (Google)

3.7.2 LUIS

Language Understanding (LUIS) is designed and developed by Microsoft. This cloud-based service exploits custom ML intelligence to provide conversational AI solutions to user's conversational interaction with the system. A user can transfer his/her command (to complete a task) or message to the service via any conversational application that has ability to communicate with LUIS. These clients include a vast range of applications, including social media applications, chat-agents or speech-enabled applications.

3.7.3 Lex

Amazon Lex, firstly offers an easy-to-use console using which it is possible to own chat bot in just few minutes. It also offers certain predefined bots in case the user is not familiar with Amazon Lex. Secondly it employs advanced deep learning functionalities so it just

[12] https://www.luis.ai/

[13] https://wit.ai/

[14] https://aws.amazon.com/lex/

[15] https://cloud.google.com/dialogflow

needs to be supplied with example phrases and Amazon Lex will train the bot to next level.

Additionally, it offers seamless deployment and scaling so the bot is always ready for the further applications. Lex allows to integrate with many other AWS services like Lambda or DynamoDB Amazon Poly. Amazon receives an input, either replies with a relevant message or it will complete the desired task for the user. It triggers a Lambda function which integrates with other services like DynamoDB Amazon Poly and many of the services and performs the necessary actions by providing a desired results to the user. There are certain typical steps that must be followed when a bot is created or bought on Amazon Lex.

3.7.4 Wit.ai

Wit.ai is the NLP tool purchased and developed by Facebook. This service also uses the power of ML to train agents and offers features like other competitors. The main advantage of this platform is its ease of use and the ability to add it to applications related to Facebook.

By providing a website and interface, this platform also provides an easy environment for users to easily meet their needs for the production and training of a dedicated agent without basic programming knowledge. Initially, the platform provided context capability for agents to learn, but later removed this important feature due to a lack of sufficient processing power.

One of the important features of this platform is the detection of entities. So that it can be considered a suitable candidate for the development of agent.

> For example, saying “I want to go from Porto to Lisboa” agent Porto and Lisboa are location entities but you can further distinguish between a fromLocation (Porto) and toLocation (Lisboa).

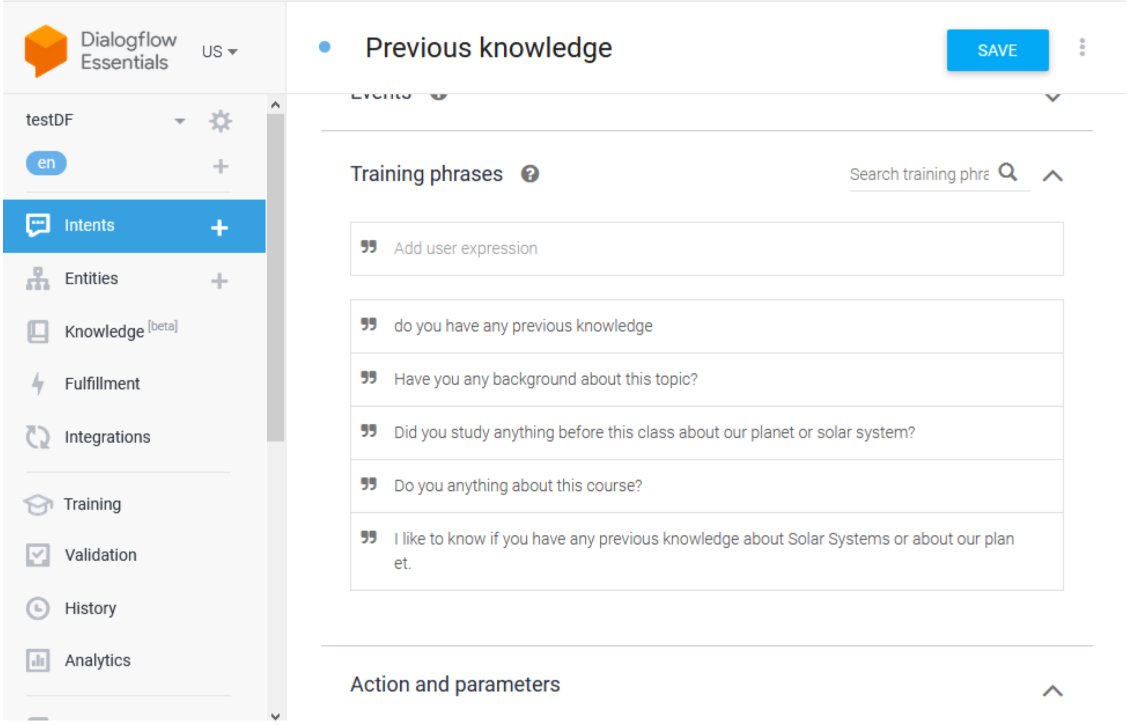
3.7.5 DialogFlow

DialogFlow is an end-to-end tool powered by NLU to facilitate rich and natural conversations. DialogFlow sits in the middle of the stack. A user can interface with it via all the common channels, including text, websites, apps, messengers, and smart voice devices like Google Home. DialogFlow handles the job of translating natural language into machine-readable data using ML model trained by the developer. Once it identifies what the user’s talking about, it can hand this data to the application (back-end) where it can use the results to make stuff happen. Application’s back-end can fulfill the request by integrating with other services, databases, or even third party tools.

DialogFlow presents solutions for building conversational interfaces on top of products and services providing a powerful NLU engine to process and understand what users are intending. With a complete website and all the features needed for a developer, DialogFlow is a complete and integrated platform that can greatly meet the needs of having a professional agent. Complete features include: creating a speech pattern (“intent”), creating entities, training, chat history, statistics and integration in other tools such as Telegram, Messenger, Google Home, etc.

One of the important and practical features of DialogFlow is the ability of context and life-span. Context allows the agent to access key information provided during the conversation and can use it to continue the conversation in order to understand better what the user is referring to. The Lifespan feature also specifies that information stored during a conversation (train) remains in agent's memory and the agent can refer to that information later in the conversation. Each intent can contain a set of follow-up intent that manages the conversation process more accurately. Intent follow-up allows the developer to control the user's different response to the agent.

For example, in case that the intent was "Order-Now-Confirm", the follow-up intents could be "Order-Now-Yes" and "Order-Now-No" with each one having different responses and associated actions.



do you have any previous knowledge

Have you any background about this topic?

Did you study anything before this class about our planet or solar system?

Do you anything about this course?

I like to know if you have any previous knowledge about Solar Systems or about our plan et.

Figure 3.8. DialogFlow interface

3.7.6 Why DialogFlow

Different platform are designed to support a subset of user's needs. There is not one single all-purpose platform that matches all the needs. The table 3.1 shows a the comparison of all these NLP platforms. The platforms are compared based on some common and fundamental features listed below.

- Developer Company
  The company that has designed and is in charge of its future development.
- Training Model
  The defined model allows the ML to predict the user's utterance. The developer can define his set of the input examples to train the algorithm for further predictions.
- Context
  The algorithm understands the keywords and keep them for further referencing. This feature is critical for natural and human-like conversations.

- Pre-defined Intents
  Has the provider prepared general intents for general use-cases?
- Pre-defined Entities
  The knowledge-base used for extracting general information from the input text.
- Custom Entities
  The ability to add new and custom entities to the knowledge-base for extracting custom information from the text.
- Composite Entities
  The feature to combine and composite different entities to have a more complex entity.
- Follow-up Intents
  Defining intents based on the previous matched intent
- Analysis and Diagnostic Tools
  The set of the tools that allow the developer to figure out and analyze how the algorithm has processed and matched the intents. Metrics on how the intents and entities are being resolved.
- Graphical User Interface
  Most of the users are novice and don't have professional or programming knowledge to use such these vast systems. A fluent and easy-to-understand is necessary.
- Conversation Flow
  The GUI that lets the user to observe the conversation and its flow between the user and the agent.
- API and Webhook
  The quality and the options that are available via programming interfaces or callbacks that can be invoked via webhooks.
- Language
  The number of languages the platform support to train the agent in different languages.
- Speech Recognition
  Possibility that let user to interact directly with voice, so the text or information is extracted from the auditory input. Mostly used in mobile devices.
- Cost
  The pricing defined by the provider.

Among the mentioned platforms, DialogFlow was selected for this study. Regarding the features it has and its price, DialogFlow is one of the most complete platforms with all the needed features without any cost. The powerful ML algorithm that is used in DialogFlow, let us develop and train the NLP service much faster and more accurately. The agent uses NLP for understanding the teacher, and it was needed to make the agent, as intelligent as possible. This gives the teacher the possibility to speak beyond the prepared script and use synonyms that have never been used for training the NLP service. During the lecture, the teacher is facing different types of challenges and distractions and it not possible to keep him sticked to the exact words in the script.

This feature, also provides the possibility to expand the current study to other scenarios where focuses on the teachers behavior and skill assessment in teaching, in future.

	Lex	DialogFlow	Wit.ai	LUIS
Developer Company	Amazon	Google	Facebook	Microsoft
Training Model	✓	✓	✓	✓
Context	✓	✓	✗	✗
Pre-defined Intents	✓	many templ.	basic	basic
Custom Entities	✗	✓	✓	✓
Composite Entities	✗	✗	✗	✓
Follow-up Intents	✗	✓	✗	✗
Analysis and Diagnostic Tools	✗	✓	✓	✓
Graphical User Interface	✓	✓	✓	✓
Conversation Flow	✗	✓	✓	✓
API and Webhook	✓	✓	✗	✓
Speech Recognition	✓	✓	✓	✓
Language	✗	15	50	10
Cost	trial	free	free	basic plan

Table 3.1. Comparison between Lex, DialogFlow, WIT.ai and LUIS

3.7.7 DialogFlow's Elements

DialogFlow uses ML as the core of its processings and set of other components as input or output to this core. Before being able to exploit the DialogFlow agent, it is needed to define some "intents" with which the system can match the input words of the user. After being trained with the intents, the agent is ready to be used in the designed activity domain. The DialogFlow's structure is not finished here and contains many more elements which makes it near perfect solution for a CA:

- **Invocation**
  The invocation kicks off the experience with the agent in a conversational manner. Just like saying "Hello, how are you?" to for example a friend.
- **Intent**
  Intents do the task of mapping the user's input to response. In each intent, we defined examples of the user's utterance, what to extract from that utterance and how to respond to that. For example, phrases like "I want pizza," "Get a pizza," or "Order pizza" all means they are indicating the same intent Pizza Ordering Intent.
- **Entity**
  Entities are DialogFlow's mechanism for identifying and extracting useful data from natural language inputs. While intents allow any agent to understand the motivation behind a particular user's input, entities are used to pick out specific pieces of information that the user have mentioned. For example, if a user's input phrase is "Please order a 12 pizza," dialogflow match "12" as Pizza Size entity.
- **Context**
  This is used to reference to the parameter values at the user moves between different intents throughout the conversation. Contexts represent the current state of a user's request and allow an agent to carry information from one intent to another. In our pizza example, if a user asks: "Order me a pizza", the agent needs to know more details about the order like: pizza size, toppings, extra sauce and few more

specifications. For collecting all these information on a multi-turn conversation and staying on the same phase we needed contexts.

- **Fulfillment**
  When a fulfillment request is added to an intention, DialogFlow will execute the rest of the intention within a programmed NodeJs environment. The programming language used within the NodeJs environment is Javascript. DialogFlow can make contact with for example Databases, Facebook and other API's by using Javascript.

- **Action and parameters**
  In this section usable parameters can be created that are used in for example the fulfillment section.

- **Response**
  The answer that will be given to the user.

3.8 How to Trigger the Agent

As mentioned earlier, NLP tools understand the meaning and intention of the user based on the interaction made by the user with them and can provide the appropriate answer based on the trainings already applied by the developer. All conversations can not be predicted due to the nature of a classroom. Although the content presented in the classroom is largely predetermined in the form of topics and questions, it is still not possible to predict the flow of conversation between teacher and student in advance. The main feature of the agent is that it can completely, automatically detect the subject discussed by the teacher. Whether the teacher is saying general phrases such as introducing himself / herself or teaching. Does the teacher address the agent or does he ask the class a question that the agent should respond to?

In similar researches, the process is in this form, the user sends his message to the agent to be analyzed by it and the agent gives the appropriate answer to the user when needed. But in this project, the main goal is to eliminate this process and automate it. Just as a real human-human conversation, the conversation takes place without the need for any additional tools or buttons. However, it should be considered that the technology used in this project is based solely on the text of user's speech, and natural conversation can never be conclusively implemented. Maybe this issue can be implemented in the future and later versions of this project.

According to the facilities provided by DialogFlow, principles for designing a proper and correct conversation have been introduced by professionals which have been used in the design of this project [28]. Conversation Design provides the principles of teaching computers to communicate more like humans and not the other way around. These processes prevent the users of having frustrating experience.

The main purpose of observing these principles is to increase the usability of NLP tools against user interactions. We can never predict what the user will want to say, so the NLP tool must be prepared in such a way that it can understand the user's true intent based on different patterns and provide the right answer.

By categorizing the sentences that may be said on the subject of the solar system in a classroom environment, the patterns required by the NLP tool can be defined. These patterns are formed based on a set of similar expressions that have the same meaning.

Phrases that are said in the classroom environment can be divided into different patterns, so that the NLP tool can understand the pattern of the phrase by comparing and examining the phrase said by the user and give the appropriate answer.

These patterns can be divided into two categories: static and dynamic. Static patterns refer to expressions that, regardless of the subject presented, before or after, always follow a fixed meaning and intent, and the speaker's purpose does not change in respect to the previous subject or utterance. For example, when the teacher says "hello, good morning" at the beginning of the class, he always has the same intention.

The second category refers to a group that depends on specific topics, keywords, or key phrases that have already been stated. This category is called dynamic certification, and the agent can not process the user's intentions without prior knowledge. This set of patterns mainly employ the entities and contextual features provided by the NLP platform and in this project have not been used.

Before explaining how to implement and transmit the teacher's words in the virtual classroom environment, the agent should be considered. It is necessary to categorize the static patterns that are designed for the classroom space and in accordance with the solar system lecture. As mentioned earlier in the introduction section, the classroom is divided into several sections. In each part of the class, depending on the type of talk the teacher is making, the agent should behave appropriately. For example, the agent always waits for the student to have ample opportunity to think and answer the teacher's questions, and then, if no action is taken by the student, the agent intervenes and asks the teacher for permission to speak. But this behavior is not always right. For example, agent does not need permission to say his name at the beginning of the class and can speak without raising his hand. Regardless of whether the student has answered or not.

The agent should understand what stage of the lecture it is or what the teacher is talking about. For this, it pays attention to the information sent to the server by the DialogFlow. When the text of the teacher's speech is sent to the DialogFlow, the DialogFlow analyzes it and matches it to the defined intents. If the match is made, a response will be sent to the server

The response received from the DialogFlow, as shown in Figure 3.9, contains important information that allows the agent's performance processing system to decide to intervene. They generally consist of two categories. Fallback and the correct answer. NLP is trained to notify or consider anything that is not related to the content of the lecture and to inform the server as a fallback. But as shown above, if the audio matches with the text in one of the patterns, the agent must behave appropriately. In the classroom environment, several types of behaviors are considered for the agent, each of them affects on the type of decision and subsequently its behavior.

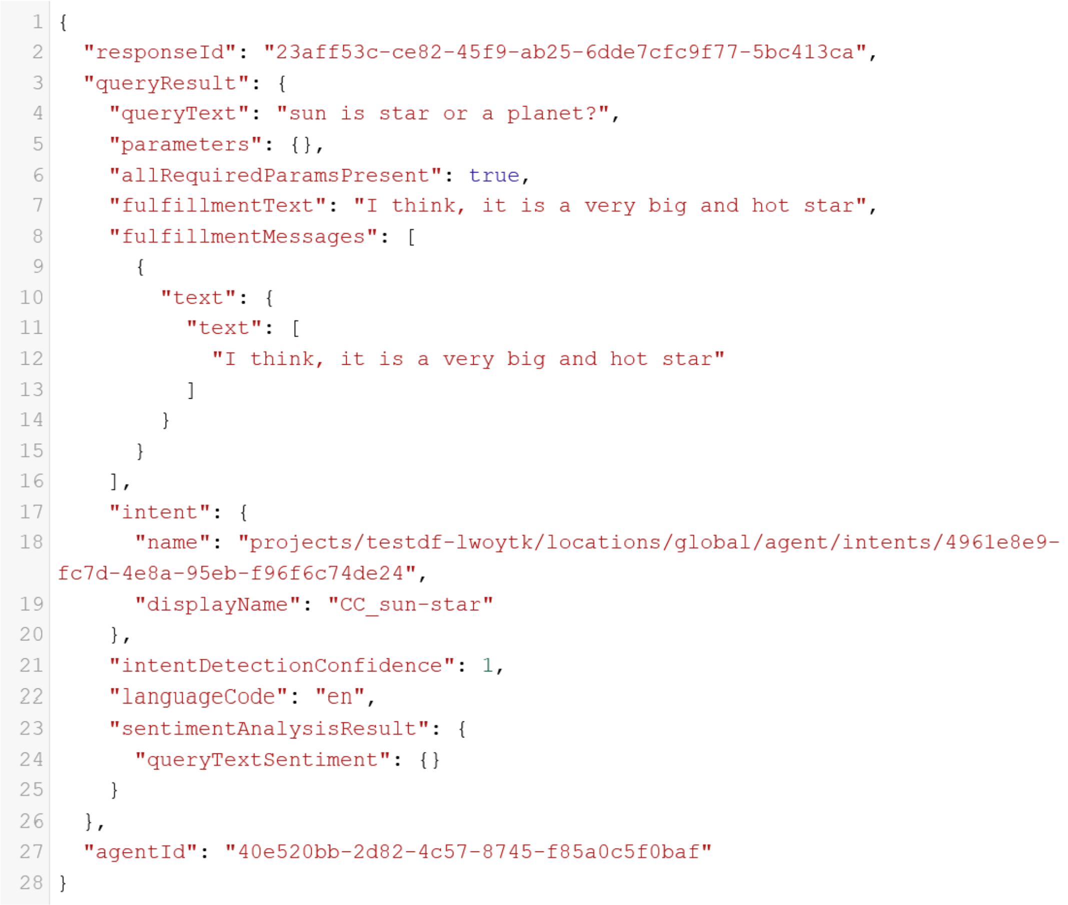
- **Repeat**

  What: It is possible, for any reason, teacher may need to ask the agent to repeat what it said. It can be for teacher's need to hear again or maybe a repeat for raising student's attention.

  Behavior: After receiving this data from server, the agent waits for 2 seconds and without asking for permission, repeats its last words.

  Interruption: Only teacher can interrupt her.

  Category: Trigger Word (TW).



```
{
  "responseId": "23aff53c-ce82-45f9-ab25-6dde7cfc9f77-5bc413ca",
  "queryResult": {
    "queryText": "sun is star or a planet?",
    "parameters": {},
    "allRequiredParamsPresent": true,
    "fulfillmentText": "I think, it is a very big and hot star",
    "fulfillmentMessages": [
      {
        "text": {
          "text": [
            "I think, it is a very big and hot star"
          ]
        }
      }
    ],
    "intent": {
      "name": "projects/testdf-lwoytk/locations/global/agent/intents/4961e8e9-fc7d-4e8a-95eb-f96f6c74de24",
      "displayName": "CC_sun-star"
    },
    "intentDetectionConfidence": 1,
    "languageCode": "en",
    "sentimentAnalysisResult": {
      "queryTextSentiment": {}
    }
  },
  "agentId": "40e520bb-2d82-4c57-8745-f85a0c5f0baf"
}
```

Figure 3.9. DialogFlow JSON response sample

- **Fallback**

  What: Agent is not able to understand what teacher meant or when it is trained (intentionally) not to recognize the meaning.

  Behavior: The agent does nothing and skips any process.

  Interruption: Not needed.

- **General One-Time Only**

  What: Some special cases happen or are designed to happen only once during the session. This category guarantees technically and covers possible misunderstandings by the NLP. Like: “What is your name...”

  Behavior: After receiving this data from server, the agent waits for 2 seconds and without asking for permission, narrates the response.

  Interruption: Only teacher can interrupt her.

  Category: General One (G1).

- **General Word**

  What: General words that can be understood by the agent. These are meant to be conversation facilitators that provide a more natural feeling and lets the agent to be more expressive.

Behavior: After receiving this data from server, the agent waits for 2 seconds and without asking for permission, narrates the response.

Interruption: Both teacher and student can interrupt it.

Category: General Word (GW).

- **Course Content**

  What: During lecture, as teacher talks and explains, the agent is always listening and analyzes if it can understand the meaning and intention of the teacher. If it happens and the conditions are appropriate, teh agent will interrupt.

  Behavior: After receiving this data from server, the agent waits for 3 seconds and raises its hand asking for permission.

  Interruption: Both teacher and student can interrupt it.

  Category: Course Content (CC).

- **Course Question**

  What: Teacher keeps asking question to keep the level of attention and engagement high. The questions asked are understood by the agent and it prepares the appropriate response.

  Behavior: After receiving this data from server, the agent waits for 7-10 seconds. If during this delay, the student did not respond, it will narrate its prepared answer, otherwise it won't talk at all.

  Interruption: Both teacher and student can interrupt it.

  Category: Course Question (CQ).

- **Answering Allowed**

  What: This special case is hidden from user's point of view and carries a trigger for the agent that makes it to respond to the teacher, only if the teacher allows it.

  Behavior: After receiving this data from server, the agent does not wait and immediately narrates its prepared response.

  Interruption: No one can interrupt it.

  Category: Answering Allowed (AA).

Table 3.2 shows the characteristic of each category. The meaning of three important variables are as following:

- CSI (Can Student Interrupt)
  if this boolean variable is true, by saying any words from the student, the agent will stop talking.

- CTI (Can Teacher Interrupt)
  if this boolean variable is true, by saying any words from the teacher, the agent will stop talking.

- MR (Must Respond)
  this boolean variable by-passes the previous variables effects.

Category	MR	CSI	CTI	Delay (s)	Type
Fallback	-	-	-	-	-
Repeat	✓	✗	✓	2	no-hand
Trigger	✓	✗	✓	0.1	no-hand
Answering Allowed	✓	✗	✗	0.1	no-hand
General One-Time Only	✓	✗	✓	0.5	no-hand
General Word	✓	✓	✓	2	no-hand
Course Content	✓	✗	✗	1	raise-hand
Course Question	✗	✓	✓	random (7-10)	dynamic

Table 3.2. Characteristic of each intent category

3.8.1 When to Talk

In the previous section, it is considered how to use the NLP tool to identify the content of the teacher's words so that the agent can choose the appropriate behavior. But this is not enough, and agent has to take other parameters into account to make the final decision on whether to intervene or not. Throughout the class and every second, the agent (server) performance section monitors the performance of the teacher, student, and himself. The server stores the complete list of information at all times:

- number of words;
- average of the number of the words;
- the last two speakers;
- last utterance of the teacher;
- Activity Rate (AR) of the agent;
- Engagement Rate (ER) of the student.

Regarding the main purpose of the presence of the agent which is to cover and compensate for the low activity of the student and in the next phase, to motivate and accompany the student for more activity, the agent always tries to give the student the opportunity to speak first. This behavior of agent adopts according to the student's performance so that there is an inverse relationship between them. If the user activity is high, the agent tries to intervene less, and if the student is not active in the class, the agent acts as a cover and, by talking, moves the classroom atmosphere towards interaction.

The agent needs to have information about the activities of the teacher, the student, and himself in order to be able to decide when and in what context is allowed to speak. This information is calculated on every second and the agent has access to it. But how they are calculated is very important. Using the audio-to-text conversion tool, the agent has access to the text of the phrases spoken by the teacher or student. It can check the number of words and letters used by them and use it as a benchmark for future calculations. In this way, the number of words said indicates the level of activity of the person in the class. Definitely the teacher as the person who talks the most in the class has the most number of words. This has been proven in experiments.

Given that the length of the classroom (in terms of duration) is not predictable, the amount of teacher's activity (number of words told) can be considered as a criteria for calculating the level of activity of other peers in the class. In this way, at each moment of time, the ratio of the number of words spoken by the student to the number of words spoken by the teacher, indicates the amount of student activity in the classroom. It is defined as follows:

$$ER = \frac{\text{number of words told by student}}{\text{number of words told by teacher}} * 100$$

Also, the ratio of the number of words spoken by the agent to the number of words spoken by the teacher shows the amount of agent activity:

$$AR = \frac{\text{number of words told by agent}}{\text{number of words told by teacher}} * 100$$

This decision-making process does not end here, and other factors also play role. By default, a constant value is set for the agent's activity rate. According to various experiments, this amount is considered equal to 20%. Also, this range can be reduced or increased by ±5%. This parameter generally specifies the range of agent's activity rate relative to the teacher activity rate. This way, the agent calculates the new value of his activity rate before saying anything (if the pattern detected by the NLP is allowed), and if the new value increases the current activity of the agent by talking above the allowable range, the agent is prevented from talking. These calculations are also performed for cases where the agent is less active to prevent the agent from decreasing.

3.9 Speech-to-Text

DialogFlow, accepts agent audio and text. In the first experiments the system was designed to send the recorded voice to the server (in specific audio format) and let the powerful Google Speech-to-Text[16] service, analyze the voice and the utterance.

Because of relying on audio transfer, one major problem was the high latency. When the teacher or student, finishes their talking, the audio is sent and the prepared response is received after about 4 to 5 seconds (tests based on university's Internet which is a high speed connection). This delay, is not acceptable at all for a simultaneous and synchronous conversation.

Because of the limitations of sending long audios to this service, another solution is designed in order to extract text from speech (on client side) and transfer raw texts to DialogFlow. This approach leads a to a much faster transmission of what the user said and in a controlled manner, any extra modifications can be applied on the raw text before any further analysis by DialogFlow. This result is accomplished by using a Web Speech API[17]. Speech recognition is done by receiving speech via microphone which is analyzed by a service against a list of vocabulary and grammar. This process is highly fast and

[16] https://cloud.google.com/speech-to-text

[17] https://developer.mozilla.org/en-us/docs/web/api/web_speech_api

synchronous that gives the possibility to recognize what is told by the users. As the user talks, this service tries to guess most matched phrases and when a word or a sentence is successfully recognized, the set of results is returned as a raw text string.

All the users and participants in this virtual classroom are non-native English speakers with good level of proficiency and oral skills. Speech recognition system works with very high level of accuracy on English language even for non-native speakers.

For receiving good results from speech recognition API, a good hardware (microphone) is also needed. It is an important factor that directly affect on the input data that speech recognition API processes. It is not possible to guarantee what quality of microphone the user uses during the experiment. To overcome this issue, we used a supportive technique is used on the NLP side. During the phase of training NLP, we also added the words meant to be said but is recognized wrongly by the speech recognition API. This let's the NLP service to work properly even if the given text is partially mal-recognized locally.

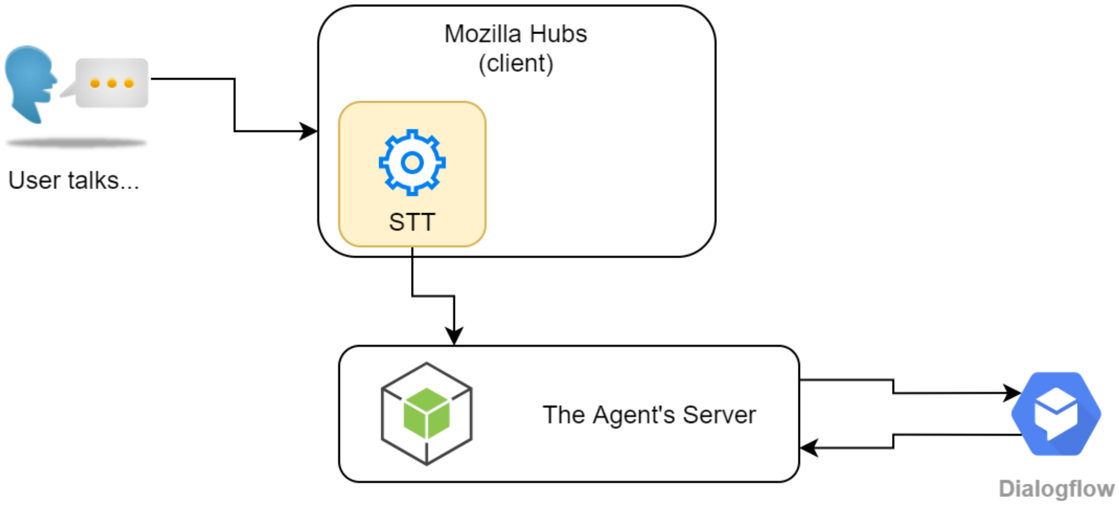


Figure 3.10. Speech-to-Text Mechanism

3.10 The Agent's Server

Separating the presentation layer (Mozilla Hubs) from the decision layer (server) has many advantages. This separation in the upper layers also leads to the separation in the lower layers (decision making logic) and implementation, and the more these two parts are not interdependent, the easier it is to expand and develop. The client's only task is to convert speech to text and text to speech, and it does so without any knowledge of how the server thinks and makes decisions. Also, the server operates without knowing the source of received text (from which client) or where the final answer will be used. In order to be able to make the right decision, the server performs many calculations that the client has no knowledge of, and also the client follows processes that the server has no knowledge of in order to properly manage the 3D space.

Agent needs to receive, process, and transmit information in order to interact effectively with other peers in the virtual classroom. As explained in the section 3.5, the Hubs client is capable of hosting 3D space and displaying in a 3D space. In this three-dimensional space, there are participants like the teacher and the student who talk throughout the class, and we expect the agent to respond verbally and non-verbally appropriately to the conversations exchanged between the teacher or the student.

On the other hand, the agent needs information from the beginning of the session to decide what action to take. Practically deciding at a particular point in time without prior knowledge, such as the level of activity of the teacher, the student, and himself, and whether the reaction he has already decided to act on, whether he has already done it, and so on. It is not possible for the agent. It is like asking a human being to participate in a conversation (even with two peers) without having memory and expecting the right conversation to take place. Here, the importance of the agent's memory and the information that it should store during class is going to be obvious.

Also, the agent needs to analyze the concepts and phrases expressed by the teacher or student, so it uses an external service called DialogFlow. The agent must transfer the text received from Hubs (client) to this service via REST-API and receive the answer. Due to the structure of the Hubs (client), it is not possible to implement this part in it and it must be designed and implemented in a separate backend system. This server is written in the language of NodeJS and by providing end-points allows direct communication between the client and the server. When the server receives the answer related to the analysis of the statements made by the teacher, re-analyzes them based on the categories defined for the agent. This will prevent a lot of unreasonable and unexpected behavior of the agent. Because the NLP service performs analysis regardless of which system it provides services to. It is the server's duty to match the received analysis to the class conditions and allow the agent to speak if the conditions are set and appropriate.

In addition, the server is designed to expand and add new features for further development. As shown in the Figure 3.11, the server acts as a bridge between the information received from the teacher and the information provided to the audience. The teacher speaks using the microphone like a regular conversation, and the words are converted into text that can be sent to the server by a speech-to-text tool. The server processes this text and returns the appropriate answer in text format to Hubs. Within the Hubs client, the text received from the server is converted to audio and expressed in agent's voice for other peers.

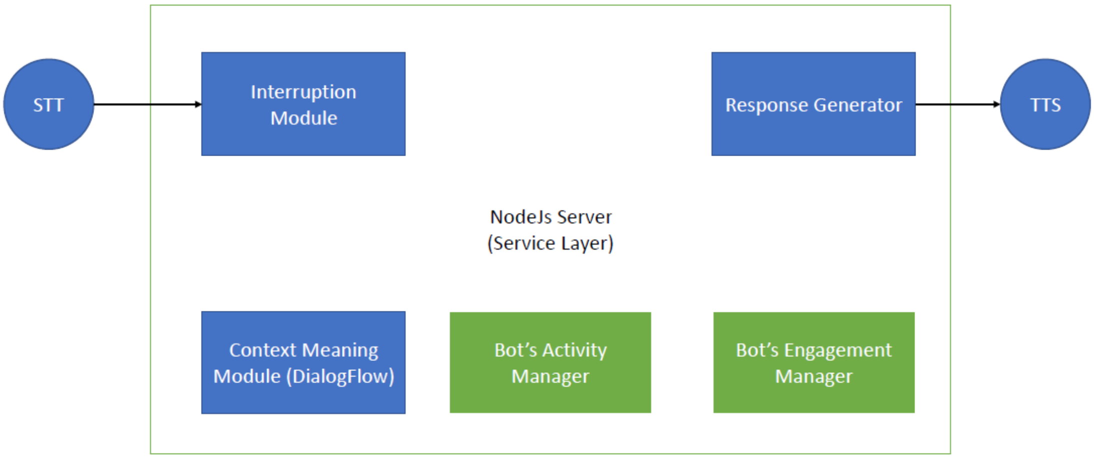


Figure 3.11. The Agent's Server

3.10.1 Decision Making Process

As mentioned earlier, from the start of the class, the server starts working and every second, a set of information is stored in the agent's memory. When the server receives a message from the client, it splits the message into two categories based on the speaker (the person who said it).

If the student has spoken, new statistical information is calculated.

- ER calculation that shows the student's activity in the classroom.
- Number of words spoken.

The main processing task takes place when the teacher has spoken. Because in this project, the agent is sensitive to the teacher's words and tries to analyze everything told by him, so that he does not miss any opportunity to interact. As soon as it receives the message, it sends it to the NLP service to find out the purpose of the teacher. First of all the agent must make sure that what the teacher has said and meant is within its knowledge domain so that he can choose the appropriate response from the set of responses and behaviors or reactions available.

If the statement made by the teacher is not understandable to the agent (the intent is recognized as "fallback"), the agent refuses to continue the calculation process and only updates the statistics related to the teacher's activity. If the teacher's words match one of the patterns already defined for the agent, the agent must perform other calculations to respond according to the conditions. In addition, the amount of teacher's activity and the number of words spoken increases, which has a direct role in changing other parameters.

After updating the parameters and changing the statistics related to the teacher, if the agent is within the allowed range of activity or certain conditions, it will send the desired reaction to the client to be applied by the agent.

3.10.2 Interaction Manager

Throughout the class, the teacher talks about a variety of topics that the agent tries to understand and respond to all of the teacher's words. In some cases, the agent can speak directly without the teacher's permission. But these conditions are not always applicable. For example, when a teacher asks a question, the agent has to wait for the student to think and answer. If the student does not respond within a certain period of time, the agent can notify the teacher that he wants to respond.

The process of preparing an answer, timing, obtaining permission, and speaking after permission by the teacher is a complex process in which the agent, the client and the server are involved. First of all, the agent must determine if the teacher has asked a question. If this happens, the agent will answer that question. Meanwhile, the teacher is waiting and we expect the student to answer the question. But that may not happen, and after a few seconds of silence, the agent has the opportunity to intervene and let the teacher and others know that he wants to answer the question.

The agent informs the teacher in three ways:

- sends text message in chat box;
- creates a hand shake sign with the hand emoji;

- makes sound.

In this way, the agent, the teacher and the student become aware of the agent's activity. The teacher, by saying, "Yes please, ..." informs the agent that he can say his word. This special phrase by the teacher activates a special pattern on the server that allows the agent to express what it has already prepared. The agent expresses his laready prepared answer in class and the process of teaching the lesson by the teacher continues.

The agent may need to ask a question for more clarification or a ask an out-of-box question which is not taught by the teacher. This is the opportunity designed in this project to let the teacher use it to redirect the agent's question to the student. In case the teacher feels that the student is not active or do not interact enough, this type of indirect questioning can stimulate the student to be more expressive and even think out of the box.

3.11 Text-to-Speech

The received (analyzed) response from DialogFlow should be uttered in a human-like voice and heard in the class. So anyone in the class, thinks that it is a real person (instead of an agent) is talking and it is one of the greatest steps in making everything more believable for other student in the class.

The received response from DialogFlow is analyzed in the agent's server to be decided if the agent has the permission to talk or not. If the agent has the permission, it finds the prepared response type from the server, which contains the message and a special instruction. On the clients (Hubs), the agent sends this response pack to all other participants (clients), then a local SpeechSynthesis[18] handles the process of converting text to speech.

As the Hubs is run in Chrome browser, we needed to use the appropriate tools supported by this browser to convert text to speech. SpeechSynthesis API is supported by default by Chrome (or any modern browser). So any client that receives the response from the agent, uses the provided synthesizer in the browser to narrate the response. The speech heard by other participates is without any noise or glitches, because the process of converting text to speech is done completely locally. This implementation lighten the load of converting text to speech and sending that to all clients, in a way that every client receives and plays it simultaneously.

It is another technical approach to speed up the agent's response and get closer to synchronous human-computer interaction. The benefits of this solution are:

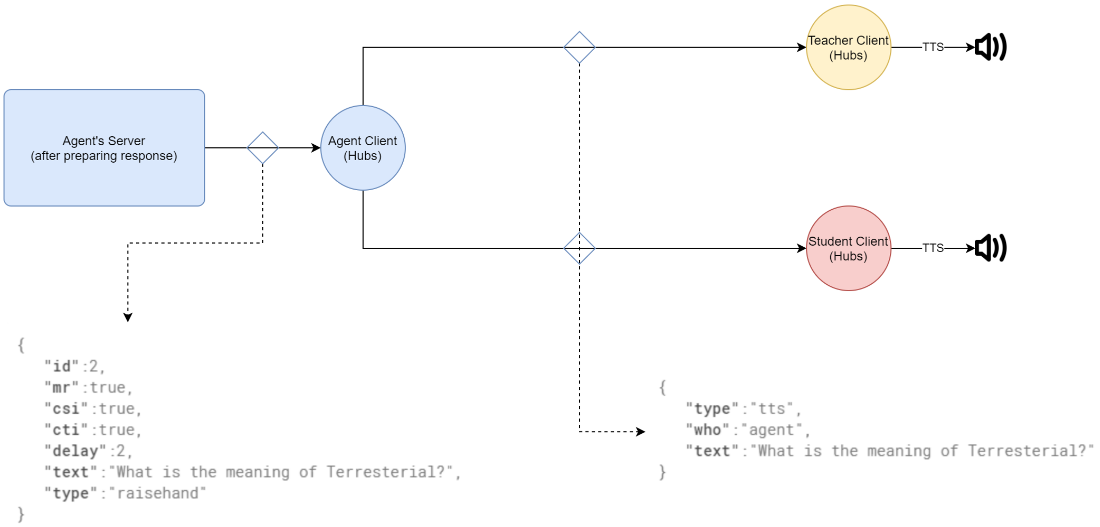
- **Light-weight Message Distribution**
  Instead of uploading or downloading heavy audio files, only text is sent to other participants.

- **High Speed and Simultaneous Message Distribution**
  This process is natively handled by Hubs, which has already prepared all the necessary tools and objects.

- **Length-free Message**
  As the response is distributed in text, there is no limitation on its length and converting it to voice.

[18] https://developer.mozilla.org/en-us/docs/web/api/SpeechSynthesis

- **Encodable/Decodable Message**
  Any further changes and modifications can be applied on the received response on each specific client. This is also a benefit to personalize the client's functionality for each participant.
- **Classified Recipients**
  For each client, specific rules or conditions can be set for further response filtering, so only a set of participants can receive the response.



```
{
   "id":2,
   "mr":true,
   "csi":true,
   "cti":true,
   "delay":2,
   "text":"What is the meaning of Terresterial?",
   "type":"raisehand"
}
```

```
{
   "type":"tts",
   "who":"agent",
   "text":"What is the meaning of Terresterial?"
}
```

Figure 3.12. Text-to-Speech Mechanism

Chapter 4

Evaluation and Results

4.1 Introduction

Previous researches [12] has reported positive experience of using VR headsets. Although setting up VR hardware can be cumbersome and needs specific equipment, the experience and the level of social presence is a motivating factor for not selecting desktop viewing method.

Basically, the project reported in this thesis, was designed to be experimented in VR using devices such as HTC®Vive^{TM}. Due to wide spread of COVID-19 and the limitations of gathering together in a closed space, it was not possible to run the experiments as planned originally, so the evaluation process changed. As mentioned before, Mozilla Hubs is designed purposefully to be used in any kind of devices and in this experiment, desktop viewing is chosen.

In this method, users open the Google Chrome[1] browser and the link provided to them to enter the virtual classroom. To prevent the problem of "shifting attention to other applications and activities" [12], we asked the users to activate "full screen" on their browser and close other unnecessary applications like messengers, audio or video player or other opened browsers or tabs. As we do not have any control on the user's computer or the experiment situation, this rules were set to minimize the level of distractions during the experiment.

As shown in Figure 4.1, the virtual classroom is hosted on Hubs server and the clients are connected to it via a special link address. The participants are:

- Real Teacher
  A real user that plays the role of the teacher in the virtual classroom. He/she has control over the lecture slide presented in the class and the agent is always listening to him/her.
- Real Student
  The real user in the experiment that is aimed to be active in the class.
- Agent
  The conversational agent that is connected to its own server and analyzes the events and the teacher's utterance.

[1] https://www.google.com/chrome/

- Fake Students 1,2,3,4
  They are dummy students to fill up the class and make the virtual classroom more crowded and natural. They do not interact neither with the teacher nor with each other.

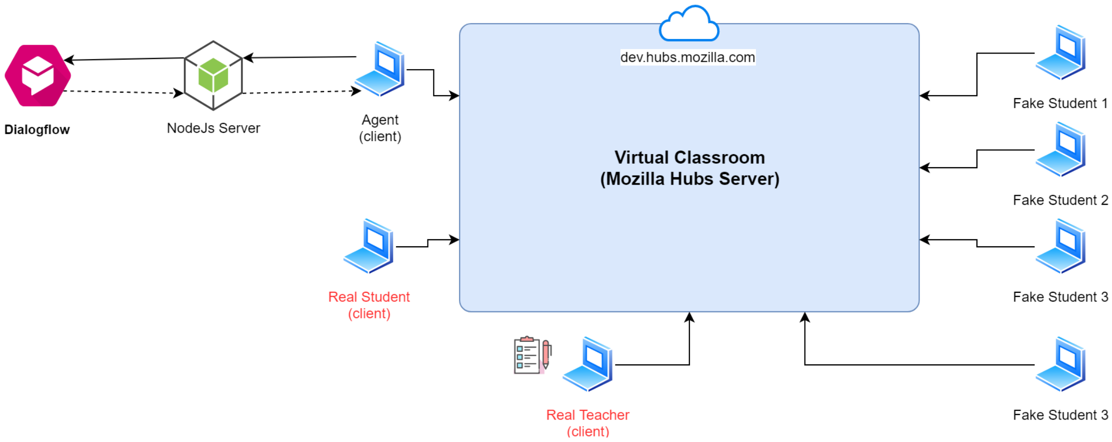


Figure 4.1. Evaluation Schema

4.2 Challenges

During the experiments we noticed the same problem mentioned by Yoshimura et al. [11]. The presence of a clock could give a better sense of control to the teacher. During the lecture, it is difficult for the teacher to calculate how long it is past since the start of the class.

The other problem that also has negative effect on the teacher's experience is lack of a second screen of the lecture slides. Teacher has to face the white board to swipe the lecture slides and loses control over the class and students' activities. Or the teacher has to repeatedly face the white board then the students during teaching or extra explanations.

In this project, the latter problem was solved to some extent. The agent is designed to ask during the lecture and it happened, most of the time, when the teacher was facing the whiteboard and the lecture slides. So when the agent raised its hand, it was not possible for the teacher to notice that. To fix this problem, extra communication methods were added. If the agent wants to ask something, it does three things:

1. creates a hand emoji, asking for permission;
2. writes in the chat box;
3. a signaling sound is played.

Chat box is always visible for every participant, as it is a fixed part of UI. So even if the teacher is not facing the students, he/she can understand who wrote the message and then, can respond the interaction. Additionally, the creation of the hand emoji and making the signaling noise, lessens the possibility og missing the agent's interaction.

4.3 Presentation

Before starting the tests, it was necessary to present the application, tools available and the environment to the user. This process was impacted by the remote access of the users to the system and made it difficult to have full control over the possible problems or mis-uses by the user.

At first, the purpose of this experiment was explained to the user, so he/she could get familiar with what he/she was facing. After that, the environment of the virtual classroom was presented and where to sit. As there are many available seats, the user was free to select one. It was suggested to them to pick a seat that was easier for them to have better observation over the class, the teacher and the whiteboard. Introducing the other classmates that were accompanying him/her was the next step. So the student understood that he/she is not alone in the class and an intelligent CA is there to bring new ideas.

One important thing was to make sure the student that the agent's presence is to help him/her, not a rival to be afraid of. There is no competition and the experiment is designed in a supportive way. It was an important point to let the students feel free before starting the test.

4.4 Data Gathering

For data gathering and future analysis, two methods were suggested in this project. The trivial solution was using a post-test questionnaire to get the users' feedback of system's usability and features. The other method which is much precise, was based of the intrinsic feature of the project. As every word said by each participant can be recorded, it gives a perfect opportunity to do analysis on them. On the other hand, the intent detection that is done by DialogFlow provides interesting data about the teaching path the teacher selected and how the agent understood him/her.

4.4.1 In-App Statistics

For system analysis and gathering the data of the participants' behavior, an internal statistical module was defined in the agent's server. This module was responsible of calculating, storing and providing the statistics at any moment of lecture, both to internal requests (to be used inside the server for making decisions) and to external requests (to be used for producing graphs and other further analysis).

This module provides two API end-points[2] dedicated for two different types of statistics.

- **General Statistics**
  This end-point provides general data mostly used internally. The agent's decision are based on the values that are stored in this part. The data is updated on each interaction made by teacher or student. Some of important elements are: last DialogFlow fulfillment text, current agent's AR, current student's ER, list of all responses given by the agent.

[2] An API endpoint is the point of entry in a communication channel. It is basically a fancy word for a URL of a server or service.

- **Statistics Per Second**
  This end-point is designed specifically for post-test analysis. Since the beginning of the lecture and at each second, precise values such as: the number of words told by each participant, the count of each detected intent and values of AR and ER are recorded.

4.5 Quality of the System

Before conducting the tests with real users, it was necessary to verify the system's quality and the intrinsic goals of the agent in this virtual class. The objective was to identify the agent's behavior during two different conditions and the provided data and results.

1. Active Student
   A situation that the real student is active and has a high rate of engagement during the lecture.

2. Inactive Student
   A situation that the real student is shy and does not tend to interact or answer to given questions.

To keep the conditions equal for both tests, a fixed script and method of teaching was used. The order of the content and topics are set in order to intrigue the student to talk. There are some topics that the agent will react to them and tries to ask some clarifying questions and also, some direct questions are asked by the teacher.

If the student answers those questions, the agent will not answer, but if the student does not provide any response in a period of 7 to 10 seconds, the agent will answer to the direct question. This approach is designed to keep the level of overall interaction of the agent in a steady condition, not keeping the agent silenced forever during the lecture.

The agent's logic tries to adapt its behavior with the behavior of the real student, so whenever the student is not active, the agent covers it and inversely.

4.5.1 Experiments

Two tests were done for simulating two different types of students. The first test simulates an active students that answers the questions of the teacher. The student does not necessarily asks new or out-of-box questions, just responds to all the questions of the teacher. In other words, there is no space left for the agent to answer to the teacher's questions.

It is notable that the agent is always listening to the teacher and if it finds the appropriate point of the topic (as trained before in the DialogFlow), it will raise its hand and asks the question. This behavior is regardless of the real student's characteristic.

The second test, simulating an inactive student, shows how the agent covers the silences and unanswered questions of the teacher.

Both tests were conducted by the same teacher (real person) and the same student (real student). The lengths of the tests were 727 seconds (≈12 minutes) and 812 seconds (≈13 minutes), respectively.

4.5.2 Conversational Activity

As it was mentioned before, any conversation is recorded and the data can be analyzed. By calculating the words told by any peer in the class, it can be understood how active each peer was. Always the teacher has the highest number of words, as he/she teaches the course, asks questions and responds to the students' questions.

Figure 4.2 shows the two calculated statistics of how many words each one has told during the lecture.

One important thing that can be interpreted is the total words told by the teacher. Although the second test was one minute longer, the teacher told almost 100 words less. This shows that the student's activity affects also the teacher's activity.

When the student is active, the agent is not forced to talk, because the main goal is to give space to the student to be active. Also, the figure shows at the beginning the agent talks and the number of the words are almost near. As the lecture continues, the agent is less active. At the end, the student has talked almost twice the agent.

On the other side, when an inactive student is in the class and does not talk, the agent is more active, by raising her hands or answering the teacher's question.

This figure also shows that the lecture is distributed almost equally for each topic and the teacher explains more when faces an active student. It is obvious, as the student responds, the teacher sometimes needs to verify (if answer is correct) or reject and then clarify (if the answer is wrong).

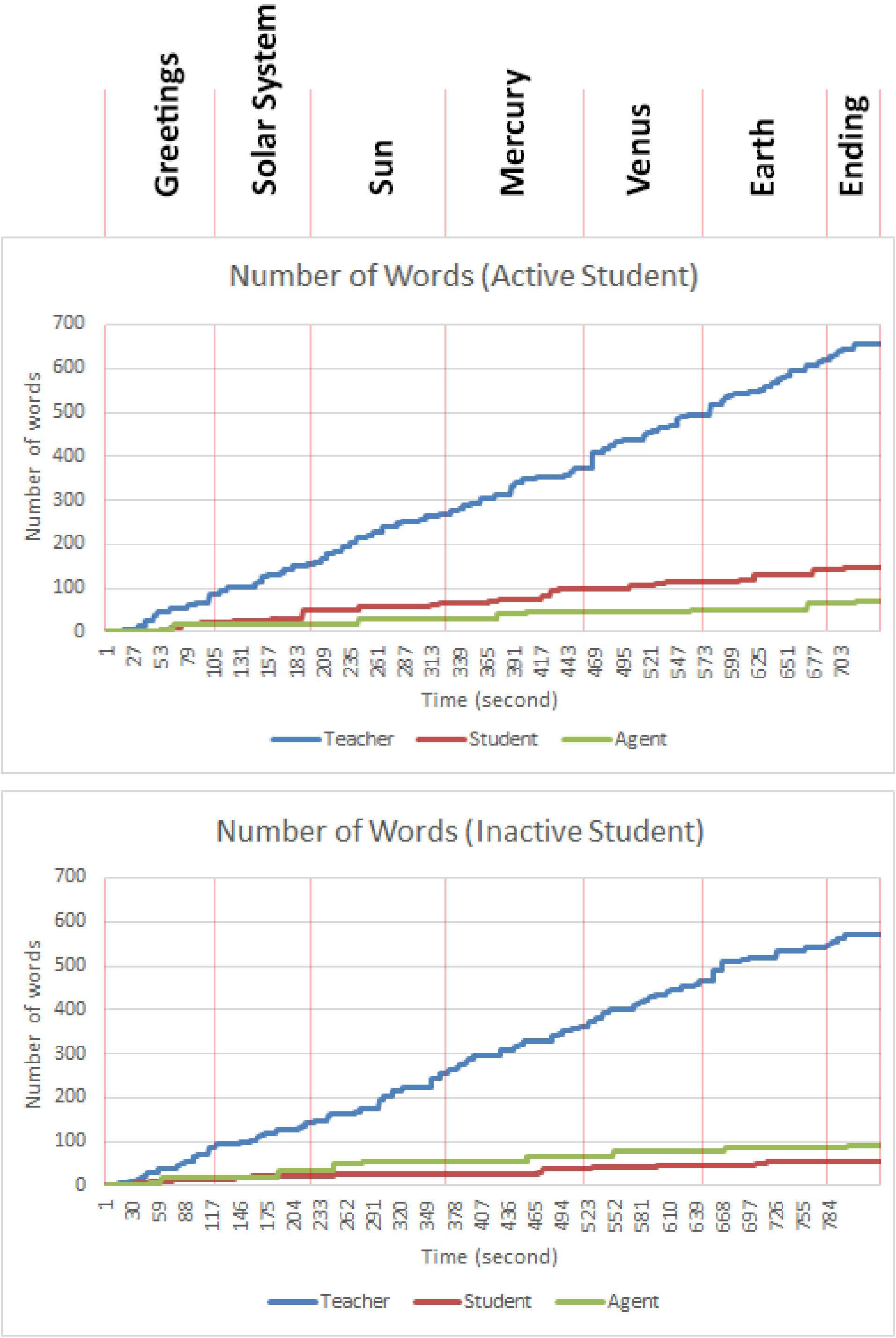


Figure 4.2. Statistics of conversational activity

4.5.3 Activity Rate and Engagement Rate

The parameters agent's Activity Rate and student's Engagement Rate were introduces in the section 3.8. These parameters demonstrate the agent's and student's behavior regarding the teacher's behavior.

The only metric available during the lecture is the number of words told by the teacher. The number of words told by either the student or the agent, shows how active they are. The goal is to keep the student stimulated to be expressive during the lecture which according to the Figure 4.3, it means to keep the red line high as much as possible.

Always at the beginning of the test, both ER and AR are very high, because the number of words told by teacher is almost equal to the student's and the agent's. But as the teacher speaks more and starts the lecture (after the Greeting section), these values are decreased significantly.

The calculations are based on the number of words told, and if a person stops talking, the line will be falling till it reaches near zero. The figure shows two different characteristics of the students clearly. The ER shows how active was the student during the test and the pulses are the sign of a long response from the student or the agent or can be a question asked by the agent.

In the script, in the topic related to Mercury, a question is designed to be asked by the agent and be redirected to the student. This is done intentionally to raise the student's attention and engagement. In the first test, it is asked between the time period 417-443, which raises ER by 5%. This happens in the second test between the time period 449-477 which is near 5%. In both cases, this leads to a higher ER value for the rest of the test.

This type of indirect question is asked only once, because the goal is to stimulate the student indirectly, not by forcing him/her to answer.

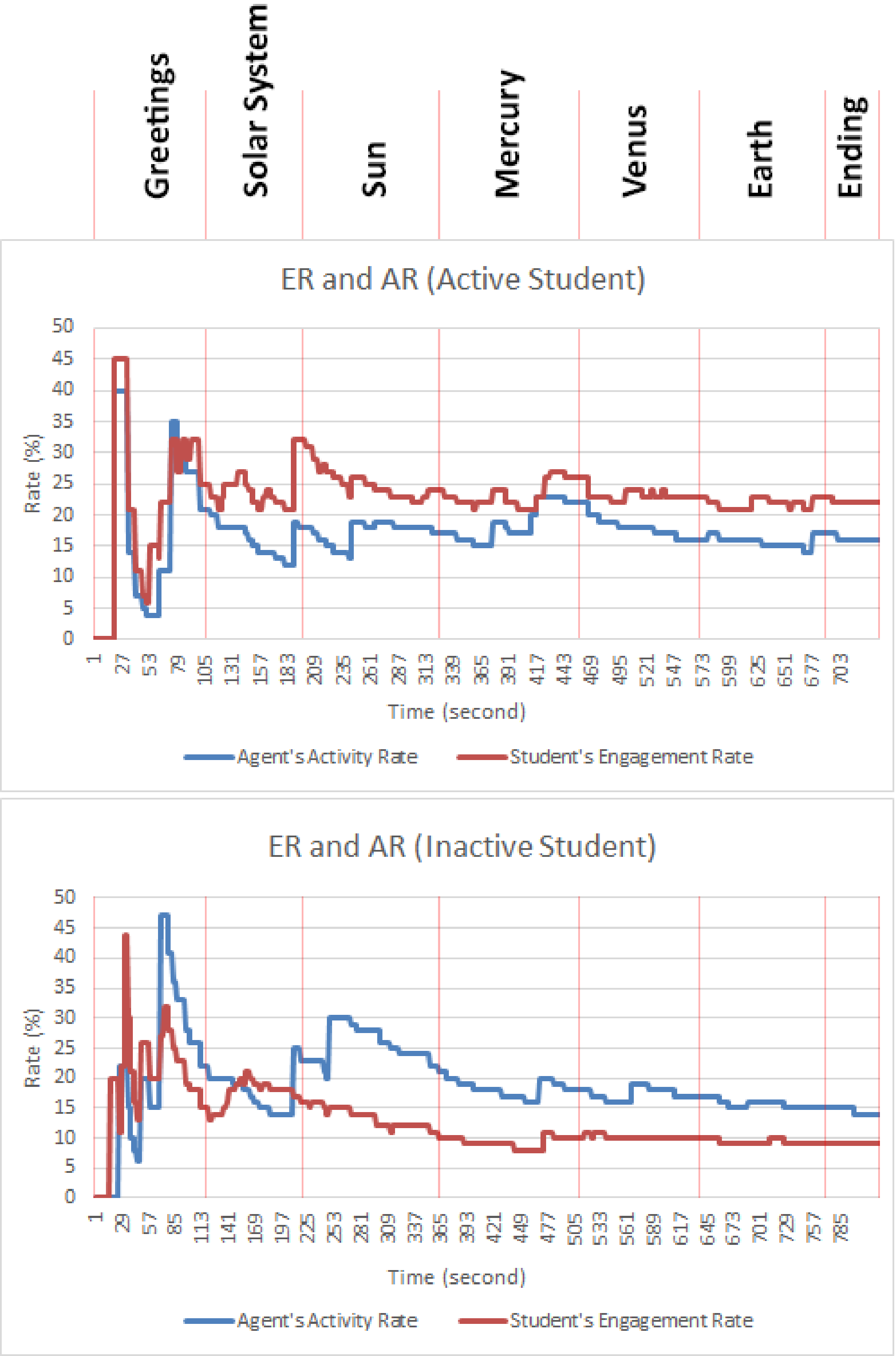


Figure 4.3. Agent's Activity Rate and student's Engagement Rate

4.5.4 Intent Detection

It is important to understand how the agent makes decision during the lecture. This figure shows the intentions detected by the agent by analyzing the teacher's words. The agent needs these analysis to understand what is happening during the lecture and based on the subject that the teacher is presenting, decides to raise its hand or is allowed to ask a question.

The category General One (G1) has always the same behavior in all the tests. It shows the four specific events in the class, the Greetings and the Ending.

Trigger Word (TW) is dependent on the interactions happen during the lecture. For example in the first test, teacher asked 4 times the agent to repeat its answer.

Course Question (CQ) points to the situations that the teacher asks a question directly. In the lecture 7 direct questions are defined, which were answered completely by the agent in the second test. As it is shown, it is detected 7 times. And in the first test, it was detected 6 times.

One important thing about this figure is that, it does not necessarily means every detected intent, is followed by an action from the agent. Like in the first test, the agent detected CQ, 6 times. Which means it found out that it has to answer the question, but whether it has the chance to answer or not is conditional. For example in the Figure 4.4, during the period 236-261, it finds 3 questions to answers, but according to Figure 4.2, it has told a few words, which means it did not respond 3 times. This is because, the student has answered and it was not necessary that also the agent responds to that question.

This is the basic characteristic of the agent, that responds to the teacher's question, if only the student does not reply. The agent is always ready to answer (detects successfully), but may not tell her answer.

Also, the second test show the normal behavior of the agent that it's behavior is not affected by the inactive student.

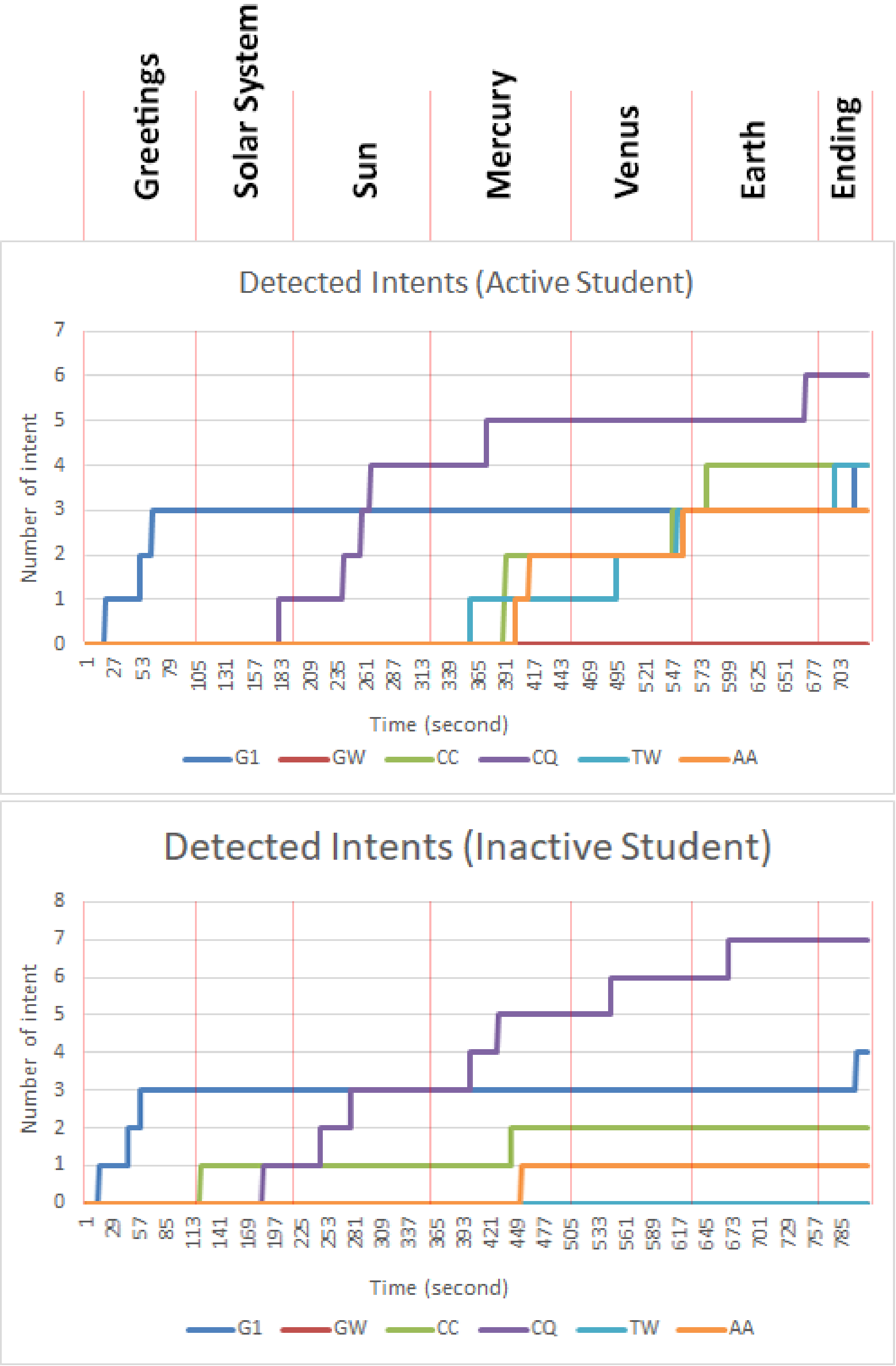


Figure 4.4. Intent detection by the agent's logic

Chapter 5

Conclusion and Future Work

5.1 Conclusion

This thesis work has led to a virtual classroom with the feature of being a synchronous on-line lecture environment. In this tool, the performance of the a real student can be analyzed while there is a CA as an intelligent classmate. The intelligent classmate that is in the class that interacts with the teacher and shifts the boring class to a stimulating environment.

The analysis of the data obtained through the two tests allowed to get interesting results about user behavior and how effective can be learning with an intelligent agent as classmate. The existence of a CA in the classroom, lowers the sense of isolation of the other learner. Also, learning with another classmate that asks out-of-box question gives confidence to other peer-learners to be active and be expressive, not to be afraid of answering or talking, which is the first step of active learning.

The analysis indicates that the agent can adapt itself with the behavior of the real student. If the student is active enough, it tends to be less active. The agents behavior like asking for permission before asking a question, has been an interesting point fo the student to learn, how he could do that. It means that real students pay attention what other classmates are doing and try to learn from them.

5.2 Future Works

Tests have been performed in a specific domain, Solar System, which has lots of possibilities while taught in virtual environments. Apart from the future researches that can be done on this topic, specially using VR hardware, more researches can be done on the role of the intelligent and CAs in virtual classrooms.

- Using multiple conversational agents as classmates that can share the role. Instead of having only one agent, multiple agents can have different verbal and non-verbal characteristics which make that classroom filled with more realistic agents.
- Intra-agents interaction that leads to more interactive classroom. While the agents can discuss or share ideas on a subject, with each other.
- Group learning with intelligent agents. Solving problems in a group of agents that focuses on social interactions and knowledge sharing.

Appendix A

Script Sample

For the lecture, a script was prepared to be used by the teacher for all the tests. This script covers all the phases of the lecture, from the Greetings to the Ending and the teacher needs to follow it.

This script is both a guideline (what to say) and the text that the agent is trained to understand it. To some extent, anything that is said out of this script, possibly is not recognized by the agent. This has two features:

1. teacher and student can talk about something that does not needs the agent's interaction. In another word, the agent listens, but does not understand the meaning of the conversation.
2. teacher should be stick to the script. This guarantees that all students learn the same content and the test condition is equal for every user.

In the following, a brief sample of the script is provided to cover all different conditions that the agent is trained for.

- General One: is told only once during the lecture
  — Teacher: "Hello everyone Good morning / afternoon"
  — Student and agent answer like: "Hello ..."
- Fallback: many of the things the teacher says, is in this category, which means that the agent does not understand. In fact the agent is not trained for these words. Because they are not related to the course content.
  — Teacher: "I am Javad, your science teacher and this semester we will have this class together. I hope we enjoy this course and learn together about our amazing solar system."
  — Student and agent say nothing.
- Course Content: the sentences that the agent is trained to understand and finds out that the teacher is teaching about a specific subject.
  — Teacher: "Turning around the sun are eight planets. The planets are divided into two categories, based on their composition, Terrestrial and Jovian."
  — Student may say something.
  — Agent raises hand to ask for permission. If the teacher allows, the agent will say: "We have only these two categories, Terrestrial and Jovian?"

- Course Question: teacher asks different questions after teaching each subject to keep or raise the attention.
  — Teacher: “What is the difference between Terrestrial and Jovian planets?”.
  — Student may answer or not.
  — Agent waits for some seconds and if the student does not answer, it will answer.

- Trigger Word: special words that are normally used in conversations, like asking for repeating the words again or prompting if everyone in the class is hearing the teacher.
  — Teacher: “Do you hear me everyone”.
  — Student and agent respond like: “Yes, I hear you well...”.

- Answering Allowed: special words dedicated to allow the agent to express its words after raising its hand.
  — Teacher: “Yes Aria, do you have any question?”
  — Student says nothing.
  — Agent will ask its question related to the subject just taught by the teacher.

Appendix B

Questionnaire

B.1 Before Experience Questionnaire (BEQ)

Before and after tests with real students, a questionnaire is answered by them.

B.1.1 Personal Data

1. ID
2. Training kind
3. Age
4. Gender

B.1.2 Background Knowledge

How familiar are you with these subjects? (Not at all familiar, slightly familiar, somewhat familiar, moderately familiar, extremely familiar)

1. How often do you play video games?
2. How often do you use video call conferencing system (e.g. skype, zoom, jitsi)
3. How often do you use/interact with VUI systems (Alexa, Google home, Siri)
4. How often do you teach to other people?
5. How often have you used immersive virtual reality tools? (HTC-Vive, Oculus Rift etc...)

B.1.3 Learning Habits

1. English Level (reading)
2. English Level (listening)
3. I take note during class

4. I put my notes in my own words to understand better
5. I put my notes copying as much as possible the teacher explaining during class
6. When I am learning a principle or definition, I try to think of at least two examples of how it might be applied or used (if not provided)
7. When I am learning material, I ask myself questions and study until I can give the answer, on two or three separate occasions, without looking at the text or my notes

B.1.4 Solutions of Learning Problems

When you don't understand what you are learning in class, you usually:

1. ask to teacher during class
2. ask to teacher after class
3. ask to a colleague
4. ask "Google" (Research online)
5. ask provided reference book

B.1.5 Learning expectancy

1. I think I know (right now) many information about the "Solar System"
2. I think I can successfully pass a test on the topic of "Solar System" without further training

B.2 Post Experience Questionnaire (PEQ)

B.2.1 Emotional Engagement

1. I enjoyed the class
2. The teaching method practiced by the instructor is enjoyable
3. I like it when the instructor asks me questions

B.2.2 Behavioral Engagement

1. I listen carefully to everything that is said in class
2. I ask questions about what I do not know
3. I interact with my peers during class
4. I always participate in discussions with my teacher

B.2.3 Cognitive Engagement

1. I always ask the instructor about difficult content

B.2.4 Co-Presence

1. In the synthetically generated world, I had a sense of presence, that is, of "being there" during the experience (Overall)

2. I noticed my partner

3. My partner noticed me

4. My partner's presence was obvious to me

5. My presence was obvious to (my partner

6. My partner caught my attention

7. I caught my partner's attention

B.2.5 Attentional Allocation

1. I was easily distracted from my partner when other things were going on

2. My partner was easily distracted from me when other things were going on

3. I remained focused on my partner throughout our interaction

4. My partner remained focused on me throughout our interaction

5. My partner did not receive my full attention

6. I did not receive my partner's full attention

B.2.6 Perceived Message Understanding

1. My thoughts were clear to my partner

2. My partner's thoughts were clear to me

3. It was easy to understand my partner

4. My partner found it easy to understand me

5. Understanding my partner was difficult

6. My partner had difficulty understanding me

B.2.7 Perceived Affective Understanding

1. I could tell how my partner felt
2. My partner could tell how I felt
3. My partner's emotions were not clear to me
4. My emotions were not clear to my partner
5. I could describe my partner's feelings accurately
6. My partner could describe my feelings accurately

B.2.8 Perceived Emotional Interdependence

1. I was sometimes influenced by my partner's moods
2. My partner was sometimes influenced by my moods
3. My partner's feelings influenced the mood of our interaction
4. My feelings influenced the mood of our interaction
5. My partner's attitudes influenced how I felt
6. My attitudes influenced how my partner felt

B.2.9 Perceived Behavioral Interdependence

1. My behavior was often in direct response to my partner's behavior
2. The behavior of my partner was often in direct response to my behavior
3. I reciprocated my partner's actions
4. My partner reciprocated my actions
5. My partner's behavior was closely tied to my behavior
6. My behavior was closely tied to my partner's behavior

B.2.10 Anthropomorphism

Rate on a scale from 1 to 5 (Leftmost/Rightmost, 3 neutral)

1. Machinelike-Humanlike
2. Artificial–Lifelike
3. Fake–Natural
4. Unconscious–Conscious
5. Moving rigidly–Moving elegantly

B.2.11 Animacy

Rate on a scale from 1 to 5 (Leftmost/Rightmost, 3 neutral)

1. Artificial–Lifelike
2. Dead–Alive
3. Stagnant–Lively
4. Apathetic–Responsive
5. Mechanical-organic
6. Inert–Interactive

B.2.12 Likeability

Rate on a scale from 1 to 5 (Leftmost/Rightmost, 3 neutral)

1. Awful–Nice
2. Unpleasant–Pleasant
3. Dislike–Like
4. Unfriendly–Friendly
5. Unkind–Kind

B.2.13 Perceived Intelligence

Rate on a scale from 1 to 5 (Leftmost/Rightmost, 3 neutral)

1. Ignorant–Knowledgeable
2. Unintelligent–Intelligent
3. Incompetent–Competent
4. Foolish–Sensible
5. Irresponsible–Responsible

B.2.14 Perceived Safety

Rate on a scale from 1 to 5 (Leftmost/Rightmost, 3 neutral)

1. Agitated–Calm
2. Anxious–Relaxed
3. Surprised–Quiescent

B.2.15 title

Rate this statements on a scale from 1 to 5 (Strongly disagree: 1; Strongly agree: 5)

1. I perceived Aria as an intelligent part of the training system
2. I think the way Aria was acting distracted me from the main goal of the experience
3. I felt like Aria was intentionally reacting to my actions
4. I liked the way Aria was moving
5. I clearly understood the suggestion provided by Aria when needed
6. I clearly understood what Aria was saying
7. I liked the way Aria was interacting with me
8. I think Aria was a Human
9. I think other classmate were Involved
10. I think other classmate were Human
11. I didn't noticed other classmate except from Aria
12. I had the feeling it was "safe" to interact with the teacher
13. I felt anxious when needed to interact with the teacher
14. I felt stimulated to interact with the teacher
15. I had no fear to interact with the teacher
16. I preferred to use the chat to interact with the teacher instead of the voice
17. I would have preferred been the only student in the classroom
18. I feel more confident when I found there is a classmate in the class
19. If my classmate was not in the class, I could not learn as much as I did with her presence
20. I felt legitimated to interact with the teacher since Aria did it first ("icebreaker")
21. If Aria would have not been there I think I wouldn't have interacted with the teacher in the same manner
22. I think Aria was no good with the teacher

Bibliography

[1] Sveinbjörnsdóttir, B., Jóhannsson, S.H., Oddsdóttir, J. et al. "Virtual discrete trial training for teacher trainees." J Multimodal User Interfaces 13, 31–40 (2019), DOI 10.1007/s12193-018-0288-9

[2] Gebhard, Patrick, et al. "Serious games for training social skills in job interviews." IEEE Transactions on Games 11.4 (2018): 340-351.

[3] Ezen-Can, Aysu, et al. "Unsupervised modeling for understanding MOOC discussion forums: a learning analytics approach." Proceedings of the fifth international conference on learning analytics and knowledge. 2015.

[4] Fatahi, Somayeh, and Nasser Ghasem-Aghaee. "Design and implementation of an intelligent educational model based on personality and Learner's emotion." arXiv preprint arXiv:1004.1224 (2010).

[5] Kostarikas, Ioannis, Iraklis Varlamis, and Andreas Giannakoulopoulos. "Blending distance learning platforms and 3D virtual learning environments." (2016).

[6] Chen, Julian ChengChiang. "The crossroads of English language learners, task-based instruction, and 3D multi-user virtual learning in Second Life." Computers & Education 102 (2016): 152-171.

[7] Maratou, Vicky, Eleni Chatzidaki, and Michalis Xenos. "Enhance learning on software project management through a role-play game in a virtual world." Interactive Learning Environments 24.4 (2016): 897-915.

[8] Tegos, Stergios, et al. "Designing Conversational Agent Interventions that Support Collaborative Chat Activities in MOOCs." EMOOCs-WIP. 2019.

[9] Caballé, Santi, and Jordi Conesa. "Conversational agents in support for collaborative learning in MOOCs: an analytical review." International Conference on Intelligent Networking and Collaborative Systems. Springer, Cham, 2018.

[10] Liao, Meng-Yun, et al. "Embodying historical learners' messages as learning companions in a VR classroom." Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. 2019.

[11] Yoshimura, Andrew, and Christoph Walter Borst. "Remote instruction in virtual reality: A study of students attending class remotely from home with vr headsets." Mensch und Computer 2020-Workshopband (2020).

[12] Yoshimura, Andrew, and Christoph W. Borst. "Evaluation and Comparison of Desktop Viewing and Headset Viewing of Remote Lectures in VR with Mozilla Hubs." (2020).

[13] Schroeder, Noah L., William L. Romine, and Scotty D. Craig. "Measuring pedagogical agent persona and the influence of agent persona on learning." Computers & Education 109 (2017): 176-186.

[14] Adamson, David, et al. "Intensification of Group Knowledge Exchange with Academically Productive Talk Agents." CSCL (1). 2013.

[15] Tegos, Stergios, and Stavros Demetriadis. “Conversational agents improve peer learning through building on prior knowledge.” Journal of Educational Technology & Society 20.1 (2017): 99-111.

[16] Dyke, Gregory, et al. “Towards academically productive talk supported by conversational agents.” Productive multivocality in the analysis of group interactions. Springer, Boston, MA, 2013. 459-476.

[17] Ferschke, Oliver, et al. “Positive impact of collaborative chat participation in an edX MOOC.” International Conference on Artificial Intelligence in Education. Springer, Cham, 2015.

[18] Won, Andrea Stevenson, Jakki O. Bailey, and Siqi Yi. "Work-in-progress—learning about virtual worlds in virtual worlds: How remote learning in a pandemic can inform future teaching." 2020 6th International Conference of the Immersive Learning Research Network (iLRN). IEEE, 2020.

[19] Gao, Hong, et al. “Digital Transformations of Classrooms in Virtual Reality.” arXiv preprint arXiv:2101.09576 (2021).

[20] Lamb, Richard, and Elisabeth A. Etopio. “Virtual Reality: a tool for preservice science teachers to put theory into practice.” Journal of Science Education and Technology 29 (2020): 573-585.

[21] Barrett, Alex, et al. “Technology acceptance model and multi-user virtual reality learning environments for Chinese language education.” Interactive Learning Environments (2020): 1-18.

[22] Hernández-Sellés, Núria, Pablo-César Muñoz-Carril, and Mercedes González-Sanmamed. “Interaction in computer supported collaborative learning: an analysis of the implementation phase.” International Journal of Educational Technology in Higher Education 17.1 (2020): 1-13.

[23] Villegas-Ch, William, Adrián Arias-Navarrete, and Xavier Palacios-Pacheco. “Proposal of an Architecture for the Integration of a Chatbot with Artificial Intelligence in a Smart Campus for the Improvement of Learning.” Sustainability 12.4 (2020): 1500.

[24] Mozilla, Hubs by Mozilla [Online; accessed March 2021]. URL: https://labs.mozilla.org/projects/hubs/

[25] UX Collective, Tips on designing conversations for voice interfaces [Online; accessed March 2021]. URL: https://uxdesign.cc/tips-on-designing-conversations-for-voice-interfaces-d4084178cfd2

[26] Canonico, Massimo, and Luigi De Russis. “A comparison and critique of natural language understanding tools.” Cloud Computing 2018 (2018): 120.

[27] McTear, Michael. “Conversation modelling for chatbots: current approaches and future directions.” Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2018 (2018): 175-185.

[28] YouTube, Google Cloud Tech - Conversation Design Best Practices [Online; accessed March 2021]. URL: https://www.youtube.com/watch?v=5vwvyi5UmP8

[29] Lin, Tsun-Ju, and Yu-Ju Lan. “Language learning in virtual reality environments: Past, present, and future.” Journal of Educational Technology & Society 18.4 (2015): 486-497.