TSI-enhanced Pedagogical Agents to Engage Learners in Virtual Worlds

Steve Leung       Sandeep Virwaney     Fuhua Lin     AJ Armstrong     Adien Dubbelboer

This paper has been accepted to be published in International Journal of Distance Education Technologies Vol. 11, No. 1, 2013. Its short version was published in Proceedings of Internet of Things (iThings/CPSCOM), 2011 International Conference on and 4th International Conference on Cyber, Physical and Social Computing, 570 - 575

This paper presents a research project developed during 2019-2011 by the team led by the applicant consisting one undergraduate student Sandeep Virwaney and one master student AJ Armstrong both supervised by the applicant, and two junior faculty members of Athabasca University Steve Leung and Adien Dubbelboer.

The project integrates game-based learning, multi-agent system architecture (MAS), and the theory of Transformed Social Interaction (TSI). The project implements a group of engaging, affectionate and effective pedagogical agents equipped with abilities of self-representation, emotional states reasoning and situational awareness. A prototype of a virtual quiz show, QuizMASter, has been implemented to realize these abilities, and will be used to test for the effectiveness of the approach.
Abstract – Building pedagogical applications in virtual worlds is a multi-disciplinary endeavor that involves learning theories, application development framework, and mediated communication theories. This paper presents a project that integrates game-based learning, multi-agent system architecture (MAS), and the theory of Transformed Social Interaction (TSI). The project implements a group of engaging, affectionate and effective pedagogical agents equipped with abilities of self-representation, emotional states reasoning and situational awareness. A prototype of a virtual quiz show, QuizMASter, has been implemented to realize these abilities, and will be used to test for the effectiveness of the approach.

Keywords- virtual quiz games, pedagogical agent, transformed social interaction(TSI) theory, BDI agent, multiagent systems.

I. INTRODUCTION

Embodied as visually pleasing creatures or agents that live in 3D virtual environments, intelligent and autonomous software agents are able to react to their environment and use multi-modal interaction capabilities to engage users. Together, these technologies have given rise to a convergent concept called intelligent virtual environments [3]. This advance has revolutionized the virtual world genre of the Internet community that provides computer-simulated intelligent virtual 3D environments whereby users can interact with one another and use 3D objects embedded in the environment. Intelligent virtual environments allow for new and exciting research into developing animated pedagogical agents.

Intelligent virtual environments are characterized by avatars that are not only controlled by human users, but also by software. In contrast to a centralized control model, we argue that a multi-agent system (MAS) architecture is particularly well-suited to environments where virtual entities are self-directed and can actively pursue their goals by means of interaction with the environment, including other entities that are also in pursuit of their own competing goals. These entities, or software agents, are commonly designed based on the Belief-Desire-Intention (BDI) model in which they are given beliefs, desires, and intentions which are expressed by an agent through defining and pursuing goals with the development of plans [23]. The BDI model makes it easier for developers to work with, taking a programming exercise that is potentially abstract and grounding them in terms that are easily relatable. Individual interests, motivations, and goals of the agents can be easily monitored, manipulated and changed to suit the dynamics of the virtual environment, including the activities of human users. Mapping the agents to visible non-playing characters (NPCs) manifested in the virtual environment in the form of avatars, agents can reap the visual and immersive capabilities of virtual environments to support pedagogical goals.

Several successful pedagogical projects attribute an increase in learning effectiveness and learner satisfaction to the immersive power of virtual worlds [20]. However, technical innovations alone do not completely explain the successes. Non-technical factors such as pedagogical design, engagement level and communication among students and instructors are also important in delivering educational applications in virtual environments [2, 13]. We argue that appropriate design and implementation of the interaction model between the agents and the human users are equally important in achieving pedagogical purposes.

Base on the theory of Transformed Social Interaction (TSI) [5], the goal of our project is to demonstrate an approach to the development of pedagogical agents to enhance the engagement of students. Applying the three dimensions of TSI theory, we subtly transform the appearance and behavior of the avatars and agents based on the human players’ emotional engagement, with the expectation of raising learning effectiveness and learner satisfaction.

To achieve the goal, we designed and implemented a quiz game show called QuizMASter on a virtual world platform. The NPCs, supported by intelligent agents, are enhanced by TSI principles to guide the flow of the game and interact with the human players. This paper details our endeavor to integrating the strengths of virtual worlds, MAS architecture, and TSI theory.

The remainder of this paper is organized as follows: Section II describes the studies and theories that compose the foundation of our work. Section III describes our design and implementation of the game system. Section IV describes the experimental design to test the system’s effectiveness. Section V concludes with a brief summary of the paper and discusses future research directions for the system.
II. RELATED STUDIES

A. Educational Computer Games

Past research on motivation in educational computer games was dominated by Malone and Lepper’s “Taxonomy of Intrinsic Motivations for Learning” [16]. Their taxonomy asserted that challenge, curiosity, control and fantasy were the motivational elements for the players of educational computer games. However, this assertion was limited to isolated individuals. In contrast, recent learning theories have emphasized the importance of social and contextual factors in the learning process [9].

Rich social and contextual factors are conveyed by face-to-face interaction in a shared space [4]. Although virtual worlds are able to provide a sense of presence with virtual shared spaces, the lack of face-to-face interaction stills presence a challenge for engaging environment, especially in pure online learning environment in which face-to-face interaction is not feasible.

B. QuizMASter

QuizMASter is a multi-agent system (MAS) based educational game developed by a group of researchers at Athabasca University. It helps students perform adaptive testing and collaborative learning through friendly competition [8, 11].

Conceptually, QuizMASter is designed to be similar to a TV game show, where a small group of contestants compete by answering questions presented by the game show host. A prototype of QuizMASter using a JADE MAS was successfully implemented on Open Wonderland (http://www.openwonderland.org). The implementation provides an opportunity to explore the capability of expressing emotions and how to apply those emotions to avatar behavior. Specifically, the QuizMASter agents can be enhanced to operate within the dimensions of TSI to maintain high levels of engagement.

C. Agents’ Roles in Affective Game-based Learning

There have been several significant empirical studies in the area of pedagogical agents that provide for learning environments that enhance the engagement level of students. Research shows that learning programs with well-designed animated virtual pedagogical agents engage and motivate students, produce greater reported satisfaction and enjoyment [2].

Conati and Kort et al. investigated the use of intelligent agents to monitor student emotion when playing educational games, so that they could intervene when students become overly despondent [10, 14]. If the student became frustrated after making a mistake, the agent intervened and attempted to make the student feel better about their performance. By focusing on the affective state of the player, the agent not only makes the learning experience enjoyable, but also more productive.

Kosinowski developed several agents to illustrate the benefits and technical issues of pedagogical agents [15]. In a study with one of the agents, Kosinowski inspected the motivational aspect of an agent’s presence in an environment. By including several modalities of simulated human-human dialogue into the tutoring process, learning environments with an agent become more appealing to use. This study shows that the persona effect is present even in mute agents and thus, for motivational purposes an agent should be considered for any developed learning environment. Although a dynamic approach (nonverbal actions, body movement) is implemented in some of the agents, environmental awareness remains an issue. Kosinowski concludes that a representation of all relevant facts of the environment has to be created for the agent with the following features: 1) The agent must know about the current task’s status, its own position and which utterances have been made in the dialogue. 2) Idle-time behavior elements (yawning, weight-shifting, foot-tapping) as well as a humanlike visual appearance are required to create a believable agent. She also envisioned that further research would require cooperative work between communication theorists (dialogues), linguists (speech), graphics specialists (agent appearance) and animators (agent movement).

Zakharov, Mitrovic, and Johnston employ a dimensional approach to track affective state along valence dimension by identifying and categorizing users’ facial features [23]. They developed a facial-feature tracking application and a set of rules that control an agent’s behavior. The agent will respond to a user’s action depending on his/her cognitive state (as determined from the session history) and affective state. To accommodate the preferences of a variety of users, two female and two male Haptek2 characters with appearances modeling young people approximately 20 to 30 years of age were created for an experiment. The results indicated a range of preferences associated with pedagogical agents and affective communication. In other words, affective interaction is individually driven and carries less importance for many learners.

Yan and Agada developed a procedure to produce head and face movements during speech by a virtual pedagogical agent by combining different voice recordings, facial expressions and head movement patterns [22]. Results of their experiments show that facial expressions and head movements have significant impact on students’ impressions and engagement with a virtual pedagogical agent. Virtual pedagogical agents that produce natural head movements and appropriate facial expressions while narrating a story generate many more positive user experiences than virtual pedagogical agents that lack these behaviors.

Although empirical studies show a relationship between the psychological or affective states of the users and visual traits or gestures of the representative characters or avatars, it does not address the problem of programming the behavior of the
non-player characters or intelligent agents to engage human players. For this we turn to the theory of mediated communication.

D. Mediated Communication and Transformed Social Interaction

Another important research area focuses on the engagement and communication of interacting parties, or interactants, in mediated environment. Also conducted in virtual world environments, empirical studies show that engagement and communication between students and instructors are essential in delivering effective online education [2, 13].

As a mediated communication theory, Bailenson et al. suggest Transformed Social Interaction (TSI) to describe the transformation of interaction in mediated communication environments. TSI postulates that interactants can be transformed in three categories or dimensions: self-representation, sensory-abilities and situational context [5]. In particular, it is feasible to reap the unique technical features supported by virtual worlds to enhance the communication capabilities of avatars and agents as well as the engagement level of the human players.

Self-representation: Self-representation is the dramatic and subtle changes to the appearances or behaviors of the avatars chosen by the users, such that emotional bonding can be established. In other words, the avatars do represent the users in virtual worlds. Self-representation is rooted in the study of mediated communication. Although the presentation of the “real” self is questioned in online environments, mediated communication may also empower the user to present an ideal self to others [4]. In particular, the use of avatar as mediation for communication increases emotional closeness and perceived communication effectiveness [7]. As an instance to support the idea of self-representation, Bailenson et al. found that the morphing of faces of political candidates with potential voters increases the affective bonding in a context with few social cues [6].

Sensory-Abilities: Avatars in virtual worlds are practically bounded by the visual and audio senses of the human users behind the avatars. In other words, avatars see and hear what the human players see and hear. However, avatars controlled by software do not share the same senses. By complementing the avatars with senses analogous to seeing and hearing, these transformations empower the avatars to engage the human players. In face to face communication, it has been argued that humans communicate with each other in a wide range of modes, including verbal and non-verbal cues. These social cues contribute to the richness of communication in real life [4].

Increasing the capability of avatars will empower the avatars to mimic the rich model of communication of humans. Taking eye gazing as example, it is estimated that 60% of the time involves gazing in face-to-face communication (Argyle, as cited in [19]). People look twice as much while listening (75%) as while speaking (41%). Vertegall and Ding concluded that people are more likely to speak when they are being gazed at more [19]. If we can add to the agent the capability of eye gazing, it is expected that we can further enhance the engagement of human users in virtual worlds.

Situational context: These transformations alter the spatial or temporal structure of a conversation. For example, the communication between agents and students can be optimally configured in terms of the geographical setup of a conference room. Each student in a class of twenty can sit directly in front of a virtual instructor and perceive all others as further away. Furthermore, by altering the flow of rendered time in the communication session, users can implement strategic uses of rewinding and fast-forwarding during a real-time interaction to increase comprehension and efficiency. Spatial and temporal structure is supported by the theories of presence such that effective communication, as in face-to-face communication, is related to a sense of shared space and time [4].

III. DESIGN OF ENGAGING PEDAGOGICAL AGENTS

Based on the original version of QuizMASter, we added to the capabilities of agents and avatars according to the following criteria according to TSI theory:

Self-representation: We implemented the self-representation dimension of TSI using facial-identity capture with a tool called FAtiMA. Each of the users’ face was morphed with QuizMASter agent’s face to capitalize on human beings’ disposition to prefer faces similar to their own [3] and general preference for appearing younger [16].

Sensory-abilities: We implemented the sensory-abilities dimension of TSI using a movement and visual tracking capability. The general challenge of sensory abilities implementation lies in two areas: the complexity of human senses and the processing of sensory data of different modality and historicity. For reasons of simplicity, we only monitored visual tracking, and thus removed the need for modality. We used a one dimensional scoring system to measure the level of engagement when the user is visually looking at or away from the host agent or the contestants. When the user is gazing at the host agent, the score is incremented, and when the user moves the viewing direction away from the host agent or the contestants, the score is decremented.

Situational-context: We implemented the situational-context dimension of TSI by using the best-view feature of Open Wonderland [17], whereby the temporal structure of a conversation can be altered. For example, each participant of QuizMASter either sees the actual behavior (e.g., facial expressions, direction of eye gaze, nodding and head shaking behavior) of other group members or sees transformed behaviors that are created to induce participants to conform to a certain behavioral model.
To test for the effectiveness of the approach, we proposed the following general hypothesis: enhancing the pedagogical agent with the self-representation, sensory-abilities, and situational context dimensions of Transformed Social Interaction theory will result in higher levels of student emotional engagement and increase satisfaction. We will conduct an experiment with two scenarios: The first scenario will use agents with none of TSI dimension capabilities applied, while the second scenario will involve agents with TSI dimension capabilities. Satisfaction of students will be compared and measured using a questionnaire.

IV. IMPLEMENTATION

Implementation of the design is technically challenging. Results from the work of Yan and Agada [22] provide a platform for creating believable pedagogical agents in an immersive and interactive virtual environment like Open Wonderland.

There is an interest in the gaming industry for AI middleware, so it seems reasonable to consider a more modular approach, integrating with existing technologies, rather than building something from scratch. With this in mind we have decided to combine a Virtual World engine with a MAS system. This has already been done to some extent with Gamebots [1], but there were issues that could not be addressed because the system being used in that implementation was not open source. With this in mind, we elected to use Open Wonderland (http://openworld.org) as the virtual world technology and Jason (http://jason.sourceforge.net/Jason/Jason.html) as the MAS platform. All of these technologies are open source, affording us full control of both the server and client code to make any changes that are required to develop a fully working integrated system.

ARCHITECTURE

QuizMASter and Open Wonderland are implemented on multiplayer online game middleware called Darkstar server, as shown in Figure 1. All actions in the server environment are divided up into sequences of short tasks that are executed within a transaction that immediately writes results to an internal database, guaranteeing that state is not lost even during server failures. It also provides an abstract communication mechanism, allowing simple messages to be sent to the server from a client, and from the server to any subset of other clients connected to the same server.

JVoiceBridge: This is a pure Java audio mixing application that was developed as part of Open Wonderland to provide server-side mixing of high-fidelity immersive audio. It runs a separate server that mixes SIP audio for multiple users, based on where the virtual space they are. It communicates directly with the Darkstar server over a private channel in order to keep all the audio in sync with the state of the world as users are added, removed, and move around.

Shared Application Server: The shared application server (SAS) is the final standard server component. The SAS allows server-hosted application sharing. Multiuser collaboration-aware applications are written specifically for Open Wonderland, and are implemented entirely in the client and Darkstar server. Moreover, the user model is stored in the SAS.

Open Wonderland Client: The Open Wonderland client is a single application that acts as a browser for connecting to an Open Wonderland server. As with the server, the client provides several core services based on existing open source components. The rendering layer of the client consists of two separate projects. JMonkeyEngine is a popular rendering framework for writing OpenGL-based applications in Java. JMonkeyEngine provides the basic scene graph and rendering framework, but is limited to working on a single thread at a time. The core services layer provides features used by Open Wonderland modules. These services include the position of objects in the 3D world, the ability to move objects, and collision detection. Extended core services, like the ability to load models, calculate real physics, and enforce security, are layered on top of the core as modules.

MAS CArtAgO Artifacts: CArtAgO (Common ARTifact infrastructure for AGents Open environments) is a general purpose framework/infrastructure that makes it possible to program and execute virtual environments for multi-agent systems. CArtAgO is based on the Agents & Artifacts (A&A) meta-model for modeling and designing multi-agent systems. Due to the fact that these artifacts can be modeled in Java, some of the core components of the Open Wonderland (OWL) such as non-playing characters (NPC) have been wrapped as artifacts in order for the Jason agents to manipulate, manage, and communicate with them.

MAS Jason-CArtAgO bridge: The capability of Jason agents to create a workspace using the bridge provided by CArtAgO is leveraged to create a workspace and run the OWL client within a workspace. Once this workspace node is exposed on a

Figure 1: QuizMASter TSI architecture.
network, it allows the Jason agent to join a workspace of other OWL clients and manage, manipulate and communicate with artifacts of the respective OWL clients, using RMI.

One of the main advantages of leveraging the capabilities of Jason-CArtAgO MAS is that it enables customization of the rendering of visible artifacts for different clients such as NPCs. The rendering messages sent to the clients within OWL simply need to be disabled in addition to setting up the above mechanisms.

Figure 2: QuizMASter in Open Wonderland.

SCENARIOS

We applied simple and generic rules for every QuizMASter game:

1) All students receive a question and have one minute to answer.
2) One minute after issuing the question, the answers given by all the students are evaluated. A point is given for each correct answer.
3) At the completion of a round of questions, the student(s) with highest score wins the game.

When a student registers to play QuizMASter, the system generates a student agent to collect data about the learner using pre-tests and questionnaires. The student agent builds the student model through continuous monitoring of the behavior and performance of the student. The information collected is used by a QuizMASter agent to generate a profile of the registering student. Based on this profile, a pedagogical agent uses an adaptive testing algorithm to determine the student’s knowledge or proficiency level and determine their starting level. Using an item bank, a student level can be estimated by grouping questions of equal difficulty in the same bin. According to the correctness of the current response by the student, the agent will select the next test item adaptively, either from the more difficult items in the left bin or from the easier items in the right bin. If there are no other students at a comparable level, non-playing characters are used.

A game comprises a number of rounds. In each round, a pedagogical agent uses a root-mean-square averaging algorithm to determine the joint proficiency level of the group. Based on the result, the agent selects a number of candidate questions from the repository. Another agent acting as the game show host then presents the questions to the contestants.

Contestants score points by correctly answering questions before their opponents [3] or before the timer runs out. In the context of the multi-agent environment, answering a question is accomplished using the Contract Net Protocol (CNP) [10].

The CNP has two stages: the proposal stage and the contract stage. In the proposal stage, the host agent as the initiator asks the contestants for the answer or “proposal” to a question by sending a CFP performative message. The contestants can either respond with an answer proposal or refuse to answer the question. In the contract stage, the host then sorts the answers according to correctness and response time. The contestant that sends a sufficiently correct answer fastest is awarded a point.

By using CNP, answering a question may be viewed as a kind of contract negotiation. Here all contestants are given the chance to answer every question while only one contestant can receive the point. This improves the pedagogical agent’s ability to determine the knowledge-base of the group. The answer given by each contestant, along with the time taken to respond and the amount of hints used is further analyzed by the pedagogical agents to determine and update information about the knowledge level of the individual and the group as well as provide feedback to the instructor. The data collected may also be used for machine learning. For example, the records of the student’s performance can be used to improve the metadata of repository questions and expand QuizMASter’s knowledge base and the quality of the student models it can generate. Finally, the contestant with the highest score is declared the winner of the game.

Contestant proficiency and quiz items

The QuizMASter game does not discriminate between contestants initially. However, based on the contestant performance and the parameters of the questions, contestant proficiency is recorded and used for further selection of quiz items.

Each quiz item has several sets of parameters:

Content related:
• Ontology or topics
• Answers

Format related:
• Problem type (e.g., multiple choice, true/false, word problem)
• Available hints
Difficulty related:
- Bloom’s Taxonomy [9]
- Difficulty level
- Discrimination of contestant level (how well the response to the question reflects the learner’s real ability [17])

When a contestant answers a question correctly or incorrectly, the difficulty and content parameters of the question, as well as the time taken to answer the questions, are recorded to build the proficiency profile of the contestant.

V. SCENARIOS AND TESTING

Two scenarios are designed to distinguish the effect of applying a TSI dimensions capability to a virtual pedagogical environment with agents:

Scenario A: Scenario A employs the TSI approach. The preliminary image of the Host agent will be morphed with students’ images and rendered differently to each student reflecting the self-representation dimension of TSI (see Figure 3). In Figure 3, Stephan is the face of the generic pedagogical agent (Host), while Wyatt is the uploaded face of a student. Below the two pictures there is an adjustment bar that can be used to adjust the amount of morphing that we need to use. According to TSI theory, 40:60 is a good ratio and was used to generate the transformed face shown on the right hand side. We used a tool called evolver (http://www.evolver.com) for this as the Agent NPC images can be easily used in OWL and OWL allows evolver avatars.

![Figure 3: Image morphing](image)

Sensory abilities of TSI will be applied by recording and observing the student avatar behaviors, especially his/her gaze direction. There will be a crowd of agents that will primarily make sounds and gestures of applause and empathy.

After all contestants enter the studio, the Host Agent will greet each and every student, addressing him or her by name with eye-gaze directed to that student. Note that the rendering of the second greeting will be customized for each student applying the self-representation and situational context dimensions of TSI.

Rules of the game are explained by the Host to all students. This is followed by the agent individually confirming with each student, again addressing them by name, whether they understood the rules or not.

If any student indicates that he or she did not understand the rules, the rules will be explained again to that particular student only, using the decoupling of rendering described by the TSI dimension of situational context. This will be followed by the question and answer session. If the student answers correctly, the crowd reaction would be to express delight (cheering and clapping with sound and animation) in that student. If the student gets the answer wrong, the crowd reaction would be to express sadness (squirming with sound and animations) for that particular student. In addition, if the student provides a correct answer, the Host Agent will congratulate the student, and for the student who gets it wrong, the Host agent will wish him luck for the next question and will give the correct answer to him or her. The renderings for each student will be different. This customized experience of crowd cheering for each student along with the Host seemingly paying more attention to him or her than anyone else will help the student feel more emotionally engaged.

All negative behaviors of the students will be hidden from other students. The displayed score of student agents will depend strictly on the performance of the real students. If Student A is leading by a large margin compared to Student B, then the Agent for Student A will score greater or equal number of points to a real student. However, the Agent for Student B will lag (in terms of scored points) in comparison to the student. Furthermore, hints will be provided to Student B in this case. The real student B, then, will never sit in last place or score 0 points, thus making sure that he or she remains encouraged and motivated. All the renderings of hints provided to a weaker student will not be delivered to stronger students.

Scenario B: Scenario B does not engage the TSI capabilities of agents. After all students enter the QuizMASt er studios, they will be greeted by the Host agent without any personalized greeting. Rules of the game are explained by the Host Agent. If any of the students do not understand the rules, the rules will be reiterated to everyone. Every question and answer session will be followed by the Host Agent and Real Student chat (banter talk). This episode of between questions chat between the Host agent and the real student will be rendered to all students. Hints will not be given for this scenario.

VI. CONCLUSIONS AND FUTURE WORK

In this paper we presented a project that employs a multi-disciplinary approach to the building of a pedagogical application in virtual worlds. While many studies support the merits of various approaches in each of the disciplines, to realize a working application requires significant work. In particular, we employed the immersive power of 3D graphics and the use of avatars supported in virtual world engine. We
followed the TSI theory to increase the engagement with users. We took a game learning perspective and design learning content approach that is in accord with adaptive principles and an ontology approach. The entire application was developed using a multi-agent system to permit the integration of the various disciplines. The result is a well-implemented and engaging game show with pedagogical content.

There are two directions to look into the future of developing pedagogical applications in virtual worlds. First, the TSI or mediated communication theory is in fact richer than we have thus far implemented. We agree with the argument that the combination of looks and behaviors constitute an important dimension of realism that enhances engagement of users [12]. To integrate behavioral transformation into the agents is a challenge. Similarly, the implication of sensory abilities and situational transformation means more than simple movement tracking and avatar interaction. Based on the evaluation of the initial results, we plan to design more of the related abilities and interactions into the agents.

Second, from an agent perspective, the intelligence of agents can be enhanced to act or react more realistically to a more complex environment. The interaction of the agents and the environment is manifold and comprises information including performance in the game show, emotional states, visual and non-visual cues, and the spatial and temporal structure. Given this large amount of available context information, a general and coherent model of the users in the game at any given point of time is significant to the effectiveness of our approach.

ACKNOWLEDGEMENT

We thank the Natural Science and Engineering of Research Council of Canada (NSERC), Canada Foundation for Innovation (CFI), and Athabasca University for sponsoring this project.

REFERENCES