Learner Modelling and Automatic Engagement Recognition with Robotic Tutors

Fotios Papadopoulos, Lee J. Corrigan, Aidan Jones, Ginevra Castellano
School of Electronic, Electrical and Computer Engineering
University of Birmingham
Birmingham, United Kingdom
f.papadopoulos, ljc228, axj100, g.castellano@bham.ac.uk

***Abstract*— In this paper we discuss the initial steps of the design of a computational model capable of automatically detecting user engagement with an artificial robotic tutor. Generally, our model will be used with a robotic tutor to support and enhance the experience of the learner by regulating pedagogical and empathetic interventions in a timely manner. We describe the requirements of the learner model in order to understand the state of the learner and facilitate the learning progress. Additionally, we propose the initial steps of the design of a suitable scenario for the learning task activity to allow the model to be tested on actual class material from UK curriculum based on teachers' feedback. Finally, we discuss an initial plan to build the model with real user data from a pilot study based on Wizard-of-Oz (WoZ) conducted in a real classroom environment.**

Keywords—affect recognition; artificial tutors; human-robot interaction; social robots; automatic detection of engagement

I. INTRODUCTION

One of the most critical challenges in education has always been the attempt of establishing and maintaining a student's engagement throughout the learning process in an attempt to accommodate efficient and continuous level of learning. This work is part of a larger project that aims to identify the role of empathic and pedagogical strategies and how these affect the quality of engagement in a learning process supported by a robotic tutor in a shared physical space with the learner. As a result, it will endeavour to measure whether a robotic tutor that perceives affective states and emits socio-emotional cues, can sustain engagement as in real school environments during the learning process with a student.

Nowadays, robots come with multiple sensors for measuring human affective states such as, distance sensors, camera, microphone arrays and touch sensors. Affective information in human interactions is transmitted through cues such as face and body expressions, voice tonality and spoken words. The significance of non-verbal behaviours in human interactions is supported by findings from Ekman [1], where it is suggested that these cues as the primary indicator for expressing emotion. Additionally, external sensors can also be utilised in order to increase the perceptive capabilities of a robot such as a Microsoft Kinect sensor or devices that measure electro-dermal activity (EDA). An artificial tutor which is capable of analysing this information could provide the learner with personalised feedback based on their overall affective state. Research has shown that students' learning capabilities are highly dependent on their affective state. Therefore, focusing on strategies that promote tutor-student engagement is of paramount importance [2]. Although a virtual tutor has shown positive learning results in student engagement [3], additional research has shown that physical embodiments are more preferable to users as they offer higher enjoyment rates [4], better performance [5] and enhanced social presence [6]. Additionally, recent research suggests that when compared to traditional learning formats, such as books and web based instruction, an educational robot can increase concentration, learning interest and academic achievement [7]. Previous studies in classrooms indicated that properly designed robot companions can successfully contribute to experimental learning, making their future very promising [8]. Furthermore, similar research has shown that students can become engaged with their learning activities if they are challenging, enjoyable and in a "zone of proximal development" [9] but without being too difficult or frustrating. Therefore, it is of paramount importance to create and maintain users 'engagement with the task and the artificial tutor for the whole duration of the learning activity by incorporating and analysing the learner's emotions into a computational system which will allow it to automatically generate the appropriate response to the user based on his current emotional states. Recognising a user's affective state is a primary concern in human-robot interaction (HRI) as it can be used to create an affective loop with the user capable of creating and maintaining the social relationship with the help of an appropriate and acceptable automated response from the robot or the agent [10].

This paper will describe a high-level scenario in order to ground the focus for the short- and long-term studies and list our methodology steps required for developing appropriate automatic engagement algorithms and learner model with adaptive difficulty levels. More specifically, it will discuss the requirements of the learner model, which is an essential step in the process of translating from a theoretical design into a computational model capable of perceiving and reacting to the affective state of the learner. Lastly, it will analyse our proposed methodology for testing and evaluating our system in order to further develop and enhance a computational model for affect recognition in real-class scenarios.

II. RELATED RESEARCH

Engagement and learning are very closely related to each other in real classrooms scenarios where a tutor needs to keep students in an engaged manner in order to increase their learning capabilities. Trained teachers can recognize learners' affect during the class and deliver the course according to their states. Learning capabilities are directly influenced from certain affective states, with the more prominent being boredom, flow, frustration and confusion [11]. However, Graig et al. suggests that the two most important and relevant affective states are flow, boredom and confusion [12]. They investigated the correlation between boredom levels and learning abilities in students supported by intelligent tutors and reported that learning was negatively correlated with boredom. On the contrary, the exhibition of confusion positively affected the learning process of the students during the session. On the other hand, flow experience, which delivers high engagement [13], has been found to be positively correlated with learning. Furthermore, recent studies have shown a direct link between affect and motivational, cognitive and behavioural processes in learners [14], while Pekrun suggests that affect greatly influences cognition and learning in terms of attention, strategy use, memory and motivation [15]. Learners' affective state has also been used in emotion-sensitive intelligent tutoring systems where they integrate students' emotions in order to guide their pedagogical approach for enhanced engagement, motivation [16] and consequently, learning [17]. Additionally, learning experience has found to be significantly affected by emotions such as frustration, anxiety, delight, surprise and confusion; particularly in deep learning [18].

In order for an intelligent system to detect engagement and other affective cues (verbal and non-verbal), a number of hardware sensors capable of perceiving socio-emotional cues are required. An example of such sensor is a device used by researchers in order to record the galvanic skin response from participants and identify their frustration during a math exercise [19] as well as stress levels [20] and generally, a number of emotions developed during a computer game [21]. Additionally, user posture information is also used on automatic affect recognition systems in order to predict engagement levels. D'Mello et al. performed experiments for automatic recognition of student engagement by utilizing a special chair equipped with pressure sensors that identify user's posture [22] where their computational model accurately predicts the learner's affecting state. Mutuality and synchronisation are two very closely related terms when it comes to recognising engagement in HRI [23]. A computational model derived from results of previous studies has been created in order to recognise the engagement between humans and robots. In particular, this model includes four event recognisers for detecting direct gaze, mutual face gaze, conversational adjacency pairs and backchannels (a brief verbal or gestural communication from the responder back to the initiator). Past and present HRI researchers have attempted to establish and maintain the cognitive connection between a human and a robot. Consequently, researchers have used an array of both pervasive and non-pervasive sensory technologies in an attempt to infer humans' affective state during interaction. Rich et al., explored the issue of recognising engagement using more pervasive methods of detecting affect, gaze and verbal utterances, building a detailed computational model of an interaction with what they call "connection events" [23]. Szafir and Mutlu adopt a non-pervasive methodology with a wearable brain-computer interface, measuring student attention from electro-encephalography (EEG), which was shown to be effective in detecting drops in attention [24]. When combined with information such as gaze and pose another effective modality for engagement and determining the desire for an interaction with a robot can be inferred from the proximity of the human with the robot [25]. Overall, the robotic tutor needs to be predictive in order to appear realistic, as a purely reactive behaviour might not reinforce the impression that the robotic tutor is a capable interactive partner. This is further supported from studies with human subjects where they were found to prefer an anticipatory agent as it appeared to be more fluent and committed than a purely reactive agent [26].

In educational settings, engagement and learning performance are often associated with each other [27]. Attempts have been made to discriminate between coincidence due to the social and situational context and changes in a learner's engagement state and task progress which are linked to recent actionable behaviour, intervention and adjustment within the platform. Using an adaptive agent another recent study found that recall performance was improved through detection and maintenance of student attention, showing that virtual agents need to be direct and immediate in their response to drops in attention and engagement [24]. Chanel et al. found player engagement in a game would decrease if the challenge and other aspects of the game were not modulated according to the emotions experienced by the user [28]. Additionally, Kapoor and Picard [29] proposed a recognition system to classify levels of interest of children in a learning task, combining facial expressions and postural shifts with information about the learner's task. Student engagement is a mix of several physiological aspects: 1) behavioural(e.g., persistence and participation), 2) emotional (e.g., interest, value and valence); and 3) cognitive (e.g., motivation, effort and strategy) [30]. Engagement with a learning task is major contributing factor in the promotion of both short and long-term learning performance [13], and considerable efforts have been made to explore social [31] and task engagement [32]. Further still, robotic tutors for educational HRI scenarios is attracting much interest [24], but the fact remains that most of this work involves robotic tutors with limited personalisation capabilities and has not been performed in the wild (i.e. within a real classroom environment).

III. LEARNER MODEL: REQUIREMENTS

An important requirement for an intelligent artificial tutor is to recognize and understand the state of the learner in order to facilitate their learning progress. A key aspect of learner modelling is the learner's affective states such as enjoyment, hope, pride, relief, anger, anxiety, shame, hopelessness, and boredom [33]. A learner's emotion could be directed towards three different targets: "emotion–for-game, emotion-for-self and emotion-for-agent" [34]. Depending on the addressing target and the status of the task, the learner model is required to understand and model the likely effect of an intervention on the learner. It is therefore essential to be able to identify the emotional effect of the intervention and the effect on the learning process. A properly designed learner model should intervene by deploying a strategy to assist the learner based on his current affective state [35]. However, studies have shown that the system should not intervene when a learner is in a state of flow or a positive learning state as this intervention might disrupt learning [36]. Additionally, interventions appropriate or not, could also cause disruption to the learning process. Furthermore, the model has to identify the engagement level with the learner by recognizing his state of "flow". Flow is when a learner is fully engaged thus making progress and enjoying the learning activity [33]. Conversely, stuck is when a learner becomes disengaged, is not making progress in the learning activity and not enjoying the task. Additionally, the model should capture the learner's progress through the learning activity in order to plan ahead and offer a personalised task.

Opening the learner model to the learner can benefit the learner by prompting reflection [33]. The learner can be prompted to think about their affect, knowledge and the process they are following as this can improve the learner's meta-affective and meta-cognitive skills [37]. It is also possible to improve the accuracy of the learner model by allowing the users to leave positive or negative feedback on the model.

In order to capture and analyse learner's interaction with the model such as time in activity, button presses, dialogue length, number of restarts, number of repeats etc., the model should be able to process and infer these input along with the state of the learner and produce appropriate behaviours, either for the task or artificial tutor.

IV. SCENARIO

In order to evaluate the artificial tutor's effectiveness, performance and engagement levels during the learner's interactions, a properly designed task oriented learning task is essential. For that reason, this project will focus on developing a suitable Geography based learning activity aimed at 11-13 year olds to be run on a touch table supported by a NAO robot (Figure 1).

Figure 1: Touch table with robotic tutor

A properly designed task should be adaptive to users' abilities; therefore, it should offer various modes of interaction. For our geography-related task, which is currently under development, we decided to use two modes of interaction: discovery and guided learning. In the discovery learning mode, the task allows the student to independently learn and discover, thus enhancing their feeling of control and independence. Discovery learning is the process whereby students self-direct their learning to achieve the learning goals, by identifying and addressing gaps in their knowledge and skills [38]. Additionally, students can enhance their learning through discovery by introducing short-term experiments and assessments, therefore, the learner can formulate their own working hypothesis to be tested through independent exploration, which is shown to result in the learner gaining a deeper understanding of the material [39]. In order for this approach to be effective, the learner should have a good understanding of his own meta-cognitive state while the system should be able to prompt meta-cognitive strategies to support this [24]. It is particularly important that the robotic tutor intervenes and guides the learner where the student is unable to solve a problem or find information themselves at any point. The learning activity should vary the difficulty level to keep users engaged and minimise fear of failure by detecting users' state of flow and respond accordingly by altering the activity or by allowing the robot to intervene.

V. METHODOLOGY

In this early stage in the project, a number of studies will be performed in order to identify and investigate indicators of engagement with the robot and learning task. More specifically, a pilot study will be performed with the purpose of grounding the initial setup and identify behavioural and contextual (e.g., robot, task) indicators of user engagement with the robot and the learning task. Additionally, it will investigate links with human-robot and human-task social bonding. Secondly, a Wizard of Oz (WoZ) study is scheduled to run at a later stage in real class environments with a more focused learning scenario but controllable tutor.

A. Pilot Study

In order to develop and run the pilot study, we will work collaboratively with psychologists in an attempt to achieve a practical identification of the most pertinent indicators in terms of behaviour and context in user engagement. Initially, we will run a short number of simplified tests involving the learning platform which will consist of the touch table, the robotic tutor and a simple learning task in an attempt to examine and identify associations among student engagement and social bonding. The primary goal of the pilot study is to identify and explore possible indicators for user engagement with the task and the robot. In order to identify these indicators a series of experimental conditions will be designed, developed and evaluated from a group of participants. These will include a battery of short tests to extract engagement indicators such as lower and upper bounds for engagement detection within a task and social bonding between the participants and the robot. A series of simple tasks will be designed with predefined difficulty levels in order to capture extremes of engagement i.e. low and high levels. Furthermore, a condition involving the robotic tutor will be presented to the participant in order to identify and capture empathy and social bond towards the robot. The setup consists of a video camera for capturing human-robot and human-task interaction, a Microsoft Kinect sensor for predicting body lean position, a real-time face recognition software for extracting facial expressions, head direction and eye gaze, a Q Sensor[1] for measuring affective states via electro-dermal recordings and finally, information from the learner task (touch, response quality etc.). Findings and feedback from this study will inform the system architecture, tutor behaviours, hardware implementation, learner model and overall the scenario. Additionally, the pilot study will include a separate task for comparing a controlled condition scenario with findings from the main task in order to ground the engagement findings.

B. *Wizard of Oz study*

The second experiment will be built upon the findings from the pilot study and will run in real classroom environments with learning content having been informed and supported by teaching experts. Additionally, teaching experts will inform and support the content of the learning task with the purpose of increasing its pedagogical value and reflect to children's needs and requirements. This study will utilize the same capturing equipment as in the pilot study but with more targeted questionnaires. Additionally, a corpus will be collected from the study and will be used towards of developing the final computational model capable of automatically recognizing and classifying engagement in real classroom environments.

C. *Learner model evaluation*

The learner model will consist of engagement detection (with task and robot) and detection and monitoring of learner progress in order to promote a personalised learning experience. Additionally, the model will be built and enhanced by taking into account critical moments in the interaction when empathic and pedagogical interventions from the tutor are required. Since the work has been designed to be modular, both the learning task and the platform can be tweaked based on our findings, particularly in terms of teaching strategies, challenge, socio-emotional intervention by the robotic tutor and in-task sensory appeal. At a later stage, the model will be integrated with an artificial tutor (both virtual and robot embodiments) and evaluated in real-classroom environment.

VI. CONCLUSIONS

In this project, we intend to advance the state of the art through the development of a new breed of supportive educational robots. We aim to achieve this objective through the design, development, investigation and evaluation of a computational learner model which automatically detects the state of engagement and adapts the task and teaching strategies to the individual, personalising and enhancing the overall learning experience. We will show through integration with a robotic tutor that our model, supported and informed by interdisciplinary research on affect recognition, learner models, pedagogical strategies and adaptive behaviour, can make robots become effective pedagogical tools for use in education.

VII. ACKNOWLEDGMENT

This work was partially supported by the European Commission (EC) and was funded by the EU FP7 ICT-317923 project EMOTE. The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

VIII. REFERENCES

[1] P. Ekman and W. V. Friesen, "Nonverbal behavior in psychotherapy research," in *Research in psychotherapy*, Washington, DC, US: American Psychological Association, 1968, pp. 179–216.
[2] J. L. Robison, S. W. Mcquiggan, and J. C. Lester, "Modeling Task-Based vs. Affect-based Feedback Behavior in Pedagogical Agents: An Inductive Approach," in *Proceedings of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling*, Amsterdam, The Netherlands, The Netherlands, 2009, pp. 25–32.
[3] S. W. McQuiggan and J. C. Lester, "Modeling and evaluating empathy in embodied companion agents," *Int J Hum-Comput Stud*, vol. 65, no. 4, pp. 348–360, Apr. 2007.
[4] A. Pereira, C. Martinho, I. Leite, and A. Paiva, "iCat, the chess player: the influence of embodiment in the enjoyment of a game," Richland, SC, 2008, pp. 1253–1256.
[5] C. Bartneck, G. T. Bünde, B. Christoph, E. Aarts, D. Bouwhuis, M. Bukowska, E. Diederiks, J. Hoonhout, H. Jun, Y. Katagiri, and A. Kohlrausch, *eMuu-An Embodied Emotional Character For The Ambient Intelligent Home*. 2002.
[6] C. D. Kidd, "Sociable robots: The role of presence and task in human-robot interaction," Massachusetts Institute of Technology, 2003.
[7] J.-H. Han, M.-H. Jo, V. Jones, and J.-H. Jo, "Comparative Study on the Educational Use of Home Robots for Children," *J. Inf. Process. Syst.*, vol. 4, no. 4, pp. 159–168, Dec. 2008.
[8] I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva, "Modelling empathic behaviour in a robotic game companion for children: An ethnographic study in real-world settings," in *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2012, pp. 367–374.

---

[1] http://www.qsensortech.com/, Affectiva Q sensor, Last accessed 25-4-2013

[9] A. L. Brown, S. Ellery, and J. C. Campione, “Creating zones of proximal development electronically,” in *Thinking practices in mathematics and science learning*, J. G. Greeno and S. V. Goldman, Eds. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 1998, pp. 341–367.
[10] C. Breazeal, “Emotion and sociable humanoid robots,” *Int J Hum-Comput Stud*, vol. 59, no. 1–2, pp. 119–155, Jul. 2003.
[11] A. Graesser and P. Chipman, “Detection of emotions during learning with AutoTutor,” in *Proceedings of the 28 th Annual Meetings of the Cognitive Science Society*, 2006, pp. 285–290.
[12] S. Craig, A. Graesser, J. Sullins, and B. Gholson, “Affect and Learning: An Exploratory Look into the Role of Affect in Learning with AutoTutor,” *J. Educ. Media*, vol. 29, no. 3, pp. 241–250, 2004.
[13] M. Csikszentmihalyi, *Flow□: The Psychology of Optimal Experience*. New York: Harper Perennial, 1990.
[14] E. A. Linnenbrink and P. R. Pintrich, “Multiple goals, multiple contexts: The dynamic interplay between personal goals and contextual goal stresses,” in *Motivation in learning contexts: Theoretical advances and methodological implications*, S. Volet and S. J, Eds. Elmsford, NY, US: Pergamon Press, 2001, pp. 251–269.
[15] R. Pekrun, “The Impact of Emotions on Learning and Achievement: Towards a Theory of Cognitive/Motivational Mediators,” *Appl. Psychol.*, vol. 41, no. 4, pp. 359–376, 1992.
[16] S. W. Mcquiggan, J. L. Robison, and J. C. Lester, *Affective Transitions in Narrative-Centered Learning Environments*. 2009.
[17] K. Forbes-riley, “Predicting emotion in spoken dialogue from multiple knowledge sources,” in *In Proceeedings of HLTNAACL 2004*, 2004, pp. 201–208.
[18] M. R. Lepper and M. Woolverton, “The wisdom of practice: Lessons learned from the study of highly effective tutors,” in *Improving academic achievement: Impact of psychological factors on education*, San Diego, CA, US: Academic Press, 2002, pp. 135–158.
[19] H. Prendinger, S. Mayer, J. Mori, and M. Ishizuka, “Persona Effect Revisited - Using Bio-signals to Measure and Reflect the Impact of Character-based Interfaces,” 2003, pp. 283–291.
[20] J. A. Healey, “Wearable and automotive systems for affect recognition from physiology,” Massachusetts Institute of Technology, 2000.
[21] C. Conati, “Probabilistic assessment of user’s emotions in educational games,” *Appl. Artif. Intell.*, vol. 16, pp. 555–575, 2002.
[22] S. D’Mello, P. Chipman, and A. C. Graesser, “Posture as a predictor of learner’s affective engagement,” in *Proceedings of the 29th Annual Cognitive Science Society*, 2007, pp. 571–576.
[23] C. Rich, B. Ponsleur, A. Holroyd, and C. L. Sidner, “Recognizing engagement in human-robot interaction,” New York, NY, USA, 2010, pp. 375–382.
[24] D. Szafir and B. Mutlu, “Pay attention!: designing adaptive agents that monitor and improve user engagement,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2012, pp. 11–20.
[25] M. P. Michalowski, S. Sabanovic, and R. Simmons, “A spatial model of engagement for a social robot,” in *9th IEEE International Workshop on Advanced Motion Control, 2006*, 2006, pp. 762–767.
[26] G. Hoffman, “Anticipation in human-robot interaction,” in *AAAI Spring Symposium Series*, 2010, p. 33.
[27] J. D. Finn, “Withdrawing From School,” *Rev. Educ. Res.*, vol. 59, no. 2, pp. 117–142, Jun. 1989.
[28] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, *Boredom, Engagement and Anxiety as Indicators for Adaptation to Difficulty in Games*. .
[29] A. Kapoor and R. W. Picard, “Multimodal affect recognition in learning environments,” in *Proceedings of the 13th annual ACM international conference on Multimedia*, New York, NY, USA, 2005, pp. 677–682.
[30] D. Zyngier, “(Re)conceptualising student engagement: Doing education not doing time,” *Teach. Teach. Educ.*, vol. 24, no. 7, pp. 1765–1776, Oct. 2008.
[31] A. L. Thomaz and C. Breazeal, “Experiments in socially guided exploration: lessons learned in building robots that learn with and without human teachers,” *Connect. Sci.*, vol. 20, no. 2–3, pp. 91–110, Sep. 2008.
[32] G. Castellano, I. Leite, A. Pereira, C. Martinho, A. Paiva, and P. W. McOwan, “Detecting Engagement in HRI: An Exploration of Social and Task-Based Context,” in *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, 2012, pp. 421–428.
[33] W. Burleson, “Affective Learning Companions: Strategies for Empathetic Agents with Real-Time Multimodal Affective Sensing to Foster Meta-Cognitive and Meta-Affective Approaches to Learning, Motivation, and Perseverance,” Massachusetts Institute of Technology, 2006.
[34] C. Conati and H. Maclaren, “Empirically building and evaluating a probabilistic model of user affect,” in *User Modeling and User-Adapted Interaction*, 2009.
[35] S. Robison, S. W. McQuiggan, and J. C. Lester, “Evaluating the Consequences of Affective Feedback in Intelligent Tutoring Systems,” in *International Conference on Affective Computing and Intelligent Interaction (ACII-2009)*, Amsterdam, The Netherlands, The Netherlands, 2009.
[36] J. C. Lester, S. W. McQuiggan, and J. L. Sabourin, “Affect Recognition and Expression in Narrative-Centered Learning Environments,” in *New Perspectives on Affect and Learning Technologies*, R. A. Calvo and S. K. D’Mello, Eds. Springer New York, 2011, pp. 85–96.
[37] S. Bull and J. Kay, “Student Models that Invite the Learner In: The SMILI:() Open Learner Modelling Framework,” *Int J Artif Intell Ed*, vol. 17, no. 2, pp. 89–120, Apr. 2007.
[38] S. Loyens, J. Magda, and R. Rikers, “Self-Directed Learning in Problem-Based Learning and its Relationships with Self-Regulated Learning,” in *Educational Psychology Review*, 2008, pp. 411–427.
[39] W. Joolingen, “Cognitive tools for discovery learning,” *Int. J. Artif. Intell. Educ.*, pp. 385–397, 1999.