The 8th International Scientific Conference
eLearning and software for Education
Bucharest, April 26-27, 2012
10.5682/2066-026X-12-098

ONE SIZE DOES NOT FIT ALL!
A THEORETICAL MODEL FOR ADAPTIVE TUTORING SYSTEMS

Silviu APOSTOL, Alexandru SOFRONIA
InSoft Development & Consulting, Sos. Virtutii 19D, sector 6, Bucharest, Romania
E-mail: silviua@insoft-dc.ro, alexandrus@insoft-dc.ro

Dan MIHAILESCU, Livia PETRESCU
Department of Anatomy, Animal physiology and Biophysics, University of Bucharest,
Splaiul Independentei, Nr.91-95, Sector 5, Bucharest, Romania
E-mail: dan.mihailescu@bio.unibuc.ro, draghicilivia@yahoo.com

Abstract*: One size does not fit all! This concept is relevant in many fields, and it constitutes a cornerstone in instruction and learning, where individual differences account for a lot of learning outcomes' variance [1].*
Considering this and other factors, we have developed a theoretical concept for a non-intrusive intelligent tutoring system that uses student facial and vocal features to customize the appearance and behavior of a virtual tutoring agent – a user's automated companion in computer-based learning environments.
It is well known that computer-based learning cannot be built entirely on legacy learning paradigms; the challenges that need to be tackled in computer-based learning require new and innovative theories and educational methods. Using the latest ITC technologies and frameworks, computer-based learning methods and educational content could be adapted to suit students' real needs while providing access to the most relevant knowledge in the most efficient way.
Educators at all levels need to address the diversity in their students' learning styles, the broad spectrum of individual abilities, and the diversity of socioeconomic and cultural backgrounds. For a while now, teachers have been called on to tailor educational activities for an increasingly heterogeneous student population [2] and there are a lot of success stories that can be discussed. Can we achieve the same flexibility in non-human instruction delivery methods? Can automated learning systems adapt and react to learners' preferences the same way excellent teachers do in real-life scenarios?
We believe so. We believe that using artificial intelligence and affective computing research results, in the near future computers will possess such abilities. What we propose is one step forward in this direction: providing e-Learning platforms with intelligent features that can read and interpret the students' state and preferences, and adapt their behavior, in real-time, to respond to those preferences. This way, the over-all learning experience becomes more natural, engaging and immersive; thus, users will choose to spend more time learning and also learn more efficiently.
This concept tries to replicate at a machine level, our subconscious human ability to constantly analyze faces, voices, gestures and nonverbal expressions of our peers in order to interpret emotions, satisfactions and dissatisfactions, attention, feelings.
We propose a theoretical model for this type of artificial intelligence used in customizing the delivery of the instruction. Details relating to user models, intelligent tutor modelling and algorithms for adjusting the tutor model are provided. This approach builds extensively on breakthroughs in intelligent tutoring systems, artificial intelligence, affective computing and evolutionary based studies of human communication. It goes beyond the state-of-the-art by adapting and applying the theoretical research and open-source multimodal emotion recognition frameworks to e-Learning environments.

Keywords: *intelligent tutoring systems, adaptive systems, online learning, artificial intelligence, learner preferences*

I. CHAPTER 1 – INTELLIGENT TUTORING SYSTEMS: OVERVIEW

Intelligent tutoring systems (ITS) or Adaptive tutoring systems as computer systems that provide direct personalized instruction to learners are not a novelty. This concept was proposed well before appropriate software and hardware technologies were available. These systems fall within the artificial intelligence (AI) field, aiming to model, simulate and ultimately replace all the aspects of a human tutor, similar to expert systems. Early ITSs have been implemented since 1970 [3], [4], but considerable progress was made only when the technologies caught up (early 90s).

ITSs present many ideal features: they give the students their own private teacher capable of offering them the learning they need [5] when they need it. ITSs learn to adjust to the user's knowledge level and expectations, and they do it without needing to use external resources. The explicit assumption is that one-on-one tutoring is the best way to learn [6]. And they really help the students to achieve higher scores as some studies show [7], [8].

ITSs aim to match the individual needs of learning by providing different routes to digital content, different interaction means and different content structuring. The revolution in cognitive sciences was integrated in the development of ITSs, not only by considering the logical and rational parts of the learners' brain, but also by taking into account their emotions and motivation, gender differences, health and special needs of students and, as further presented, even the evolutionary history of the species.

ITS and AI fields are not only using advances in research from a number of different fields, but also inform and extend the scope of those fields providing specific research issues, hypothesis and questions (Self, 1988):

- How can we represent and structure the knowledge?
- How can each individual be helped to learn?
- Which strategies of teaching are usually used by human tutors, and how can these be replicated in ITS?
- What is the intuitive knowledge of learners about the different domains of knowledge and how can these be used to enhance learning?

ITS's scope is to produce a "teacher for every student" or even more than one teacher per student. Thus ITS-based learning is not far away from another ideal - making learning a social activity.

Results include the usage of novel ways of interacting with computers, using speech, facial expressions and body language. From a human user's point of view, these are the usual ways of interacting with other humans. Thus computers are learning to behave more like humans, understanding their natural ways of behaving. A framework for this endeavor has been proposed and there are some notable results in this direction [9], [10].

Using virtual human models as agents of ITS, computers are also "speaking" the human language! This includes speech, facial expression, voice modulation and body language. Virtual teachers use such communication means to implement a wide range of teaching strategies such as collaboration, inquiry, and dialogue.

Here are some examples of features implemented in ITSs:

- Learning languages implies pronunciations skills, so that the students can speak into a microphone and ask the virtual tutor to give them feedback. The system can withhold the feedback until the end of the speaking session in order not to interrupt the student [11].
- First-year medical students are learning how the barometric (blood pressure) response works. Their conversation with a computer tutor does not involve a microphone or avatar, yet they discuss the qualitative analysis of a cardio-physiological feedback system and the tutor understands their short answers [12].

The analysis of facial and body dynamics, gestures, verbal and nonverbal expressions is used unconsciously by humans to interpret emotions, satisfactions / dissatisfactions, attention, feelings. One of the goals of AI and affective computing research is to endow the computers with such abilities.

Recently the CALLAS FP6 project [13] has registered a great leap in the field of intelligent systems, results that are re-usable in education to a great extent. The CALLAS consortium has developed means of affective computing i.e. automated interpretation of human emotions via gestural and other body language means. We list some of the Callas technologies relevant to the model proposed in this paper:

- Real-time emotion recognition from speech: a framework for building an emotion classifier and for recognizing emotions in real-time. It extracts from speech signals a vector of emotion-relevant acoustic features (e.g. derived from pitch, energy, voice quality, pauses, and spectral information) and then it uses a statistical classifier, trained by examples, to assign emotion labels.
- Video Feature Extraction: extracting faces from video sequence or live camera feed to derive information about the emotional state, content and context, keeping track of the amount of people looking towards the camera, and deriving interesting cues about the state of the audience's interest. The component consists of a video player that can play video files and capture live feed from a camera.
- Video-Based Gesture Expressivity Features Extraction: a video-based component detecting and tracking the user's hands to extract and transmit expressivity features' values such as overall activation, spatial extent, temporal, fluidity, power.
- Gaze detection and Head Pose estimation: a component estimating human head movements (yaw, pitch, roll), and direction of the eyes related to a user position in front of the content displayed on a computer monitor, deriving information about his state: attentive, distracted or nervous.
- Facial feature detection: detecting and tracking different facial features (such as eye centers, eye corners, top-down eyelids) based on facial geometry and prototypes of natural human motion.
- Facial Expression Recognition: recognizing facial expression in real-time by localizing and tracking facial features/movements, based on the appearance of the expression of a person when interacting with a camera, and providing feedback regarding emotion recognition based on dimensional or Ekmanian emotions.

Companies such as ISN Virtual Worlds have been working intensively on efficient implementation of intelligent avatars for training, marketing and customer care applications. AVATARS applications enhanced with an artificial intelligence module obtained a lot of interesting results. The AI modules for the AVATARS can be easily integrated with any existing knowledge base. By linking a company’s knowledge structured repository to the AI module of the AVATARS, the artificial agents are able to interact with people and interpret their actions or determine the depth of the information required by a given user. This includes managing access privileges and recognizing different users, guiding them to the information they need. Intelligent AVATARS can be integrated in different platforms, including online 3D virtual worlds.

Despite all the developments made in the last decade, ITSs are still are not used as widely as envisioned. There are a lot of causes behind this, including the high costs of such systems. A lack of personalization of the pedagogical agents also holds back the usage of ITSs. Learners are not yet provided with a trustworthy natural behavior and appearance of tutoring agents.

II. CHAPTER 2 – CONSORTIUM BACKGROUND AND ACHIEVEMENTS

InSoft Development&Consulting, founded in 2008, has developed a wide range of ICT solutions for the educational and other sectors. Among the wide range of software solutions in the educational and business sectors, *InSoft D&C* has developed a proprietary eLearning suite – called *learnIN*. The suite includes integrated eLearning applications: a complete eLearning platform (LMS/LCMS) and digital educational content (eContent). The LCMS *LearnIN* platform offers multiple facilities for teaching and learning, can be accessed from various IT devices, has a flexible educational content management system, that doesn’t require advanced technical knowledge and it is compliant with standards (such as SCORM 1.4, SCORM 1.2, AICC, and QTI).

InSoft D&C has also been developing elearning content using both 2D and 3D technologies and now has an extensive library of SCORM compatible learning objects. INSOFT uses specialists in each knowledge domain (subject matter experts) together with content developers that use cutting-edge software technologies in order to emphasize the important aspects in each and every educational field. For example, accurate 2D and 3D numerical simulations are a part of the portfolio that INSOFT has developed and added valuable experience to the entire team.

Most of the e-content is build using Flash, Java and HTML-based technologies. Some of the content represents simulations of real world and graphical representations. Developing simulations in virtual 3D environments helped by a virtual tutor can increase the immersiveness of the user and, hence, improve the educational progress.

The Department of Anatomy, Physiology and Biophysics, Faculty of Biology (DAFAB) is a leading organization within the University of Bucharest - a department specialized in high level education and research. DAFAB has expertise in:

- Teaching methodology, experienced over many years of working with students;
- Creating educational content in different areas of biology: anatomy and physiology, neurobiology, cell biology, anthropology, ethology, biophysics, molecular modeling.
- Ethological science, with results in processing biometric information (facial and vocal).
- Using computer and processing of statistical data. One of department's main goals is to develop and efficiently analyze data structures and algorithms, tools for visualization and communication, in order to integrate a large amount of information and create models of biological systems, both at macro and micro-structural level. This approach involves the use of computer simulations for analysis and visualization of complex connections of cellular processes.
- Collection and processing of visual information: patch clamp, modeling of systems.

III. CHAPTER 3 - THE PROPOSED THEORETICAL FRAMEWORK

Existing software systems such as CALLAS use adaptive 3D models and AI that can sense emotional states, attention, and personality and/or can dynamically adjust the behavior of the systems. Our model adds the use of existing theoretical research in multimodal emotion recognition and open-source frameworks applied to the educational field. Using learners' immutable features to synchronize and provide an animated interactive tutor, our model aims to increase the user satisfaction and the efficiency of learning process.

Thus we propose a non-intrusive ITS: by using student facial and vocal features to customize a virtual tutoring agent, we provide the students a better fitted user's companion in the eLearning environment. Our paradigm is to make the interactive learning experience more natural, engaging and immersive, incorporating user needs to ensure that content delivery is fitted for each learner. Thus, users choose to spend more time learning and also learn more efficiently using the personalized learning environment.

Our proposal of the model is based on several key ideas:

- *Diversity:* Learners always benefit from personalized learning experiences. There is no "*one size fits all*" principle valid for learning. People have different ways of learning, different preferences and experiences, and by individually addressing their needs, learning efficiency can be increased.
- *Unity:* There are universals regarding the way users behave and those universals have the same significance across all populations and ages. People communicate more reliable and predictable their inner states through non-verbal features compared with verbal communication. The patterns of non-verbal communications could be used to assess the inner states of learners, states that are connected with their learning behavior (e.g. presence, attention, interest).
- The evolutionary history of humans as social beings shaped our mind to expect agents in every element around. *Humans respond differently* to other humans based on their *appearance*. People respond to facial resemblance by increasing pro-social behavior and positive attributions toward self-resembling images. This is consistent with evolutionary theories, such as natural selection theory [14], inclusive fitness [15] and kin selection theory [16]. The same is valid for voice resemblance. Those two channels of communication are unconsciously used by people to judge the resemblance with others. Recent studies about facial resemblance indicate that humans act much more trustful when exposed to human faces resembling themselves [17]. *Trustworthiness* is a relevant factor for the

learner-teacher relationship. Trust has been shown to increase cooperation, improve flexibility, lowering the cost of coordinating activities, and increasing the level of knowledge transfer and potential for learning [18].

- *Animated virtual pedagogical agents* increase the efficiency of learning. Human generally enjoy other humans' presence and interaction helps in making the computerized learning resources more efficient.

Our proposed theoretical model aims to increase the efficiency of online learning by using information gathered from the learners. It interprets and feeds it to the ITS to adjust the appearance and behavior of the tutoring agent.

The components of the proposed theoretical model are:

- The user model (UM), which contains information about the learner. It is filled with data via user interaction with the system (whether the user merely types information into a keyboard or uses other audio/video channels of interaction). The UM is permanently updated as the user interacts with the system. It also represents the knowledge of the learner as it stores and updates the knowledge of the student, finds the errors and their causes and indicates possible routes to improve [19]. Inexpensive commercially available web cams and microphones are used to gather information about users. The software component responsible for the UM uses input from the user via audio and video sensors to interprets gestures, postures and other nonverbal features of the learner's behavior.
- The agent model (AM), which contains information about the virtual pedagogical agent: appearance (facial and vocal features), behavior, tutoring strategies etc. The AM is fed with data from the UM. The information from AM it is used to construct a virtual agent at runtime.
- The content model (CM), which contains information about the learning domain. It is constructed by modeling the expert knowledge as conceptual maps of the information prior to the system running. The virtual tutor is integrated as a dynamic part of the content model (CM).

We present a typical scenario using this theoretical model: Learners start a session in their asynchronous eLearning environment using a typical personal computer with a web cam and a microphone. The system gathers relevant information non-intrusively and without distracting the users' attention. The users have an initial interaction with the ITS, where they are informed that their personal data (audio and video) will be used to customize the learning environment. After the user gives his/her consent via a form, a user modelling module, using Audio & Video feed from the web camera and microphone, generates a custom UM. Features from this UM are used to construct and update the corresponding AM: the facial and vocal characteristics of the animated pedagogical agent are generated in this step. Learners can then proceed to access the educational resources, continuing their learning paths. In the eLearning environment, the presentation of content within each learning object is done using the agent.

In such a scenario, the virtual agent will be able to assess student's emotion and respond appropriately. During each interaction, the virtual tutoring agent updates its model of presumed student knowledge and current misconceptions.

SCORM is a de facto standard of elearning used in many academic, military and corporate environments. The model proposed could be implemented in a SCORM compliant LMS. This will imply that SCORM features needs to be expanded to fit this model and thus make it better suited to students' needs and preferences.

IV. CONCLUSIONS

In this paper we presented the Intelligent Tutoring Systems state of the art considering virtual agents as deliverers of tutoring.

As an expansion of the ITS capabilities and benefits we proposed a theoretical model that takes into consideration the natural communication and interaction between humans, endowing the pedagogical virtual agents with better sensing and adapting properties.

Asynchronous learning users, such as students and employees, will benefit from runtime-tailored content in the learning-friendly environment proposed by the model, which will lead to an increase in the users' cognitive gain from eLearning experiences.

Besides the educational market, the innovative use of the ITS with virtual customized assistant might be used in interaction with digital content in areas such as flight training, explosive detection and destruction training, specialized military training, etc.

Clearly further research and development is needed in order to implement the theoretical model, so we would like to extend an invitation to conference participants to send us their feedback and proposals, for improving or augmenting the theoretical model and / or contributing with specific expertise to prototype development.

References

[1] Jonassen, David H. & Grabowski, Barbara L. 1993. Handbook of Individual Difference, Learning, and Instruction. Lawrence Erlbaum Associates, Publishers, Hillsdale, NJ

[2] Self, J.A. 1974. Student models in computer-aided instruction, International Journal of Man-Machine Studies, 6, 261-276.

[3] Carbonell, J. R. 1970. AI in CAI: an artificial intelligence approach to computer-assisted instruction, IEEE Transactions on Man-Machine Systems, II, 190-202.

[4] McArthur, D., &: Lewis, M. 1998. Untangling the Web: Applications of the Internet and other information technologies to higher learning. Santa Monica, GA: Rand.

[5] Bloom, B. 1984. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher 13.

[6] Koedinger, K. R., Anderson, J.R., Hadley, W.H., & Mark, M. A., 1997. Intelligent tutoring goes to school in the big city. International Journal of Artificial Intelligence in Education, 8, 30-43.

[7] VanLehn, K.; Graesser, A.; Jackson, G. T.; Jordan, P.; Olney, A.; and Rose, C. P. 2005. When is reading just as effective as one-on-one interactive human tutoring? In Proceedings of CogSci2005.

[8] Gunes, H., M. Piccardi, et al. 2004. Face and body gesture recognition for a vision-based multimodal analyzer, Australian Computer Society, Inc.

[9] Tian, Y., T. Kanade, et al. 2011. Facial Expression Recognition. In Stan Z. L., Anil K. J. *Handbook of Face Recognition*, pg. 487-519, Springer.

[10] Mostow, J., Beck, J., Winter, S. V., Wang, S., and Tobin, B., Predicting oral reading miscues. In *Proceedings of the Seventh International Conference on Spoken Language Processing*, ICSLP'02, Denver, USA. p. 1221–1224.

[11] Freedman, R., Evens., M.W., 1997. The Use of Multiple Knowledge Types in an Intelligent Tutoring System. *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society, Stanford.*

[12] CALLAS, http://www.callas-newmedia.eu

[13] Darwin, C. 1859, On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life, London: John Murray.

[14] Hamilton, W. 1964. The genetical evolution of social behaviour. I. Journal of theoretical biology 7 (1): 1–16

[15] Smith, J. M. 1964. Group Selection and Kin Selection. Nature 201 (4924): 1145–1147

[16] DeBruine. L.M. 2002. Facial resemblance enhances trust. Proceedings of the Royal Society of London B, 269(1498): 1307-1312

[17] Nielsen, B.B. 2004 The Role of Trust in Collaborative Relationships: A Multi-Dimensional Approach, M@n@gement 3/2004 (Vol. 7), p. 239-256

[18] Woolf, B. P. 2008. Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning, Morgan Kaufmann.

[19] Bransford, J.D., Brown, A.L., Cocking, R.R., 2000. How People Learn: Brain, Mind, Experience, and School. Expanded Edition, National Academy Press, Washington, D. C.