![DiVA logo](http://www.diva-portal.org)
This is the published version of a paper presented at *2nd International IBM Symposium on Human Factors, Software, and Systems Engineering, 26–30 July 2015, Las Vegas, United States.*

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-109205

1. Introduction

In teaching and learning process, the role of teacher is undeniably vital and his/her presence stimulates the social, emotional and cognitive interactions especially when it comes to cross-cultural communication and education [5, 31]. The researchers agreed that in cross-cultural education, face-to-face interaction and communication deliver the best experience in developing skills, knowledge and competence [3, 4]. But, the recent technological advancements have given birth to online or distance education setting where text, audio and video forms of interactions are mostly used. These forms of distance education setting have greatly benefited the education process [6]; however when it comes to pragmatics, prosody, and non-verbal behavior based communication, which are considered important components of emotional and cognitive presence of teacher, are not being fully addressed using internet based text, audio and video communication [2]. The emotional and cognitive based embodied presence of a teacher not only assists with planning, conducting, intervening in teaching process but also improves the student's ability to relate and create meaning from the taught contents [7,8]. It has been reported that teacher's behavioral cues such as facial expression, eye contact, proximities, direction of attention, postures and gestures, have significant impact on student performance and improves teacher's influential role [9,10,12].

The most common computer mediated distance education is based on standard audio-video conferencing system. These video conferencing systems based distance educational setting are considered economical and motivational for students but are restrictive when it comes to non-verbal communication [1,2,32]. In face-to-face learning, head gesture (and eye contact) conveys a serious communicative messages especially where speech is ambiguous, hard to hear and/or understand [11,13]. Recently, researchers have proposed various robotic-agent based solutions to narrow the gap between face-to-face learning and computer based distance learning. These robotic-agents based distance learning is also termed as robot-mediated learning. These robotic-agents improve the instructional effectiveness while presenting an embodied presence of a teacher in distance educational setting (see [19]). It is reported that this inclusion can provide an effective blend of human-behavioral cues and computer-mediated standard video-conferencing systems. This embodied presence increases with social behaviour of robotic-agent which is directly related to its movements [28]. Human head movement is very important in general conversation and in-class communication, tutor's head movements repersent the contextual information as well as an integral part of a deictic gestures [26] . Despite the influential role of the head gestures in teaching and learning, very little research has examined gesture's role in the robot-mediated learning process.

In this work, we present experimental studies that investigate the role of embodied head gesture (and eye contact) developed for a robotic agent in distance learning scenarios. For these experimental studies, we have used our embodied telepresence system (ETS) which represent the head gesture of a human tutor in distance education setting. We provide a comparison study on learning experience using ETS and standard video conference system. Through such scenarios we want to test our hypothesis that '*robotic agent with expressive head gesture (i.e., mimic the human head movement) can improve the students learning process and hence have positive impact on their performance*'.

2. Background

The massive growth in communication technologies, in last two decades, has impacted on distance education and training, which is expressively different from the traditional classroom interactions. It is estimated that more than 6.7 million students are registered in online distance education [14]. During in-class interaction, people exchange non-verbal cues, facial expressions, gaze direction, bodily gestures, and tone of the voice to create presence and perform various interaction patterns to aid information transmission. It is believed that nonverbal cues accomplish two distinct purposes 1) a direct passage of information from one person to another; 2) the 'integrational aspects' of the communication process [18]. The 'integrational aspects' contains all the non-verbal physical manifestations of information exchange that regulate the interaction process and keep the conversation going and providing semantic meaning as well as relation to larger contexts. The most dominant form of computer mediated communication is based on the standard video conferencing software which is often described as a medium that is limiting in non-verbal cues and social context [15,16,17]. Considering these limittaions, researchers are proposing the robotic and

animated agents which would be assiting the remote students and may offer computer medicated education with flavour of human tutor's beahvioural cues.

Recently, Robotic-agents, also known as educational assistive robots (EAR), have been used to interact with students in order to help them develop educational skills both in distance education and in-class setting [19, 20]. However, the effectiveness of these robot agents is not only measered by the task performance but also its social behaviour which is perceived and understood by its interacting parnters [21, 22]. Hence, the perception of social behaviour of a robotic-agnet is directly related to its movements which also repersents the contextual information. Therefore, the design and modeling of robot-agent's mechincal part movements (eyes, head, arm, etc.) must be approached carefully. When it comes to educational assistive robots in distance education, it is highly desirable that their head movement (and hence eye contact) must 'look like' or 'similar' to human tutor head-gesture; as human gestures which are considered very cruial in social interactions as well as teaching and learning settings [9-11]. In-class setting, tutor's head movement communicates levels of gaze, physical proximity, and other behaviors indicative of interactivity. This feedback communicates relevant information to synchronize rhythm between participants as well as provides the contextual information [23-27].

In this work, we study the role of head movements of educational assistive robots due to its significance in face-to-face settings. For these experiments, we have used our robotic agent namely; Embodied Tele-presence System (ETS). ETS is a three degree-of-freedom tele-robot which mimics the human neck movements; i.e., ETS embodies the remote participant head-movements to give more levelheaded presence and can be perceived by his collaborators as being equally present by gaze direction and by head embodiment. For more details about ETS design and functionality see [28].

2.1. Hypothesis

For the following studies, we want to see how the head-gesture will affect the focus of attention and improve the learning experience of remote participants for distance education. We felt that the remote participants will be more interested-in and adaptively engaged by embodied head gesture based interactions as compared to standard video conferencing based distance education setting. Therefore, we consider the following two hypothesis;

H1*: Participants will have positive reaction towards ETS, which will increase over time.*

H2*: Participants engagement with ETS in real time will improve participants' attention and involvement hence motivate them toward learning.*
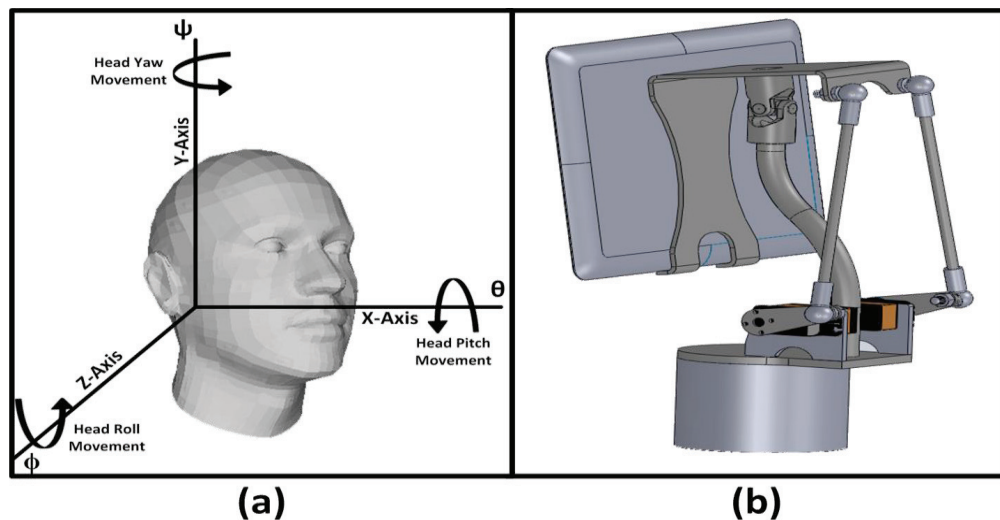

Fig. 1. a) Human head orientation modelling used for head pose estimation and designing a 3 DOF neck/head robot; (b) Embodied Telepresence System (ETS)- our educational assistive robot.

3. Embodied telepresence system (ETS): Modeling head gesture

The simulation of head gesture through ETS consists of two modules; software module and hardware module. In software module, the pose of the human head is estimated with a constraint that the human head is a 3 DOF rigid object which has yaw, pitch and roll movements as shown in Fig. 1(a). We have used our geometric head pose estimation (GHPE) algorithm which estimates the head pose through a standard webcam of the computer. The implementation of GHPE algorithm is done in VC++ and the details of this algorithm can be found in [29].

The hardware module of ETS consists of a 3 DOF neck/head robot for exhibiting the real head movement. We named this robot as Embodied Telepresence System (ETS) as shown in Fig. 1(b). The design of ETS consists of three servo motors attached in a configuration to give all three degrees of head motion (yaw, pitch and roll). The tablet PC is used to present the audio and video of the person.

We have used ETS in a distance education scenario where tutor's head gestures are presented to the students through the combination of software and hardware modules of the ETS. The deployment of the ETS system in real distance education scenario is explained in section 3.1.

3.1. System deployment

The ETS system is deployed for a distance education setting where we have two sites: a student site and a tutor site as shown in Fig.2. The left column in Fig.2 shows the tutor site and the right column shows the student site. At tutor site, we have two computer screens; where one screen displays the lecture slides and the other shows a real-time video stream of the student. Here at tutor site, a webcam based GHPE algorithm is installed to estimate the head gesture of the tutor. At student site, we have an ETS which could be used to present the audio-video and head gesture of tutor during lecture. The similar lecture slides are also being presented on the computer screen of the student and is controlled by the tutor. The real time deployment of the system consists of following steps:

- Audio-Video communication is setup between the tutor and student site through Video-over-IP software (i.e. Skype).
- Wireless data communication is done through Xbee wireless transceiver.
- The GHPE algorithm is used to calculate the Yaw, Pitch and Roll angles of the tutor.
- The pitch and roll angles are mapped directly to ETS to present the head gesture of the tutor at student site; where, the Yaw angle decides where the tutor is looking, i.e. is he looking at lecture slides or at the student video.
- Based on the provided angles to ETS-controller, it turns toward the lecture slides or toward the student at student site- 'showing' eye-contact and head gesture.
- ETS controller generates PWM signals to perform these yaw, pitch and roll movements.
- The whole system performs real time communication with 25 frames per second.
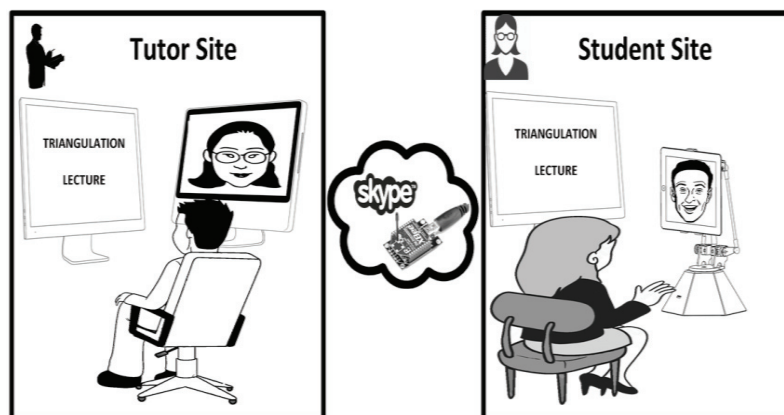

Fig. 2. Application scenario: one-to-one distance education setting; the left is a 'tutor site' and the right side depicts the remote student participant with educational assistive robot.
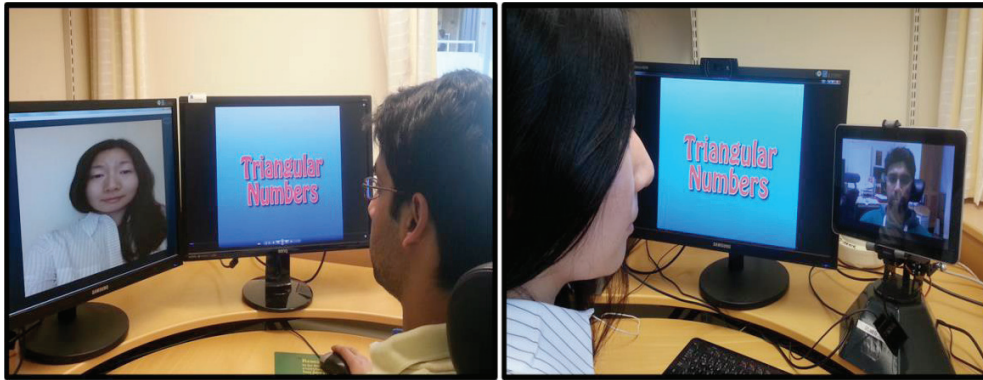
Fig. 3. Real experimental setting: one-to-one distance education setting; the left is a 'tutor site' with two screen and head gesture estimation algorithm installed and the right side depicts the remote student participant with one screen and a Embodied Telepresence (ETS).

4. Experimental studies

The goal of the experimental studies are to investigate the effectiveness of our novel distance education scenario where the distant located student is assisted by the tutor's head gestures, gaze and focus of attention. Furthermore, this study focuses on the effects of head gestures in distance education over time.

4.1. Participants and procedure

Ten students (5 boys and 5 girls) were recruited from the campus of Umeå University, Sweden, ranging in the age of 15 to 23. All the students were directly involved in distance education and they have the basic knowledge of mathematics. Furthermore, we have hired one mathematics teacher to deliver a lecture on *triangulations*. A training was given to the teacher to make him familiar with the ETS and head gestures based distance education setting. All the students plus the teacher were told about the purpose of the experiment. The experiment was setup according to the system deployment steps. The real experimental setup is shown in Fig. 3. The experiment was run 10 times for 10 students; where teacher delivers a lecture for 5 minutes on *triangulations* for each student. There are two scenarios for this experiment; one where teacher delivers a lecture through simple skype conversation for 2.5 minutes and the other where teacher delivers a lecture through ETS for the next 2.5 minutes. The order of these sessions was random; i.e., some students took skype lecture first and other ETS base lecture. At the end of the experiment there was informal interview and each student was given a questionnaire to fill in questions related to our hypothesis.

4.1. Questionnaire

Our modified subjective questionnaire was adopted from a previously developed questionnaire; i.e., studies related to distance education by [30]. We have selected the following most important questions which are directly related to our experiment. The questionnaire used Likert style 7-point rating system, which scales from 1 to 7. The value 1 represents strong disagreement (negative) and 7 represents strong agreement (positive). The questions in the questionnaire are:

1) Motivation toward learning: Does movement capability of the educational assistive robot (ETS) motivates you to learn more during online lecture?
2) Monitoring participants: Do you feel to be monitored by your tutor during lecture?
3) Trust: Does the gaze direction of tutor helps to build trust during lecture?
4) Understanding: Does ETS help you to understand more as compare to traditional online teaching method?

5) Disturbing: Do you feel disturbed by the ETS? (1 = disturbed and 7 = not disturbed at all)
6) Track: Can you keep a track of the lecture?
7) Welcome and Comfortable: Is ETS welcoming and comfortable during teaching?
8) Time: Do you forget the role of technology over time?
9) Physical Presence: Do you feel the physical presence of the tutor at your site?

All these questions are compared with the traditional online teaching method (i.e. through video-over-IP software).

5. Results and discussion

The questionnaire results were analyzed by calculating the means and standard deviations of the ETS vs Skype systems used in distance learning scenario. Table 1 shows the mean questionnaire response for all the questions on 7 point Likert scale. The results are graphically shown in Fig. 4, where red bars show the ETS based user-responses and the blue bars show the Skype based user-responses.

If we compare the standard deviation of ETS with Skype, we see a large variation for ETS as compared to Skype only setting. The less variation in case of Skype can be due to the fact that the students were 'used-to'/aware of the traditional distance learning methods and were previously using them. But it was the first time they experienced distance learning through ETS, hence large standard deviation as compared to Skype setting. However, if we compare the mean values of the questionnaire, it can be seen that ETS setting outperform Skype-only setting in every question except in *Disturbing*. The students found that the ETS based scenario *(μ= 4.5) is unusual* which can be more disturbing as compared to Skype-setting *(μ= 5.0)*. This was expected as movement based interaction are sometime distracting but with *time (μ= 5.9)* students forget the role of technology and showed the 'feeling' of *physical presence (μ= 6.3)* of the tutor when using ETS. The students felt to be *monitored (μ= 6.0)* as the ETS sometime moves toward the computer screen and sometime makes an eye contact with the student.

One of the student mentioned that '*now he feels the influence of the teacher during lecture….*' Because of *monitoring/influence* and *head gesture*, students felt *motivated* to learn and these non-verbal cues help them to *understand* better.

There was more excitement in students as they found this learning method is more fascinating. The *trust* factor through ETS learning was also high as compared to skype-only based communication. The students keep an equal *track* of the lecture with both system. Finally, the students found ETS *welcoming* and *comfortable* and they showed willingness to buy this product in informal interviews. Based on these results, it can be confidently said that our both hypothesis holds for the setting i.e., *H-1-- Participants have positive reaction towards ETS, which increases over time.* Furthermore, *H-2 – participants' engagement with ETS in real time enhances participants' attention and involvement hence motivate them toward learning.*

Table 1. Statistical Analysis of survey questionnaire.

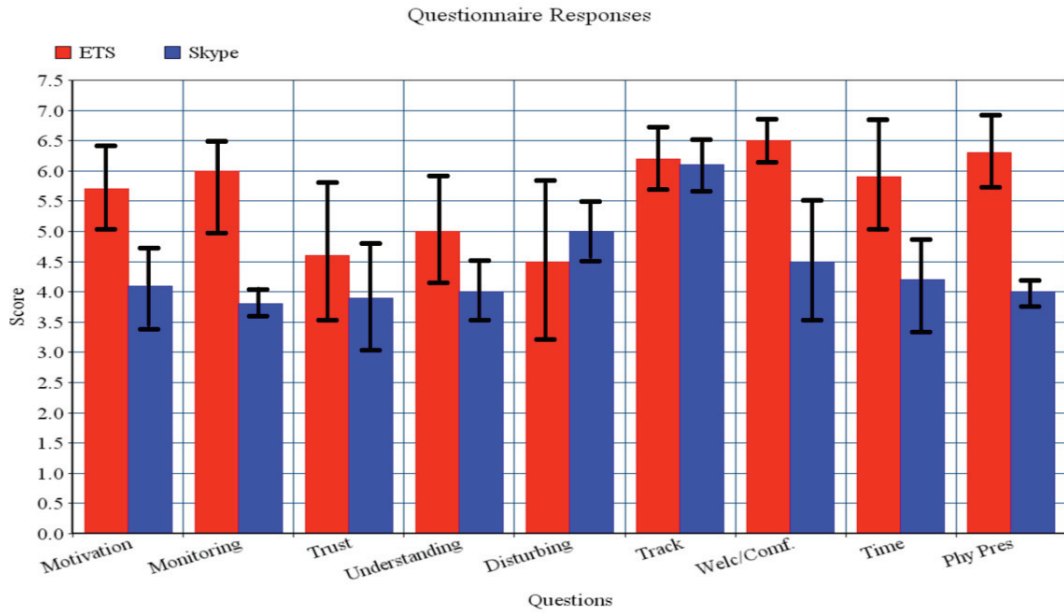| Questions	ETS		Skype	
	Mean (μ)	Std. dev. (σ)	Mean (μ)	Std. dev. (σ)
Motivation	5.7	0.7	4.1	0.6
Monitoring	6.0	0.5	3.8	0.2
Trust	4.6	1.1	3.9	0.9
Understanding	5.0	0.9	4.0	0.5
Disturbing	4.5	1.3	5.0	0.5
Track	6.2	0.5	6.1	0.4
Welcome/comfortable	6.5	0.3	4.5	1.0
Time	5.9	0.9	4.2	0.7
Physical Presence	6.3	0.6	4.0	0.2

Fig. 4. Mean Questionnaire Score.

6. Conclusion and future direction

Most of the traditional methods for online teaching are limited to standard audio-video and text based communication. These methods have limited ability to transmit certain nonverbal cues. In this paper, we have proposed a novel scenario for distance education by introducing educational assistive robot named embodied telepresence system (ETS). ETS is a physical representation of the tutor at student site, where head gestures of tutor are mapped to ETS. Furthermore, ETS imitates a shift in its focus of attention according to the tutor's focus of attention. Our user study shows the effectiveness of the proposed approach in distance education scenario. The results suggest that the nonverbal cues provide a vital feedback such as head nod, head shake etc. and these cues are useful supplement to audio-video communication. Our hypothesis upheld on the basis of experimental results i.e. *the robotic agent with expressive head gesture improves the learning process and have a positive impact on the student performance.*

The present study is subject to limitations that should be addressed in future research. The present study focused on a small number of university students in Sweden with little cultural diversity, therefore future research is required to determine whether the same results would be obtained in distance education environments with learners of different ages, gender, grade, intellectual level, and diverse cultural background. Similarly, future research is needed to determine whether the proposed ETS based online learning environment foster/affect the two way presence with positive affect on student learning outcomes.

References

[1] S. Yarosh, K. M. Inkpen and A. Brush, "Video playdate: toward free play across distance," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2010.
[2] 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, 2012.
[3] D. Batchelder and E. G. Warner, Beyond Experience: The Experiential Approach to Cross-Cultural Education., ERIC, 1977.
[4] K. Cushner and R. W. Brislin, Intercultural interactions: A practical guide, Sage publications, 1995.
[5] H. L. Dreyfus, On the internet: Thinking in Action, Routledge, 2008.

[6] M. Merryfield, "Like a veil: Cross-cultural experiential learning online," Contemporary issues in technology and teacher education, vol. 3, no. 2, pp. 146-171, 2003.

[7] C. F. Aust, "Face-to-Face Communication outside the Digital Realm to Foster Student Growth and Development," in Teaching, Learning, and the Net Generation: Concepts and Tools for Reaching Digital Learners, CH5, IGI Global, 2012, p. 74.

[8] D. M. Christophel, "The relationships among teacher immediacy behaviors, student motivation, and learning," Communication education, vol. 39, no. 4, pp. 323-340, 1990.

[9] W.-M. Roth, "Gestures: Their role in teaching and learning," Review of Educational Research, vol. 71, no. 3, pp. 365-392, 2001.

[10] J. R. Nelson and M. L. Roberts, "Ongoing reciprocal teacher-student interactions involving disruptive behaviors in general education classrooms," Journal of Emotional and Behavioral Disorders, vol. 8, no. 1, pp. 27-37, 2000.

[11] R. H. Thaler and C. R. Sunstein, Nudge, Yale University Press, 2008.

[12] J. A. Fredricks, P. C. Blumenfeld and A. H. Paris, "School engagement: Potential of the concept, state of the evidence," Review of educational research, vol. 74, no. 1, pp. 59-109, 2004.

[13] R. B. Church, S. Ayman-Nolley and S. Mahootian, "The role of gesture in bilingual education: Does gesture enhance learning?," International Journal of Bilingual Education and Bilingualism, pp. 303-319, 2004.

[14] I. E. Allen and J. Seaman, Changing Course: Ten Years of Tracking Online Education in the United States., ERIC, 2013.

[15] A. Feenberg, "The written world: On the theory and practice of computer conferencing," Mindweave: Communication, computers, and distance education, pp. 22-39, 1989.

[16] J. Wuther and J. Andersson, "Interpersonal Effects in Computer-Mediated Interaction," Communication Research, pp. 460-480, 1994.

[17] D. R. Garrison, T. Anderson and W. Archer, "Critical inquiry in a text-based environment: Computer conferencing in higher education," The internet and higher education, pp. 87-105, 1999.

[18] R. L. Birdwhistell, Kinesics and context: Essays on body motion communication, University of Pennsylvania press, 2010.

[19] L. Brown , R. Kerwin, and A.M. Howard, "Applying behavioral strategies for student engagement using a robotic educational agent," in IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2013.

[20] G. Gordon and C. Breazeal, "Bayesian Active Learning-based Robot Tutor for Children's Word-Reading Skills," in Proceedings of the 29th AAAI Conference on Artificial Intelligence, Austin, USA, 2015.

[21] M. Saerbeck and A. J. van Breemen, "Design guidelines and tools for creating believable motion for personal robots," in the 16th IEEE International Symposium on Robot and Human interactive Communication (RO-MAN 2007), 2007.

[22] M. Saerbeck and C. Bartneck, "Perception of affect elicited by robot motion," in The 5th ACM/IEEE international conference on Human-robot interaction, 2010.

[23] A. T. Dittmann and L. G. Llewellyn, "Relationship between vocalizations and head nods as listener responses," Journal of personality and social psychology, vol. 9, no. 1, p. 79, 1968.

[24] J. Lemke, "Meaning-making in the conversation: Head spinning, heart winning, and everything in between," Human Development, pp. 87-91, 1999.

[25] C. Pelachaud, V. Carofiglio, B. De Carolis, F. de Rosis and I. Poggi, "Embodied contextual agent in information delivering application," in the first international joint conference on Autonomous agents and multiagent systems: part 2, 2002.

[26] C. Goodwin, Conversational organization: Interaction between speakers and hearers, Academic Press, 1981.

[27] L.-P. Morency, C. Sidner, C. Lee and T. Darrell, "Contextual recognition of head gestures," in 7th international conference on Multimodal interfaces, 2005.

[28] M. Khan, H. Li and S. Rehman, "Embodied Tele-Presence System (ETS): Designing Tele-Presence for Video Teleconferencing," in HCI International 2014, 2014.

[29] M. Khan, S. Rehman, Z. Lu and H. Li, "Head Orientation Modeling : Geometric Head Pose Estimation using Monocular Camera," in Proceedings of the 1st IEEE/IIAE International Conference on Intelligent Systems and Image Processing, 2013.

[30] P. Youngblood, F. Trede and S. Di Corpo, "Facilitating online learning: A descriptive study," Distance Education, vol. 22, pp. 264--284, 2001.

[31] R. M. Lehman and S. C. Conceicao, Creating a sense of presence in online teaching: How to 'be there' for distance learners, John Wiley & Sons, 2010.

[32] M. Khan and S. Rehman, Distance Communication: Trends and Challanges and how to resolve them, HandBook: Strategies for a Creative Future with Computer Science, Quality Design and Communicability. Blue Herons, 2014.