Multimodal Assessment of Teaching Behavior in Immersive Rehearsal Environment - TeachLivE™

Roghayeh Barmaki
Computer Science Division
University of Central Florida
Orlando, FL, USA
barmaki@eecs.ucf.edu

ABSTRACT

Nonverbal behaviors such as facial expressions, eye contact, gestures, and body movements in general have strong impacts on the process of communicative interactions. Gestures play an important role in interpersonal communication in the classroom between student and teacher. To assist teachers with exhibiting open and positive nonverbal signals in their actual classroom, we have designed a multimodal teaching application with provisions for real-time feedback in coordination with our TeachLivE test-bed environment and its reflective application; ReflectLivE.

Individuals walk into this virtual environment and interact with five virtual students shown on a large screen display. The recent research study is designed to have two settings (7-minute long each). In each of the settings, the participants are provided lesson plans from which they teach. All the participants are asked to take part in both settings, with half receiving automated real-time feedback about their body poses in the first session (group 1) and the other half receiving such feedback in the second session (group 2).

Feedback is in the form of a visual indication each time the participant exhibits a closed stance. To create this automated feedback application, a closed posture corpus was collected and trained based on the existing TeachLivE teaching records. After each session, the participants take a post-questionnaire about their experience. We hypothesize that visual feedback improves positive body gestures for both groups during the feedback session, and that, for group 2, this persists into their second unaided session but, for group 1, improvements occur only during the second session.

Categories and Subject Descriptors

H.5.1. [Information Interfaces and Presentation (e.g. HCI)]: Multimedia Information Systems

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ICMI '15, November 09-13, 2015, Seattle, WA, USA

©2015 ACM. ISBN 978-1-4503-3912-4/15/11 ... \$15.00

DOI: <http://dx.doi.org/10.1145/2818346.2830604>

Keywords

Immersive rehearsal environment; Teacher preparation; Gesture recognition; Reflection; Multimodal Analytics

1. INTRODUCTION

Interpersonal communication includes verbal and nonverbal elements. Nonverbal communication (NVC) refers to all of the components of communication excluding the actual words used [19]. Para-language (pitch, volume, etc.) is the vocal element of nonverbal communication. On the other hand, body language (gestures and movements, facial expressions, eye contact, etc.) is the nonvocal element of nonverbal communication [14]. Upwards of 65 percent of the meaning derived in interactions comes from nonverbal communication [19, 10].

In the classroom and student-teacher communication, NVC also plays a critical role in successful interaction [20]. Teachers are required to comprehend the nonverbal cues they are sending and receiving in the classroom setting [6]. Teachers may use NVC in the classroom for different purposes. Alibali et al. [2] indicated that teachers' gestures have impact on student comprehension and student learning, specifically in instructional discourses. Gestures are defined as a form of nonverbal communication that associated visible bodily actions are used to communicate messages, either in place of speech or together and in parallel with spoken words [13].

The presented research investigates the nonverbal behaviors (body movements, especially gestures and postures) of participants in an immersive rehearsal environment - TeachLivE. The objective is to perceive the relations of participant's teaching experience with NVC feedback provision in simulated classroom. In addition, the participants are asked to take pre and post questionnaires during the experiment to determine their perception of the experience. The eventual goal of the research is to design a semi-automated annotation system from multimodal features including the body language and gesture tracking of participants in the TeachLivE environment that leads to assessment and reflection procedures for professional development. There have been one completed and one ongoing case studies related to the research topic that is covered in this paper. The methodology details and results of the first case study along with the design and preliminary results the second case study are described after the background section.

2. BACKGROUND

2.1 TeachLivE Rehearsal Environment

TeachLivE is a teaching and learning environment that was developed at the University of Central Florida to provide an interactive learning experience for teachers to rehearse and hone their classroom skills. It is an avatar-mediated system that consists of at least five student avatars in a virtual classroom setting. A human in the loop (interactor), orchestrates the behaviors of student avatars in response to a participant's actions. Figure 1 shows the architecture of the TeachLivE system [21] with some extended features that is explained in methodology section. TeachLivE has been widely used for training purposes. Over 12,000 users at 70 universities in the U.S, experienced the environment in 2014-15. Many of these universities have carried out research projects that use TeachlivE as the basic study environment. One of the largest of these studies was funded by the Bill & Melinda Gates Foundation. The results showed positive changes in teaching strategies after just four ten-minute sessions in the simulated classroom [9].

In the previous study [4], we used TeachAARS (TeachLivE After Action Review System) for manual annotation of nonverbal behaviors including proximity, body language (open versus closed postures) and head movement (nodding). Moving forward in our research, the nonverbal communication coding procedure is supported by the Microsoft Kinect V2 and 3D-sensor technology [1]. In addition, a new application, called ReflectLivE, is integrated into TeachLivE for multimodal data collection of participants' experiences.

2.2 Nonverbal Communication in Context

Communication scholars have compartmentalized nonverbal communication into nine categories: kinesics, more commonly referred to as body language and physical appearance; chronemics or communication through time; proxemics or communication using space; para-language, communication through tone of voice; artifacts, communication by physical objects; haptics, communication through touch; facial expressions; olfactics, communication by means of smells; and oculistics, more commonly referred to as eye contact [23]. The presented research focuses on kinesics and body language, especially gestures from defined NVC categories. McNeill defines gestures as four main types. Iconics are related to semantic content of speech; metaphors tie to an abstract concept; deictics or pointing gestures relate to the identity of individuals to whom the utterances refer; and beats keep the rhythm of speech [18].

NVC plays important role in various professions. In the case of physician-patient interaction, scholars presented results on the relation of nonverbal communication skill of the physicians and the patient satisfaction [7]. Also in the classroom environment, the successful student-teacher communication is affected by the NVC skills of the teacher [15, 20].

2.2.1 NVC and Immersive Environments

There exist an extensive amount of work on the impact and application of gestures in virtual and interactive environments. The context of deictic expressions and the impact of pointing gestures in virtual immersive environment have

been investigated by Pfeiffer et al. [22]. Two participants were asked to play an object identification game together and the session was recorded and annotated based on gaze finger pointing and index-finger pointing to analyze the success rate of object identification.

In another research [11], scholars introduced an automated system for social skills training called MACH (My Automated Conversation Coach). The system includes a virtual agent that reads facial expressions, speech, and prosody (tone, pitch) of the participant and responds with verbal and nonverbal behaviors in real-time. During the interview training session as an example application, MACH asks interview questions, automatically mimics certain behavior issued by the user, and exhibits appropriate nonverbal behaviors [11].

2.2.2 NVC and Classroom

Nonverbal communication strategies is as expressive as verbal communication skills and this has been also noted in teacher training [16]. The strategies like eye contact, prolonged gaze, and proximity can have positive or negative effects on student behavior and classroom management, depending on the situation and context [15, 6]. Teachers can use gesture to be effective in several fundamental aspects of their profession, including communication, assessment of student knowledge, and the ability to instill a profound understanding of abstract concepts in traditionally difficult domains such as teaching foreign languages and mathematics [12].

Alibali et al. showed that teachers use gestures in an effort to scaffold students' understanding, especially on abstract concepts. They worked on a mathematics lesson session that focused on algebraic relations [3]. They manually annotated the gestures and transcribed the teacher's conversation in the study and used the annotation data to analyze the nonverbal behavior of the teacher. In [5], we investigated the way teachers use gestures in the virtual classroom settings of TeachLivE. Our research had been developed based on the hypothesis reported in [3]. Analysis of video recordings from real and virtual classroom environments from biology and algebra teachers indicated that algebra teachers gesture significantly more often than biology teachers.

3. METHOD: GESTURE RECOGNITION

Manual annotation on video recordings by experts can provide accurate annotations; however it is extremely time consuming and sometimes costly. Hence, some automated techniques have been presented for gesture recognition from video clips and other resources. In general, most gesture recognition techniques have a set of training data to learn and extract the patterns of the gesture from the training set and then use the trained engine for recognition. In this research, we also use a similar technique with skeletal data in the Visual Gesture Builder (VGB) environment [1]. This approach is fast and inexpensive in comparison to the manual gesture annotation procedure.

VGB generates gesture databases that are used by applications to perform run-time gesture detection. VGB is defined as a data driven machine learning solution for gesture detection. Visual Gesture Builder uses skeleton tracking data (25

body points in Kinect V2) representing body movements, and AdaBoost ensemble classifier (for discrete gestures) or Random Forest Regression (for continuous gestures). At first, for each gesture, some clips from different subjects must be recorded with the Kinect Studio application from the Microsoft Kinect V2 Software Development Kit. Every clip contains an individual in front of the Kinect who exhibits the gesture (with some variations) 15 to 25 times. Then, the VGB tool is used to give meaning to the recorded clips. All the frames in the recordings that define a gesture, must be tagged manually. All the frames exhibiting the gesture get the label of 1 and the remaining frames in the clip get the label of -1 [1]. At this point, the problem is similar to a supervised binary classification, in which the classifier will recognize the frames in which a gesture is performed. For each gesture, the recognition engine have an associate's binary classifier. After the training step, VGB is able to generate the tags automatically in the analysis step from the test clips. In the following section, we discuss more details about the case studies and the targeted gestures to recognize according to interpersonal communication perspective.

3.1 Case Study 1

3.1.1 Participants and Study Setting

The experiment was a within-subjects study, fitting in a single session lasting at most 12 minutes (session range = 8-12 min., mean= 10 min.). The subjects were self-selected from a group of visiting UCF scholars composed of instructors and university students. Participants were given a consent form and a pre-questionnaire that were used to determine demographic information, teaching and educational background, English speaking level, and topic of interest for the teaching session. The participants (n=14, 6 female, 8 male) had an average age of 28.1 years and all were Hispanic/Latino with English as their second language. Two of the subjects majored in English Language Teaching (ELT). Six of 14 subjects had taught for at least five years. The other eight subjects were students with no teaching experience. Half of the participants selected Spanish Language as a teaching topic and the others picked different topics from science, marketing and algebra. All of the participants, except one, were novices in the TeachLivE environment. Each participant was introduced to the virtual classroom prior to the recording procedure and then were asked to start their teaching session.

3.1.2 Recording

Two types of data were recorded during the study. (I) skeleton data (IR, depth and body frame) and (II) virtual scene paired with the user experience from the observer station (video recording). Figure 1 presents the TeachLivE setting for the study and data collection specifications. The skeleton data of the participants were recorded using a Kinect V2 sensor and the Kinect Studio utility application. The recorded clips with Kinect Studio are required as data inputs to Visual Gesture Builder (VGB) to do automated tagging of gestures [1]. The second type of recorded data was the virtual scene paired with the user experience from the classroom in mp4 format. This was used for teaching proficiency analysis by experts.

Some of the closed postures from previous work [4] are shown in Figure 2. Caswell and Neill in their study on effects of

Figure 1: TeachLivE setting with some adjustments from the original architecture [21]. Two types of recorded streams are: (I) skeleton data (IR, depth and body frame) shown with orange arrow and (II) virtual scene paired with user experience recording from the observer station (video stream) with red arrow.

Figure 2: Reference example of closed postures/gestures; left to right: unreceptive (arms folded in front), seductive (hands clasped in back), skeptical (hands placed on hips), protective (hands clasped in lower front), pensive (thoughtful pose), and submissive (hands clasped in upper front).

body language of competent teachers showed that using reference postures as a communication medium, teachers employ open instead of closed poses in their classroom [6]. They also hypothesized in their study that closed poses indicate defensiveness and avoidance of the speaker/teacher to listener. A corpus of closed body poses (reference poses in figure 2 and some variations of them) was recorded from the performances of three male and two female students using the Microsoft Kinect V2. Subjects for the main study were not involved in the corpus collection. The corpus was used as training clips in the Visual Gesture Builder (VGB) application. All the clips were hand-coded and tagged in order to feed the automated analysis procedure of the participant's clips. There were six targeted closed gestures/poses to train/evaluate in this research. For training each pose, four 30-sec clips were manually annotated and added to the gesture database. 7 x 5 short clips were annotated; a thinking or pensive gesture is a sided-gesture, and we needed to record left and right versions of that gesture. Hence, we had seven classifiers rather than six at the end of training for the targeted six gestures. The process of creating the gesture recognition engine is shown in reference [1].

Figure 3: The average time of exhibiting closed gesture for two different groups in the study.

3.1.3 Performance Measures

The raw data from the user experience recordings (video clips) were reviewed by two supervisors to measure the teaching proficiency aspect and the communication aspect as an interested sub-domain of teaching evaluation framework. There are some teaching frameworks [8, 17] as evaluation instruments of teachers in real classrooms, but due to the specifications of TeachLivE as a virtual classroom, some domains may not be measurable. We created a teaching and communication proficiency checklist as a reference for observers with some adjustments to existing teaching evaluation frameworks. The checklist from the observers are available at the TeachLive web site (TeachlivE.org). Space limitations prevented them from being included here.

3.1.4 Results

All valid clips were added to the VGB solution created with different projects trained for each closed gesture. The automatically tagged data were reviewed to minimize the tagging error. For each clip of the participant, the length of exhibiting gesture(s) was reported as a percentage of the total session time. Figure 3 presents the average time of exhibiting six different closed gestures for teachers ($n=6$) and students ($n=8$) who participated in the case study.

This bar chart indicates that experienced teachers did not exhibit closed gestures as much as students who had no teaching experiences, considering the total time of having closed gesture (shown in the last two columns of figure 3). In addition, submissive gesture was employed frequently rather than other closed gestures by all participants. On the other hand, experienced teachers exhibited skeptical gesture more than students. As figure 3 shows, the results for all of the gestures, except for the seductive and submissive, were close to each other for both of the groups.

From a performance prospective, there was a significance positive correlation between some demographics data and teaching proficiency (TP) and communication proficiency (CP) scores reported by observers. For instance, there was a significant correlation for those in the ELT major, having teaching experience, and being older ($p < 0.01$ for all these variables) with TP score. For CP, age and English speaking level had significant correlations. Other demographics had weak correlations with TP and CP.

3.1.5 Discussion

In this pilot study, we hypothesized that there is a relation between teaching/communication proficiency and a subject's body movements/gestures during the interactive experiment. Among the detected gestures, submissive and seductive gestures were much more prevalent than other closed gestures. There might be cultural and personal influences for the prevalence of submissive and skeptical gestures among participants. In addition, since English was not the participants' first spoken language, they frequently used beat gestures to keep the rhythm of the speech [18]. There was significant correlation between a subject's teaching experience, age and TP/CP scores.

There may be multiple confounding factors involved in this study. All of the subjects except one were new to the simulated classroom and the experiment was single-session. All the participants were non-native English speakers with average level of English proficiency. The other limitation was the small number of participants and also having selective open topic for teaching sessions because the sample was from a group of convenience.

3.2 Case Study 2

3.2.1 Study Setting

In this case study, which is an ongoing project, a maximum of 50 student participants will be recruited from the affiliated University. Only those College of Education students who have seen the TeachLivE environment before are eligible to be recruited. Participants will be ones to have a native level of English proficiency, with some teaching experience to be able to talk and interact with virtual students in the TeachLivE simulator.

The new case study is designed to have two settings with and without feedback provision to participants (7-minute long each). Feedback is in the form of a visual or perhaps haptic indication each time the participant exhibits a closed posture. The feedback application is designed based on an existing closed body-posture corpus and a new interface for automated online feedback provision during the sessions. All the participants experience both of the settings. Half the participants receive real-time feedback (setting 2) during this first session while the other half receive no feedback (setting 1) during this session. After the first teaching session, all participants fill out a post-questionnaire. They reenter the virtual classroom then to present the same lesson plan. During this second session, those who did not receive real-time feedback in the first session will receive it during the second, and those who received it in the first will receive none during the second. All participants will then be asked to complete the same post-questionnaire that they filled out after the first session.

3.2.2 Recording

In the new experiment, in addition to full-body and video recording, we will keep a track of virtual classroom's interactions as well. Virtual avatars' NVC signals and their talk-time will be recorded in a log-file using the comma-delimited format in an observer station for further multimodal analysis such as understanding the correlations of teacher's wait-time, student's talk-time or teacher's talk-

time with teacher's body movement data in the two settings or groups. We are using our reflection tool called ReflectLivE for saving and analyzing the related log-data.

3.2.3 Preliminary Results and Analysis

We have recruited 12 participants so far. As mentioned before, half (6) of the participants received automated real-time feedback about their body poses in the first session and the other half received such feedback in the second session. Based on recorded skeletal data, we observed that only those who received feedback in their first session demonstrated a significant number of open poses in the session containing no feedback. However, the post-questionnaire information indicated that most of the participants were more mindful of their body poses while teaching after they had participated in the study. Some strict adjustments about the case study 2 recruitment group has been made. The discussion section from case study 1 explains the reasons for those changes to minimize the negative effect of confounding conditions.

4. FUTURE WORK

For future steps, we intend to complete the new experiment phase. Multimodal data analysis will be performed next, and correlations will be reported. This phase involves multimodal signal processing and related machine learning approaches.

5. ACKNOWLEDGMENTS

The author wish to thank TeachLivE principal investigators Drs. Lisa A. Dieker, Michael Hynes and Charles E. Hughes (faculty supervisor), as well as Cynthia Hutchinson, Joyce Nutta, Thomas Owens, Jacqueline Rodriguez, Mrs. Pam Jones and Mr. Michael Hopper for their support during the data collection and analysis processes. Their insights and their assistance with subject recruitment were invaluable. I also wish to express my appreciation for the support of the entire TeachLivE team and the participants. Finally I wish to acknowledge support for the evolution of the TeachLivE project from the Bill & Melinda Gates Foundation (OPP1053202) and the National Science Foundation (CNS1051067, IIS1116615). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

6. REFERENCES

- [1] Visual gesture builder: A data-driven solution to gesture detection, July 2014.
- [2] M. W. Alibali, S. Kita, and A. J. Young. Gesture and the process of speech production: We think, therefore we gesture. *Language and cognitive processes*, 15(6):593–613, 2000.
- [3] M. W. Alibali and M. J. Nathan. Teachers' gestures as a means of scaffolding students' understanding: Evidence from an early algebra lesson. *Video research in the learning sciences*, pages 349–365, 2007.
- [4] R. Barmaki. Nonverbal communication and teaching performance. In P. Z. M. M. B. M. Stamper, J., editor, *Educational Data Mining 2014*, pages 441–443, 2014.
- [5] R. Barmaki and C. E. Hughes. A case study to track teacher gestures and performance in a virtual learning environment. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, LAK '15, pages 420–421, New York, NY, USA, 2015. ACM.
- [6] C. Caswell and S. Neill. *Body language for competent teachers*. Routledge, 2003.
- [7] L. G. Collins, A. Schrimmer, J. Diamond, and J. Burke. Evaluating verbal and non-verbal communication skills, in an ethnogeriatric osce. *Patient education and counseling*, 83(2):158–162, 2011.
- [8] C. Danielson. *The framework for teaching evaluation instrument*. Danielson Group, 2011.
- [9] L. A. Dieker, C. L. Straub, C. E. Hughes, M. C. Hynes, and S. Hardin. Learning from virtual students. *Educational Leadership*, 71(8):54–58, 2014.
- [10] L. K. Guerrero and K. Floyd. *Nonverbal communication in close relationships*. Routledge, 2006.
- [11] M. E. Hoque, M. Courgeon, J.-C. Martin, B. Mutlu, and R. W. Picard. Mach: My automated conversation coach. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 697–706. ACM, 2013.
- [12] S. D. Kelly, S. M. Manning, and S. Rodak. Gesture gives a hand to language and learning: Perspectives from cognitive neuroscience, developmental psychology and education. *Language and Linguistics Compass*, 2(4):569–588, 2008.
- [13] A. Kendon. *Gesture: Visible action as utterance*. Cambridge University Press, 2004.
- [14] M. Knapp, J. Hall, and T. Horgan. *Nonverbal communication in human interaction*. Cengage Learning, 2013.
- [15] R. Laslett and C. Smith. *Effective classroom management: a teacher's guide*. Routledge, 2002.
- [16] J. Mahon, B. Bryant, B. Brown, and M. Kim. Using second life to enhance classroom management practice in teacher education. *Educational Media International*, 47(2):121–134, 2010.
- [17] R. J. Marzano and M. D. Toth. *Teacher evaluation that makes a difference: A new model for teacher growth and student achievement*. ASCD, 2013.
- [18] D. McNeill. *Hand and mind: What gestures reveal about thought*. University of Chicago Press, 1992.
- [19] A. Mehrabian. Communication without words. *Psychological today*, 2:53–55, 1968.
- [20] P. W. Miller. Body language in the classroom. *Techniques: Connecting Education and Careers*, 8:28–30, 2005.
- [21] A. Nagendran, R. Pillat, A. Kavanaugh, G. Welch, and C. Hughes. A unified framework for individualized avatar-based interactions. *Presence: Teleoperators and Virtual Environments*, 23(2):109–132, 2014.
- [22] T. Pfeiffer, M. E. Latoschik, and I. Wachsmuth. Conversational pointing gestures for virtual reality interaction: implications from an empirical study. In *Virtual Reality Conference, 2008. VR'08. IEEE*, pages 281–282. IEEE, 2008.
- [23] V. P. Richmond, J. C. McCroskey, and M. L. Hickson III. Nonverbal behavior in interpersonal relations. 2014.