University of Central Florida

STARS

Gesture Assessment of Teachers in an Immersive Rehearsal Environment

2016

Roghayeh Barmaki
University of Central Florida

Find similar works at: https://stars.library.ucf.edu/etd

University of Central Florida Libraries http://library.ucf.edu

Part of the Computer Sciences Commons

University of
Central
Florida

STARS
Showcase of Text, Archives, Research & Scholarship

للاستشارات

www.manaraa.com

GESTURE ASSESSMENT OF TEACHERS IN AN IMMERSIVE REHEARSAL
ENVIRONMENT

by

ROGHAYEH BARMAKI
B.S. Kharazmi University, 2008
M.S. Iran University of Science and Technology, 2012

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Computer Science
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Summer Term
2016

Major Professor: Charles E. Hughes

ii

ABSTRACT

Interactive training environments typically include feedback mechanisms designed to help trainees improve their performance through either guided- or self-reflection. When the training system deals with human-to-human communications, as one would find in a teacher, counselor, enterprise culture or cross-cultural trainer, such feedback needs to focus on all aspects of human communication. This means that, in addition to verbal communication, nonverbal messages must be captured and analyzed for semantic meaning.

The goal of this dissertation is to employ machine-learning algorithms that semi-automate and, where supported, automate event tagging in training systems developed to improve human-to-human interaction. The specific context in which we prototype and validate these models is the TeachLivE teacher rehearsal environment developed at the University of Central Florida. The choice of this environment was governed by its availability, large user population, extensibility and existing reflection tools found within the AMITIES [1] framework underlying the TeachLivE system.

Our contribution includes accuracy improvement of the existing data-driven gesture recognition utility from Microsoft; called Visual Gesture Builder. Using this proposed methodology and tracking sensors, we created a gesture database and used it for the implementation of our proposed online gesture recognition and feedback application. We also investigated multiple methods of feedback provision, including visual and haptics. The results from the conducted user studies indicate the positive impact of the proposed feedback applications and informed body language in teaching competency.

In this dissertation, we describe the context in which the algorithms have been developed, the importance of recognizing nonverbal communication in this context, the means of providing

---

[1] Avatar-Mediated Interactive Training and Individualized Experience System

iii

semi- and fully-automated feedback associated with nonverbal messaging, and a series of preliminary studies developed to inform the research. Furthermore, we outline future research directions on new case studies, and multimodal annotation and analysis, in order to understand the synchrony of acoustic features and gestures in teaching context.

To *my parents and my spouse*

for their dedicated love.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

ix

LIST OF FIGURES

LIST OF TABLES

CHAPTER 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 [83]. Having good skills in nonverbal communication is a fundamental part of social competence [67]. "We speak with our vocal organs, but we converse with our whole body" [5]. The ability to receive or to accurately decode nonverbal cues, as well as the ability to send them expressively and unambiguously matters greatly in our daily lives [66].

More specifically, in classrooms and educational environments establishing good communication between students and the teacher is a critical step for both the learning and teaching processes [26, 66, 86]. Teaching and learning is an inherent communication process where teachers and students engage in both verbal and nonverbal relations [91]. Research indicates that the majority of interaction between individuals, including students and teachers, is nonverbal, encompassing between 65 and 93 percent of what occurs related to learning [51, 85]; therefore when preparing teachers, it is worth mentioning and practicing NVC skills for interpersonal interactions.

There are a number of essential skills needed for pre-service teachers to develop prior to entering the real classroom. Apart from the theoretical courses and references that help novice teachers to passively learn about teaching proficiency basics such as communication and classroom management skills, simulation-based training systems provide a safe and comfortable environment for them to interactively practice teaching skills in a realistic setting. TeachLivE is an immersive, virtual environment, designed and implemented at the University of Central Florida, for teachers to rehearse and hone their classroom skills. In this virtual classroom, teachers interact with student avatars that are controlled in real time by a human-in-the-loop system that blends human and computer agency [34].

For purposes of this dissertation, TeachLivE provides an environment in which to develop and validate the associated research by using its innate ability to provide immediate feedback for

1

teacher training. The main goal of this dissertation is the development of a semi-automated assessment and feedback system that assists in-service and practicing teachers to improve their communication skills as well as their classroom management and pedagogy. Here the behaviors of teachers from nonverbal communication aspects, including body posture, proximity, and their correlations with teaching preparedness and competency, are analyzed using machine learning methodologies. The dissertation also presents measures of user satisfaction, accuracy and validation of the proposed automated nonverbal communication feedback application.

1.1   Research Questions

The following questions guided the research underlying the dissertation:

1. What communication strategies do participants employ to interact with the avatars and exhibit real-world skills?

2. What types of data (parallel input modes) and information can be collected in these contexts?

3. What algorithms and methodologies from the perspectives of data mining and machine learning are most effective in these contexts?

4. What presentation modes and timing provide optimal feedback for participants to understand and improve their verbal and nonverbal interactions, especially in gesture and posture evaluations?

5. How effective and valid is the feedback application? How does purely objective (automated) feedback compare with that provided by human experts?

6. Can choices made by human experts help to evolve a reliable and valid automated mode of feedback?

7. How do we most effectively combine data collected by mining and machine learning algorithms to automate or at least semi-automate tagging used in reflective processes?

1.2   Research Contribution

Education researchers have investigated the impact of nonverbal behaviors in classrooms [7, 62, 77, 130] where they rely on manual annotation of video recordings of teachers in their classroom settings, or on-site observers (also known as subject matter experts) who collected notes for analysis [101]. In other words, most of these studies are qualitative by nature, or require researchers to manually annotate hours of video recordings [21]. As the theoretical underpinnings of nonverbal communication (in education) have been established already based on the extensive existing body of literature, now is the right time to speed up discovery and data analysis by automating some of these hand-coded procedures. There are some research projects that are working on automated feedback provision in virtual environments for job interview [56] and social presentation skills [117]; however the context is very different from teacher training. The feedback provision method in these previous works is also limited to some objective categories, such as voice volume [117], or facial expression mimicking [56]. The existing research in education has shown that performance feedback is an essential component of effective professional development and staff training packages that target workplace behavior change [72]. Considering the advantage of having an existing successful teacher rehearsal environment already developed by our research group, we focused on teacher performance feedback as the main application.

The approach of this dissertation is to contribute to the design of a (semi) automated tagging system for teacher assessment in the TeachLivE rehearsal environment. More specifically, the automated gesture recognition component of tagging is the problem that this research investigates. This is a machine learning problem, because in order to process the enormous quantities of data produced by such a system, machine learning interfaces must be used. The system has recorded

3

different types of data (including event logs, video, and body tracking) that must be evaluated to have high-level features for a gesture recognition feedback system. For example, proximity-based outlier detection techniques have been applied to improve the accuracy of the Microsoft tool for gesture recognition, called Visual Gesture Builder.

We have used the Microsoft Kinect V2 to capture body movements as low-level data which we transform into higher level features as part of an assessment procedure. In the teaching effectiveness content, the associated tags typically fall into the following categories: purely objective ones that require no actual understanding of intent of the behavior, for example, proximity and talk time; those with shallow semantics associated with intent, for example, open versus closed posture or use of open versus closed words, (such as why versus what) that can be part of a deeper semantic analysis, for instance, showing respect/disrespect or asking open versus closed questions; and yet others that require deep analysis to acquire any semantic meaning, for example, providing encouragement versus closing off a conversation. Even in the latter, deep semantic situations, cues can be observed through body postures that may indicate a desire to continue a dialogue versus a desire to close off all further communication.

We have chosen to use and improve a set of existing tools for classifying postures and to determine the degree of correlation between such pose recognition and the tags that would be chosen by subject-matter experts. As this relates to the use of tagging for reflection, we also are investigating the user interaction issue associated with form and time of feedback. For example, is it better to give information about performance in real-time (immediate feedback intervention) versus after a session is completed. The other example is haptic or vibration form of feedback versus visual cues.

4

1.3   Thesis Motivation and Organization

Even though education technologies are growing and most educational environments are becoming student-centered, it is important to consider the impact of teaching competency in K-12 Education. Most research on educational data mining focuses on student learning models; the key role of a teacher in the learning process is often neglected by such analysis. In contrast, the work reported here focuses on preparing (in-service and practicing) teachers to learn essential communication behaviors in the TeachLivE simulated environment.

Different applications and machine learning methodologies were used in this effort to design an automated posture feedback application. The eventual goal is to use this application for teacher assessment as an automated tagging (behavioral annotation) system based on posture recognition and correlations to expert human tagging.

Chapter 2 introduces the user study infrastructure, TeachLivE rehearsal environment. In this research, we needed to have a sense of teachers' interactions, and particularly their stances in the virtual environment first. Based upon existing literature on the nonverbal behaviors of teachers from actual classrooms, and recordings from the TeachLivE sessions, we collected a corpus of target gestures.

The results from two formative studies A & B (presented in Chapter 5) and related literature were the input for the next phase of the research, which was to automate the annotation procedure using the Microsoft Kinect V2 sensor and its Software Development Kit (SDK) in case study 1 (Chapter 6) and the following studies.

In case study 2, we improved the system to review the postures and gestures of the user in real-time, and provide immediate visual feedback intervention to the participating teacher. The details of the conducted case study and results are presented in Chapter 7.

One of the research goals is to observe the impact of the feedback application over time. Therefore, we conducted a pilot with a series of sessions by recruiting some participants from a

previous experiment to investigate the impact of feedback application over a four-month period. We also conducted two separate pilot studies to evaluate the effect of vibration (haptic) feedback and the teaching scenario on body posture changes. The details of these three pilots are described in Chapter 8. In Chapter 9, we conclude the research reported in this thesis and describe research limitations and future directions. Further information about the conducted studies is presented in Table 1.1 for reference.

Table 1.1: Thesis organization for conducted studies.

Chapter	Study Name	Data Analysis	Aim
Chapter 5	Formative A	TLE Video Recordings	NVB overview
	Formative B	TLE Video Recordings, YouTube, Teaching Channel	Dynamics of gesturing for Biology vs. Algebra teachers
Chapter 6	Case Study 1	RGB-D Recordings	Semi-automated gesture annotation
Chapter 7	Case Study 2	RGB-D Recordings	Automated gesture recognition and feedback application
Chapter 8	Pilot *I*	RGB-D Recordings	Feedback impact on gesturing over time
	Pilot *II*	RGB-D Recordings	Vibration vs. visual feedback
	Pilot *III*	RGB-D Recordings	Impact of teaching role playing on gesturing

The approved office of human research outcome letter, informed consent and debriefing form, pre and post questionnaires and the teaching plans of the studies are presented in the appendices.

CHAPTER 2: TEACHLIVE - A DIGITAL PUPPETRY
INFRASTRUCTURE

In this chapter, we introduce the main components of our basic research environment, TeachLivE, its underlying platform and its related reflection tools.

2.1   TeachLivE Rehearsal Environment

A human surrogate is any object, virtual, physical or even a blend of virtual and physical that acts as a stand-in for a human. Surrogates can be directly controlled or just given a specific task to carry out on behalf of a human. In the context of a virtual environment, a surrogate is often referred to as an avatar, reflecting that it is intended to represent a person in some context, rather than just carrying out a specific task on his or her behalf. In many instances, the person controlling the avatar is referred to as an inhabiter, implying that the person is representing themselves in the virtual environment.

In contrast, for the virtual learning environments we are employing here, the human is referred to as an interactor, in that he or she is playing a fictional role with distinct objectives for achieving specified emotional and behavioral responses in the participant(s). Generally, the interactor controls all critical actions, verbal and nonverbal, of his or her avatar, although the specific manifestation of the avatar, e.g., a robot in a real setting, may place constraints on how it carries out some of these desired behaviors.

The research presented here employs avatars as remote entities in scenarios involving interpersonal skills. More specifically, we are interested in situations that involve the development or honing of interpersonal skills that are needed by a K-12 teacher. An example of such use of avatars can be seen in Figure 2.1. Here we focus on rehearsal of classroom management, content delivery and pedagogical techniques in the TeachLivE environment. TeachLivE stands for teaching and

learning in a virtual environment and it has been developed at University of Central Florida to help potential and existing teachers learn new skills and hone old ones.
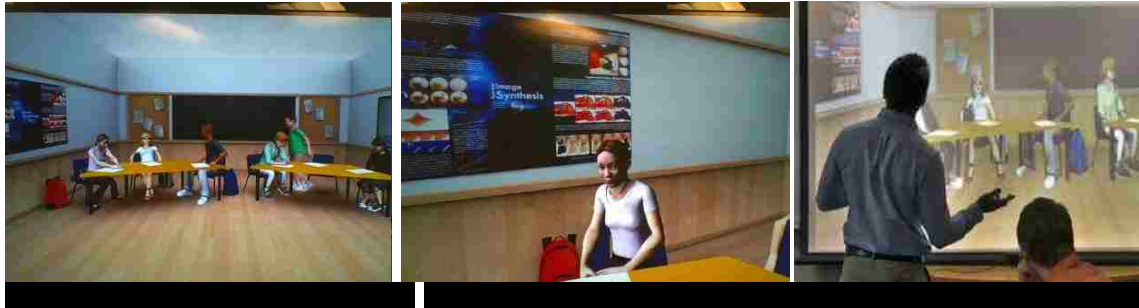
Figure 2.1: Different examples of avatars for teaching preparation purpose.

The context delivered by TeachLivE has synthetic entities, including students that are modeled individually to great detail including their clothing. This attention to detail increases the fidelity of each student, as they have assigned behavior types based on research that identifies commonly encountered personalities [75]. Moreover, each individual child has unique expressions consistent with Ekman's facial action coding system [41]. The behaviors and appearances of each synthetic entity are idiosyncratic to that character; these include certain poses that are unique to each and are designed to be representative of some specific personality. Moreover, each individual child has unique expressions consistent with Ekman's facial action coding system [41]. Two extreme avatars in the current TeachLivE middle and high school scenarios are a virtual student named Sean who is an aggressive-dependent in constant need of his teacher's attention and approval, and Maria who is an exceptionally talented passive-independent with no perceived need for the teacher's help or approval. For inclusive classroom preparation, Bailey (shown in Figure 2.1.b) is a student with intellectual disabilities and Martin (standing in Figure 2.1.a) is a teen with autism. Beyond personalities and cognitive strengths/weaknesses, each virtual student has a deep backstory and a variety of representative misconceptions related to subject matter appropriate for the student's age.

8

TeachLivE is in current use at over 80 universities and six school districts in the US. In comparison to learning skills in an actual classroom, TeachLivE harms no real children and it gives teachers the opportunity to reflect objectively on their performances, reentering the virtual classroom at a later time to improve their abilities. Furthermore, participating teachers or their supervisors can request the level of misbehavior, detailed session features and teaching plan of the classroom prior to teaching sessions in order to leverage a teacher's professional development.

Over the past several years, the TeachLivE team has been running a series of experiments to determine its effectiveness in conveying new strategies to teachers. Our initial focus was on math instruction in middle schools in the United States. These students are typically in the age range from 11 to 14 years-old. The experiment started with 157 in-service teachers at 10 distinct sites across the country. Due to attrition, 22 dropped out of the experiment resulting in a final total of 135 participants.

Each participating teacher received four levels of professional development, including computer simulation, synchronous online instruction, and lesson resources based on the common core standards [59]. We initially observed the teachers in their classrooms, then in the TeachLivE simulation and then back in their classrooms. The goals were to see if improvement occurred in the simulation and if that improvement continued once the teachers returned to their classroom settings. The first specific skill addressed was the use of describe/explain versus short-response versus yes/no questions; here we want high-order questions that involve students in the process of analysis and thinking about their own learning processes. The second was the provision of specific versus general feedback; here, we wanted feedback that relates to the student's actual performance rather than generalities.

Results indicated that four 10-minute professional learning sessions in the TeachLivE classroom simulator improved targeted teaching behaviors while in the simulator, and those improvements were seen in the real classroom as well [35].

2.2   The AMITIES Framework

TeachLivE is built on a platform called AMITIES (Avatar-Mediated Interactive Training and Individualized Experience System) [92] that provides an efficient and effective infrastructure for human-in-the loop simulated experiences.

Much of the system's uniqueness lies in its flexibility, allowing for a single person to inhabit multiple avatars as well as multiple people to control multiple or even a single avatar. Achieving this has involved the development of interactor paradigms that focus on low cognitive and physical demand, and participant paradigms that focus on situational plausibility and place illusion [99], plus a network protocol that delivers animated behaviors, including facial and body gestures, with very low bandwidth requirements [92]. This paradigm allows avatars to be inhabited by interactors who are generalists or who are specifically trained to suit fixed application domains. The basic architecture is displayed in Figure 2.2.

The Interactor Station(s). AMITIES provides a multi-functional interface for interactors who are controlling their avatar counterparts. An interactor station consists of a location in which a person can be tracked via several sensors and perform actions using a wide variety of user-interface devices. The data from the devices and the sensors together form the sensory affordances of that particular interactor instance. AMITIES is responsible for interpreting these data and encoding them into a single packet with sufficient information to capture an interactor's intent during avatar control, i.e., the system processes the individual data streams for all sensors and devices to identify a behavioral intent for the interactor such as "waving". This constructed packet is then transmitted over the network to the remote location where the avatar resides. At the avatar's end, the information in this packet is interpreted to obtain the desired behavior that must be executed by the avatar instance. AMITIES then takes into account the number of affordances of that particular avatar instance to decode the data into sub-components required by the avatar, following which the avatar executes the interpreted behavior in real-time.

10
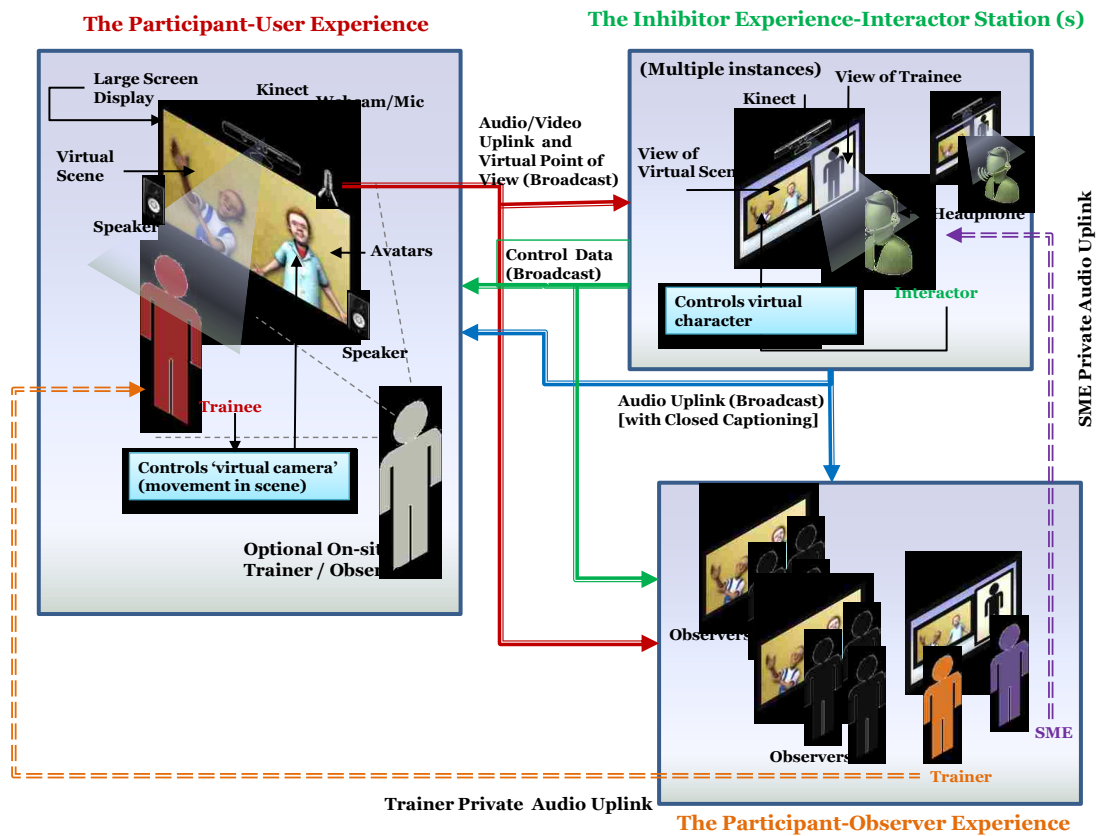
Figure 2.2: AMITIES structure and information flow [92].

Although the normal mode of interaction involves just one interactor who controls multiple avatars, typically using voice morphing software to match the pitch and timbre (speed, nasal involvement, vocal vibration, etc. are generally controlled by the interactor) for each avatar, the AMITIES infrastructure supports multiple interactors, one of whom is considered primary (everyone connects through this interactor) and the others secondary. When multiple interactors are involved, each may have his or her dedicated subset of avatars, or there may be overlap. Having multiple interactors controlling the same avatar is typically used when one is a master interactor and the other is an interactor in training. The master interactor can function much like an acting coach, taking over only to demonstrate a better approach to delivering some aspect of the perfor-

11

mance. Additionally, we have used multiple interactors of both genders in cases where we felt that voice morphing was not giving us the subtleties of gender-specific voice qualities.

The User Interface: Participants and Observers. AMITIES classifies users into two categories, depending on their interaction capabilities with the avatars. The first category is the participant who is directly involved in bi-directional conversations and actively engages in behaviors with the avatar (s). AMITIES provides an interface that allows a participant to be immersed in the environment in which the avatar-mediated interaction is occurring by tracking their motion and body poses, and correspondingly adjusting the system's response. Examples include altering the viewpoint of virtual cameras in synchrony with a user's movement to give them a sense of "immersion" in the virtual environment or autonomously altering an avatar's gaze to look at the user as they move around in the interaction space. This latter capability is included as eye gaze has been shown to be an important factor in determining the perceived quality of communication in immersive environments [47]. Additionally, AMITIES captures and transmits bi-directional audio streams to allow conversations between the participant and the avatar (that is controlled by its interactor). Selective video-streaming capabilities are also offered by the AMITIES interface at this end, allowing an interactor to view the user and the remote environment during interactions. While the system supports bi-directional video, the stream from an interactor is traditionally not required, since the avatars are the focal points of interaction for a user.

The second category of users is referred to as observers. Observers come in two modes, simple and tagging (capable of annotating events). In neither case does the observer directly affect avatar-mediated interactions. Simple observers include trainees or bystanders who wish to witness the interaction with a view to gathering information. Tagging observers (representing coaches or raters) have an interface through which they can record the participant and the avatar environments, providing annotations to categorize events that should be reviewed during self or guided reflection. Note that the tagging process also triggers automatic data recording, including reports on proximity to avatars, and talk time of the participant versus avatars.

12

2.3 Reflection in TeachLivE

In order to facilitate the process of teacher assessment, multiple software systems have been developed. The first software called TeachAARS, or TeachLivE After Action Review System, was integrated into the an earlier TeachLivE system in 2012. TeachAARS does direct video/audio capturing that contains both the virtual classroom and the participant from the screen of the main interactor's machine (server). In addition to directly recording sessions, TeachAARS has the capability to support behavior tagging. Each tag is associated with a sequence of frames, and thus allows selective viewing during reflection or debriefing. Feedback can be presented in tabular or graphical forms and can be archived (anonymously) for aggregate analysis. Figure 2.3 displays the TeachAARS environment for teacher assessment.
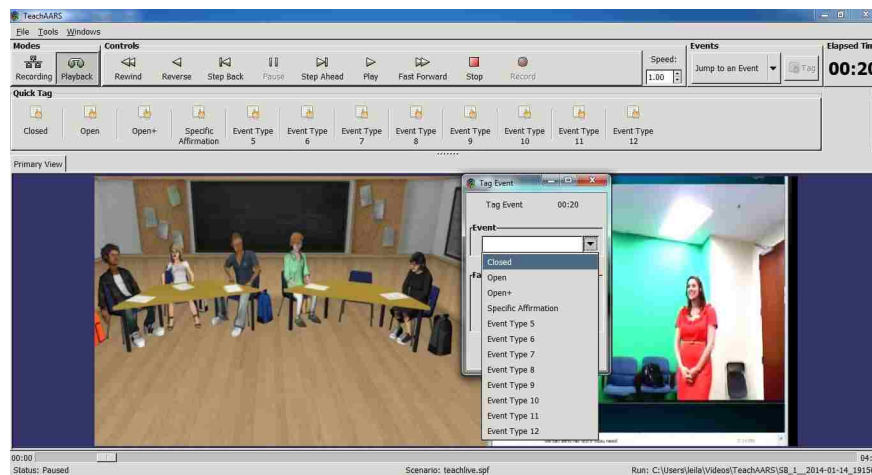
Figure 2.3: TeachLivE after action review system (TeachAARS).

The new reflection software that is integrated into the AMITIES architecture is called ReflectLivE which is shown in Figure 2.4. AMITIES has an embedded annotation/tagging component and a stand-alone analysis component to support the reflection process. This system provides support for reflection in the form of automated and manually entered tagging of events.

13
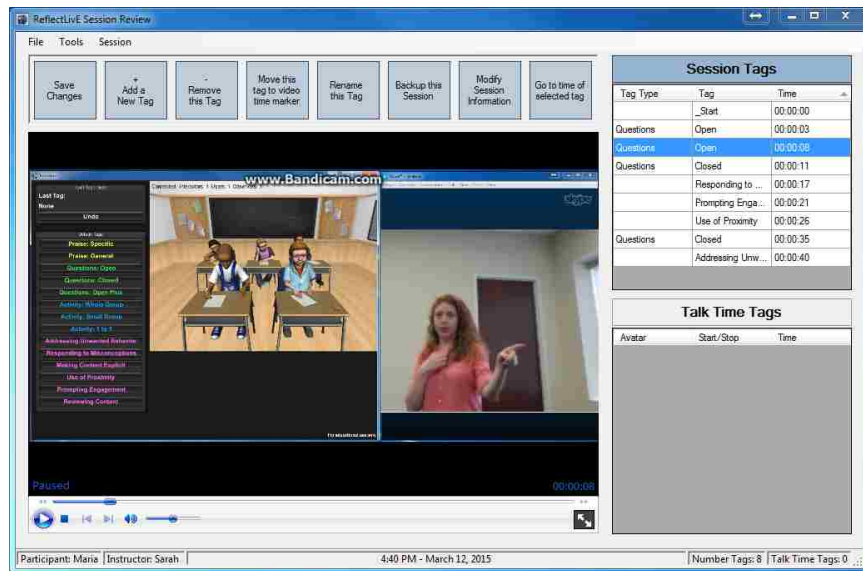
Figure 2.4: ReflectLivE application for online and offline reflection.

Automated data can include time spent in front of the class, time spent in proximity of each character, percentage of time spent in each zone of the classroom and time spent talking by the user versus that spent by the virtual characters. Data that presently requires human judgment includes (i) number of high-level questions asked, (ii) number of low-level questions asked, (iii) specific praise offered to students, (iv) general praise offered to students and (v) time from asking a question to giving an answer (Figure 2.5).

As we progress in our research, we are finding that we are gathering enormous amounts of data that we can now mine for other actions that correlate with success. For instance, a study reported later in this dissertation has shown positive correlation between body posture and perceived success, which appears to also correlate to better performance (Chapter 6). Other studies are looking at metrics related to perceived social presence [99], and its correlation with performance [55], and physical presence when remotely inhabiting an avatar. The terms presence and social presence are described in section 3.1.3.
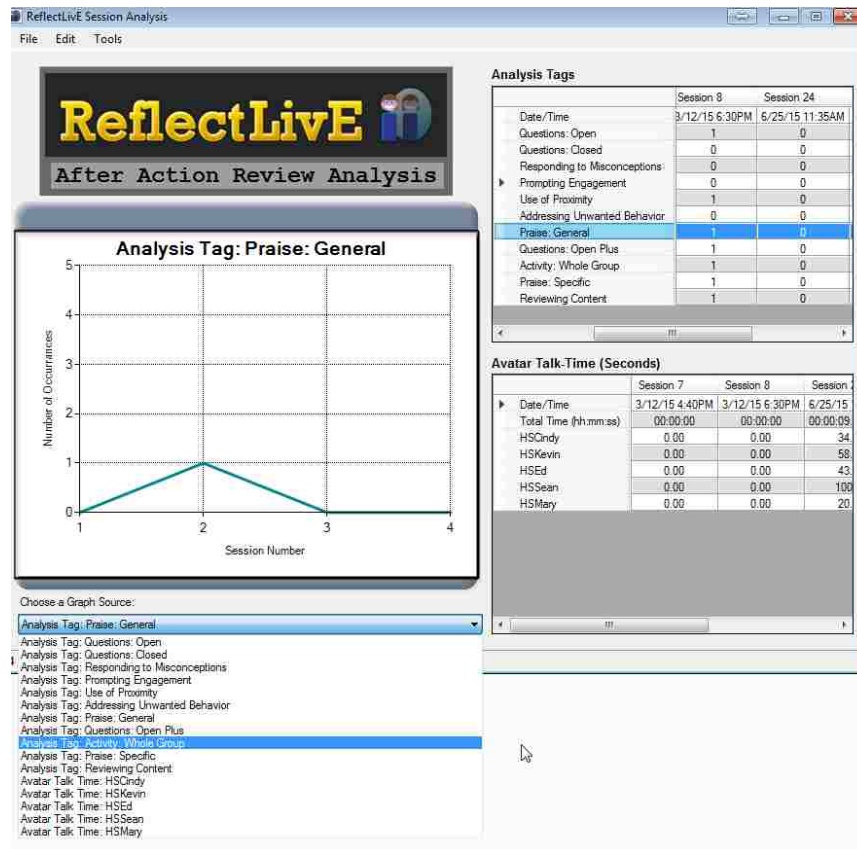
Figure 2.5: ReflectLivE session analysis.

The ReflectLivE system employs these annotations to allow participants and coaches to reflect on selective playback of a session. Annotations are typically designed to match application domains, and within domains to support specific research studies or pedagogical goals. For example, a study funded by the Bill and Melinda Gates Foundation used this system to determine if sessions in the TeachLivE virtual classroom could positively change the classroom strategies of teachers.

Currently, ReflectLivE relies on screen recording programs to record the interaction between the participant and the virtual environment. Using the ReflectLivE application, an observer can review any past or ongoing sessions and manage existing tags.

15

A new capability of ReflectLivE that distinguishes it from TeachAARS is the online tagging property in an observer station of the AMITIES architecture. Other capabilities include domain and study-specific tags, offline tagging and a diverse set of analysis and presentation tools. Figure 2.4 presents the tagging mode of ReflectLivE.

As mentioned earlier, ReflectLivE provides the ability to visualize recorded session data through charts. It combines two types of information for analysis, one from the manual annotation procedure, presented as analysis tags, and the other from automatically generated log data from session key-stroke analysis for talk-time reports, presented as avatar talk-time. Figure 2.5 displays a snapshot from the ReflectLivE session analysis main window. In this Figure, the number of tags for positive praise is reported from four sessions. The session analysis window has four components:

- Graph source selection: For reviewing the sessions, a user selects an item. The items in the selection combo-box are customized based on the input study-specific tags.

- Visual analysis: It presents the number of tag occurrences in selected sessions for manual annotations, or avatar talk time (in seconds) based on session number.

- Analysis tags: It presents a table with the detailed number of tags for each session.

- Avatar talk-time (seconds): It presents a table with detailed information about session length, and avatar talk-time for each session.

In the following chapter we describe related work on virtual environments, nonverbal behaviors and teaching evaluation. More specifically, in section 3.2, we provide a detailed list of items as teaching behavior tags that has been widely used in the ReflectLivE system and our teaching performance evaluation studies (Chapter 5 and Chapter 6).

CHAPTER 3: RELATED WORK

The nature of this study is interdisciplinary. As a result it includes a review of the literature within a variety of disciplines to find similar research and relevant previous work. The literature presented in this chapter and the following chapter are collected from the education, social and cognitive science, communication, psychology, simulation and training, educational data mining, learning analytics, technology enhanced learning, computer-mediated communication [1] (particularly virtual and mixed reality), affective computing, and multimodal data analysis communities. It is challenging to report the existing literature since the related projects are very broad. In addition, the TeachLivE environment has unique features that make it difficult to find and compare its contributions to similar projects. In general, the literature is divided into real-world and simulated environments and we have organized the flow of the discussion of the literature based on this dichotomy.

In section 3.1 we introduce some of the projects similar to TeachLivE designed for simulation and training, as well as the properties of a virtual learning environment. Afterwards, we indicate the significance of teacher assessment, current evaluation methods and projects in real classrooms and our virtual classroom in section 3.2. We then introduce the measures that interest us to provide feedback to teachers, which are all related to nonverbal communication. The importance of nonverbal communication strategies in the classroom and education environments is also presented. Many works exist on different nonverbal communication aspects of teaching and learning in the classroom, so we only focus on the most related research projects.

In Chapter 4, we introduce some multimodal, multisensor interfaces, multimodal data collection and annotation, related software, frameworks and equipment for data collection, and the

---

[1]This terminology arises from the comparison of the avatar and the agent for the virtual representation of a human from control prospective: Avatars are controlled by humans, whereas agents are controlled by computer algorithms. Hence, interaction, with an avatar qualifies as computer-mediated communication, whereas interaction with an agent qualifies as human computer interaction [89].

methodologies and projects from the perspectives of multimodal interactions in the classroom and other settings.

3.1 Virtual Rehearsal Environment

Here, we introduce some relevant projects in virtual reality in order to provide context for our test-bed environment TeachLivE. We then describe the TeachLivE system, the paradigm it embodies and its unique properties as a virtual rehearsal environment for teacher preparation.

3.1.1 Virtual Environments for Simulation and Training

Simulation-based training systems provide learners low-cost and hazardous-free environments in which they may practice and develop their skills. As a result, simulation and modeling are broadly used in a variety of fields and across many disciplines.

There are many research on simulated environments [17, 18, 22, 30], but in an example similar to our research, Luciew et al. [76] present the details of developing interview procedures for Immersive Learning Simulations (ILS). Concurrent research on body language, facial expression and proximity relative to the interview and interrogation processes are discussed in the research. Their work is focused on nonverbal expressions of human and avatar subjects that indicate the impact of nonverbal expression studies in simulation.

Another similar application is public speaking rehearsal with virtual audiences. Virtual audiences have already been used successfully in virtual reality exposure therapy to mitigate public speaking anxiety [103]. In a survey paper, Vanni et al. address the applications of virtual environments for public speaking fear and anxiety [129]. Chollet et al. present an interactive virtual audience platform for public speaking training. Each user's public speaking behavior is automatically analyzed using audiovisual sensors. The virtual characters display indirect feedback depending on each user's behavior descriptors correlated with public speaking performance [27].

18

There are many other applications of the use of modeling and simulation in education. For instance [32] provides a virtual science laboratory and tutor, and [50] presents virtual classrooms modeled for educational purposes. However, none of these provide quite the same sense of being present in the virtual classroom that is achieved in TeachLivE [34]. The concept of presence is briefly introduced in section 3.1.3.

3.1.2 Assessment, Reflection and Feedback for Practical Learning

A key factor to development and improvement of one's skills is feedback, which is one of the most influential interventions in learning [53]. The means to present feedback vary greatly and several dimensions of feedback have been identified. One of these dimensions refers to the timing of feedback, which can be delayed or immediate [53, 90]. Most of the studies conducted comparing both types of feedback, concluded that, for the majority of learning situations, the impact of immediate feedback is more positive, since delayed feedback tends to defer the acquisition of needed information [90, 117].

A tool to improve nonverbal communication skills for the purpose of public presentation is introduced by Schneider et al. [117]. They use the Microsoft Kinect V2 to track the trainee's body movements during the presentation and provide visualized immediate feedback about each subject's nonverbal communication skills. Another project called MACH is designed for interview rehearsal and provides both immediate and delayed feedback to the trainee [56]. More details about these research projects and their used sensory devices are presented in Chapter 4.

Some simulation-based training systems are paired with an after action review (AAR) tool that makes it possible for supervisors and reviewers to observe the trainee's simulation sessions and provide subsequent feedback [114, 119]. Additionally, these systems allow trainees to reflect on their performances, seeing their actions from an objective rather than personal, subjective point of view.

3.1.3 Presence in Virtual Environments

The term presence is a truncated version of "telepresence" that was introduced by the U.S. cognitive scientist, Marvin Minsky, for the first time in 1980 [88]. Telepresence refers to a set of technologies that allow a person to feel as if they were present [54], to give the appearance of being present, or to have an effect, via telerobotics, at a place other than their true location [2].

Initially, presence meant the sensation of being at the remote work-site rather than at the operator's control station [88]. Over time, presence has been simplified to be "the subjective experience of being in one place or environment, even when one is physically situated in another" [133]. In a virtual environment, presence refers to experiencing the computer-generated environment rather than the actual physical locale [133].

Researchers differentiate sub-categories for presence as physical presence, co–presence and social presence. Physical presence is defined as a sense of "being there," co–presence as "being there in a shared space with another person," [122] and social presence as the experience of being together with another individual in a technology mediated experience without acknowledging or noticing the technology that is connecting the individuals [20, 54, 94].

Immersion is a psychological state defined as "perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences" [133]. A virtual environment that produces a greater sense of immersion on the individual(s) will produce higher levels of presence as well [133].

3.2   Teaching Evaluation and Feedback

Principle # 9 from the Interstate New Teacher Assessment and Support Consortium (INTASC) states the importance of reflection and feedback in the teaching with this definition: "The teacher is a reflective practitioner who continually evaluates the effects of her/ his choices and

---

[2]https://en.wikipedia.org/wiki/Telepresence, accessed March 2016

20

actions on others (students, parents, and other professionals in the learning community) and who actively seeks out opportunities to grow professionally" [2]. To support the assessment in teaching, some teaching frameworks [31, 79] are used as evaluation instruments for teachers in K-12 education.

Danielson's framework [31] is a research-based set of components of instruction, aligned to the INTASC standards, and grounded in a constructivist view of learning and teaching. The complex activity of teaching is divided into 22 components clustered into four domains of teaching responsibility: planning and preparation, classroom environment, instruction and professional responsibilities [3].

Based on specific characteristics of the TeachLivE, and our clients' needs for teaching assessment measures, the project researchers used Delphi strategy [24] to finalize the current teacher behavior tagging options of ReflectLivE. In this method, our research partners (as panel of experts) were given a list of items; which was selected from teaching evaluation frameworks [31, 79]; and they were asked to rate each item on a Likert scale ranging from totally unimportant (= 1) to very important (= 5) [24]. The objective of using this method is to reach consensus, after providing a summary of opinions and iterating the process for few rounds. The agreed list of behaviors for teaching assessment, and their assigned tag names for ReflectLivE system is shown bellow (especial thanks to Claire Donehower and Caitlyn Bukaty for conducting the Delphi study and sharing this checklist).

- Clear communication: Communicating with students (e.g., providing clear directions, developing procedures and routines)

- Student practice opportunity: Helping students practice skills, strategies, and processes

- Question Type: Using questioning techniques (e.g., open vs. closed)

---

[3]https://danielsongroup.org/framework/

- Address unwanted behavior: Responding when students are not engaged or are displaying inappropriate behavior

- Respect/rapport: Creating an environment of respect and rapport (e.g., building relationships with students, valuing all students, understanding interests and backgrounds)

- Manage classroom procedures: Managing classroom procedures (e.g., transitions, materials)

- Teacher flexibility: Demonstrating flexibility and responsiveness (e.g., lesson adjustment, response to students, persistence)

- Feedback to students: Giving positive feedback to students (e.g., celebrating successes, acknowledging adherence to rules and procedures)

- Assessment: Using assessment in instruction (e.g., formal and informal, formative and summative, monitoring of student learning)

- Discussion techniques: Using discussion to help students elaborate on new information

3.2.1 Video Research for Teaching Assessment

Video records have been instrumental in theoretical developments by researchers contributing to studies of learning, as well as studies of nonverbal behaviors including gaze, kinesics and gestures [101, 109].

Video annotation tools offer the potential to support both the reflection and analysis of one's own teaching [101] as well as the ability to associate captured video with related student and teaching evidence. Rich and Hannafin [109] compare and contrast video annotation tools and describe their applications to support and potentially transform teacher reflection. They introduced emergent video annotation tools that have been used to address a wide range of teacher preparation and development concerns, including board certification, e-portfolios, detection of active student

22

engagement, and teacher and administrator assessment and evaluation. Two video annotation tools are introduced in section 4.1.

A study from Alibali et al. [7] indicated gestures as a means for teachers to scaffold their students' understanding. In this research, there was a scaffolding hypothesis based on the work of Lakoff [69], which states that teachers use gestures to "ground" their instructional language, especially in abstract concepts. The analysis on selected video sessions of a mathematics lesson indicated that the teachers' gestures were used most frequently for new materials, for referents that were highly abstract, and in response to students' questions and comments [7].

3.3 Nonverbal Communication

Nonverbal communication (NVC) refers to all of the elements of communication excluding the actual words used [42].

Communication scholars have compartmentalized nonverbal communication into nine distinct categories: kinesics, more commonly referred to as body language; physical appearance; chronemics or communication through time; proxemics or communication using space; paralanguage, communication through tone of voice; artifacts, communication by physical objects; haptics, communication through touch; facial expressions; olfactics, communication by means of smells; and oculesics, more commonly referred to as eye contact [110].

Gestures are a form of nonverbal communication in which visible bodily actions are used to communicate important messages, either in place of speech or together and in parallel with spoken words [64].

There is some debate about the interpretation of gesture among cognitive scientists. Even though most researchers agree that gestures are produced as part of the cognitive processes that underlie thinking and speaking, some scholars consider only movements produced along with speech as gestures [48] while other researchers have a broader view and include the movements

produced when thinking in silence as gestures [28, 121].

McNeill defines gestures as consisting of four main types [82]: iconics that are related to semantic content of speech; metaphorics that tie to an abstract concept; deictics or pointing gestures; and beats that are used to keep the rhythm of speech [62, 82].

Nonverbal communication dynamics play important roles in several professions with face-to-face interactions. In the case of physician-patient interaction, scholars presented results on the relation of nonverbal communication skills of physicians and their patients' satisfaction [29, 36]. In the classroom environment, successful student-teacher communication is also affected by the nonverbal communication skills of the teacher [61, 66]. The following section presents some of the related work on the impact of nonverbal communication in education.

3.3.1   *Nonverbal Communication in Education*

Scholars who study classroom communication emphasize the teacher's nonverbal behavior as information to students [123, 136]. Good communication between students and the teacher introduces successful steps for both the learning and teaching [38]. Communication is more than words, and it is important for teachers to understand the nonverbal messages they are sending and receiving in the classroom [31, 79].

The functions of nonverbal behavior in teaching according to Woolfolk et al. [136] are categorized as: a) indicating expectations and attitudes [84]; b) revealing emotional states and attraction [40, 41]; c) supplementing, reinforcing or regulating verbal exchanges [39]; d) being persuasive [81]; and e) influencing the performance of others [58].

Nonverbal communication strategies are consistently noted in approaches to teacher training [66]. 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 [71]. In a recent study from Wang and Loewen [130] the employment of nonverbal behaviors of teachers as feedback in second language acquisition was investigated. They explored teachers'

24

nonverbal behaviors in corrective feedback during 48 observations (about 65 hours of recordings) of nine classrooms for English as a Second Language. The results indicated that effective teachers used a variety of nonverbal behaviors in their corrective feedback, including hand gestures (specifically iconics, metaphorics, deictics, and beats), head movements, affect displays, kinetographs, and emblems.

3.3.1.1 *Body Language and Posture*

Teachers send and receive messages through their bodies [68, 26]. It is important for teachers to use open postures while interacting with their students in the classroom [9, 123]. Open posture is often used as a measure of closeness, receptivity, and interest. Open postures illustrate positive feelings to others and show that the person is open and positive to the listener [83]. For hands, some open postures are exposed (neither crossed nor folded) arms and palms that are not close to the body. For legs, if they are not crossed and the body weight is distributed equally on the feet, this represents another form of open body posture. In contrast, closed postures are often cited to indicate defensiveness, aggression, and avoidance [83]. In general, closed body stances demonstrate negative feelings to the other person. When somebody folds and crosses her arms, she seems to protect herself from the other person and her listener feels that she is not open and comfortable in the communication.

In addition, recent research on social skill training and virtual agents' behavior design [74, 93] defines a measure called bounding volume (BV) for expressive gesture quality analysis. According to Niewiadomski et al. [93], BV is the normalized volume of the smallest parallelepiped enclosing the participant's body. The BV can be considered as an approximation of the user's degree of body openness. For instance, if the user stretches her arms outside or upside, the BV increases. Figure 3.1 presents some frequent standing open and closed body posture models. Conceptualizing BV for different open and closed stances is easier with this Figure.

Figure 3.1: Reference example of standing closed postures (first row), and open postures (second row).

There are also some ambivalent body postures that may cause confusion and misunderstanding in the communication. It is suggested that teachers not display ambivalent postures. As an example, we know that having a hand on one's hip and legs crossed indicates threatening and insecurity; in contrast, an open palm shows openness. A teacher in an ambivalent posture with one hand on her hip (closed), and one open palm on her side (open) sends contradictory messages to her students and the students can get confused about what their teacher means during the interaction [26]. Some ambivalent postures from the TeachLivE recording sessions are shown in Figure 5.2.

Head movements, as a component of body expressiveness may convey the degree of approval or disapproval between the speaker and listeners [37, 61]. Most of the nonverbal behaviors vary from culture to culture; in American culture, up-and-down movements or nods indicate affirmation, and side-to-side sweeps, called shakes, are used to signal negation [80].

3.3.1.2 Gesture

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 Language and Mathematics [62].

A teacher's gestures influence student comprehension and student learning, especially in instructional discourses [15, 71]. Some studies have shown that speakers' gestures facilitate listeners' comprehension of speech [6, 68].

There is a significant body of research on the positive impact of gesturing behaviors in classroom. In a survey by Roth in 2001, the role of gestures was studied in teaching and learning. This especially addressed the role of gestures in knowing and learning scientific and mathematical concepts in school-aged children [112]. Kelly et al. showed that employing iconic gestures during a teaching session had a significant impact on English-speaking adults attempting to learn new Japanese words [63]. In a similar research study, Macedonia et al. [77] explored the impact of iconic gestures in foreign language word learning. Their research indicated that iconic gestures in comparison to meaningless gestures helped the memorization of foreign language nouns in a significant fashion.

Pozzer-Ardenghi et al. in [106] explored the videos of science lectures (subject: human body parts in biology high school classes). Their research indicates multimodal resources and nonverbal aspects of teaching may help students to be able to better articulate their conceptions and understandings with peers.

3.3.1.3 Proximity

Proximity can be used to encourage student participation and strategically redirect them [71]. In addition, proximity helps teachers to have better management in the classroom because

the students' disruptive behaviors can often be controlled by approaching them [52]. Moreover, proximity means attention, affirmation and closeness of the teacher to the speaking student [26].

Kale presented a research paper on the levels of interaction and proximity [60]. The main purpose of this study was to examine the types of classroom interaction and proximity levels between students and teachers observed in online video-based classroom cases. The findings regarding proximity showed that the highest number of reciprocal interactions between teachers and students was observed when the distance was the greatest (public). It was also found that most of the teacher-group interaction occurred at the closest level of proximity, "Intimate".

In TeachLivE, as our test-bed environment, the simulation has been designed to enable the teacher to move close to the student avatar within the virtual environment. While moving, the visual perspective moves with the teacher, even allowing eye-to-eye communication. This feature assists the participant teachers to experience a more immersive and intimate interaction [54].

3.3.1.4 *Facial Expression and Eye Contact*

Facial and eye behavior of teachers has been the focus of several studies. Gaze direction affects the degree of emotionality just as facial expressiveness conveys cues indicating the emotional and attitudinal states of interactants and can function as reinforcing events [61]. The literature supports the ability of the majority of students to perceive their teacher's faces; this is why it is very important for the teacher to make eye contact with the class [43, 123].

Eye contact and facial expressions of the participant teacher may be a future pathway to the research presented here. The current TeachLivE setup needs some adjustments in order to fit the limitations of existing facial expression coding software, as most facial expression tools need a very short distance from the facing camera and they are very vulnerable to a user's body/head movements.

28

3.3.2   Nonverbal Communication in Virtual Environments

There have been a number of prior attempts to develop social skill training and feedback applications using interactive environments.

The context of deictic expressions and the impact of pointing gestures in virtual immersive environment have been investigated [104]. 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 a similar study [46] from the same research lab the applications of iconic gestures to express shape-related references to objects in a virtual construction environment were investigated. They showed how to make use of the Imagistic Description Tree (IDT) formalism to enhance virtual construction parts with shape-related information. This formalism helps the user to specify and reference construction parts with the help of iconic gestures [46].

There are some working examples of research on nonverbal behavior in virtual rehearsal environments [76, 103] that we have already introduced them (section 3.1.1). In the following chapter, we report the literature on nonverbal behavior analysis from practical and technical point of view, with specific attention to multimodal interfaces.

CHAPTER 4: MULTIMODAL, MULTISENSOR INTERFACES

In this chapter, we define and describe the properties of multimodality, review the existing literature on multimodal data collection and annotation, discuss current software and sensors for data collection, and finally present the concept of multimodal learning analytics.

The word multimodal[1] is used in different fields and generally with different meanings. In the field of Human-Computer-Interaction (HCI), a mode or modality is a natural way of interaction (speech, vision, face expressions, gesture, handwriting, and head or body movements). In general, the word multimodal is associated to the input rather than the output of information [108]. Multimodal interfaces process two or more combined user input modes, such as speech, pen, touch, manual gesture, head and body movements, gaze and eye direction [108]. The paradigm aims to recognize human behavior and language in a natural way using input sensors.

State-of-the-art multimodal input systems are generally limited to two to three modal input channels. The choices of modes and number of modes are very application specific. We explain some related multimodal research projects as follows.

Presentation trainer [117] collects multimodal data using the Microsoft Kinect and provides immediate cues about the trainee's body posture, embodiment and voice volume during her presentation. Figure 4.1 shows the user-interface for this application. In this Figure, the check-mark shown on the TV screen indicates positive feedback about the trainee's hand gestures and body orientation.

Similarly, Dermody and Sutherland [33] present a multimodal prototype for public speaking purposes that uses the Kinect sensor. Their system provides real time feedback on gaze direction, body pose and gesture, vocal tonality, vocal dysfluencies and speaking rate.

---

[1]The word "multimodal" is a free (but etymologically precise) translation of a Greek adjective which, for Homer's Odysseus or any man, means the "man of many ways" or the "man of many devices"[78].

30
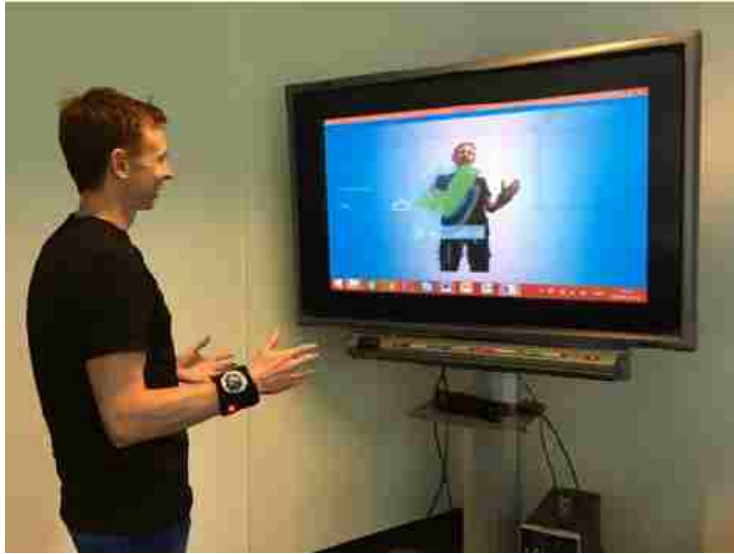
Figure 4.1: Presentation trainer user interface [117].

Figure 4.2: MACH user interface [56].

For job interview training, Hoque et al. have presented a system called MACH: My Automated Conversation coacH [56]. This system uses a web-cam facing the participant and a TV screen showing the MACH avatar with the interaction occurring in a seated position. During the training session, MACH (an automated agent) asks some predefined interview questions, mimics certain behaviors issued by the user, and exhibits appropriate nonverbal behaviors. The user

31

interface is shown in Figure 4.2. Following the interaction, MACH provides a visual summary (Figure 4.3) of the user's performance on nonverbal behaviors, such as smiles, head movements, and intonation change over time.
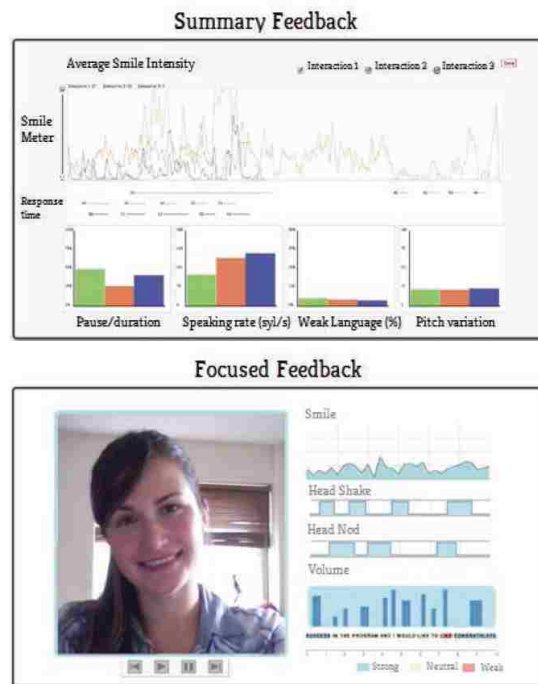
Figure 4.3: MACH after session visual summary [56].

Saleh et al. [113] presented an interaction model for human-robot nonverbal communication. They focused on head movements and facial expressions in the social interaction scenarios. Multimodal channels of input including RGB imagery and depth were used to model a robot's perception based on that of a human. This perception includes face detection, face tracking, head pose estimation, and facial expression recognition of the human in the HRI (Human-Robot-Interaction). They specifically focused on recognizing universal facial expressions (disgust, happiness, sadness, fear, neutral, anger and surprise) and reported successful recognition rates from the existing 3D face database (Bosphorus). For future research, their proposed system might be extended to in-

32

clude hand gestures as well as body postures in the nonverbal communication [113].

4.1 Multimodal Data Collection and Annotation

The capture of a multimodal corpus requires complex settings, such as instrumented lecture and meeting rooms containing capture devices for each of the modalities that are intended to be recorded, but also, most challengingly, requiring hardware and software for digitizing and synchronizing the acquired signals [105].

There are a few public frameworks for multimodal data collection. We briefly introduce two of them. The iMotions [2] software integrates biosensors and synchronizes eye tracking, facial expression analysis, EEG, GSR, EMG, ECG and Surveys in one unified software platform. The MultiSense [3] is a free perception framework developed by the MultiComp Lab at the University of Southern California that enables multiple sensing and understanding modules to inter-operate simultaneously, broadcasting data through the Perception Markup Language. MultiSense supports multiple modules, including a vision module (for tracking face features like smile, gaze, attention, activity, etc) gesture recognition module, and a speech recognition module.

After capturing the data, we need to analyze them. There are some multimodal annotation software applications, such as ANVIL [4] and ELAN [5] that we introduce them in this section.

ANVIL is a free video annotation tool that offers multi-layered annotation based on a user-defined coding scheme [65]. ANVIL can import data from phonetic tools like Praat which allow precise and comfortable speech transcription (see section 4.1.2 to learn about Praat). It can display waveform and pitch contour. ANVIL's data files are XML-based. Exported tables can be used for analysis in statistical tool-kits. The next scheduled version will also be able to import ELAN files.

---

[2]https://imotions.com/

[3]https://confluence.ict.usc.edu/display/VHTK/MultiSense

[4]http://www.anvil-software.org/

[5]http://tla.mpi.nl/tools/tla-tools/elan/

33

The software is written in Java and runs on Windows, Mac and UNIX platforms. Figure 4.4 shows the ANVIL tool. The video clip is one of the recorded video sessions from our case study that presents the participant front view and virtual classroom scene.
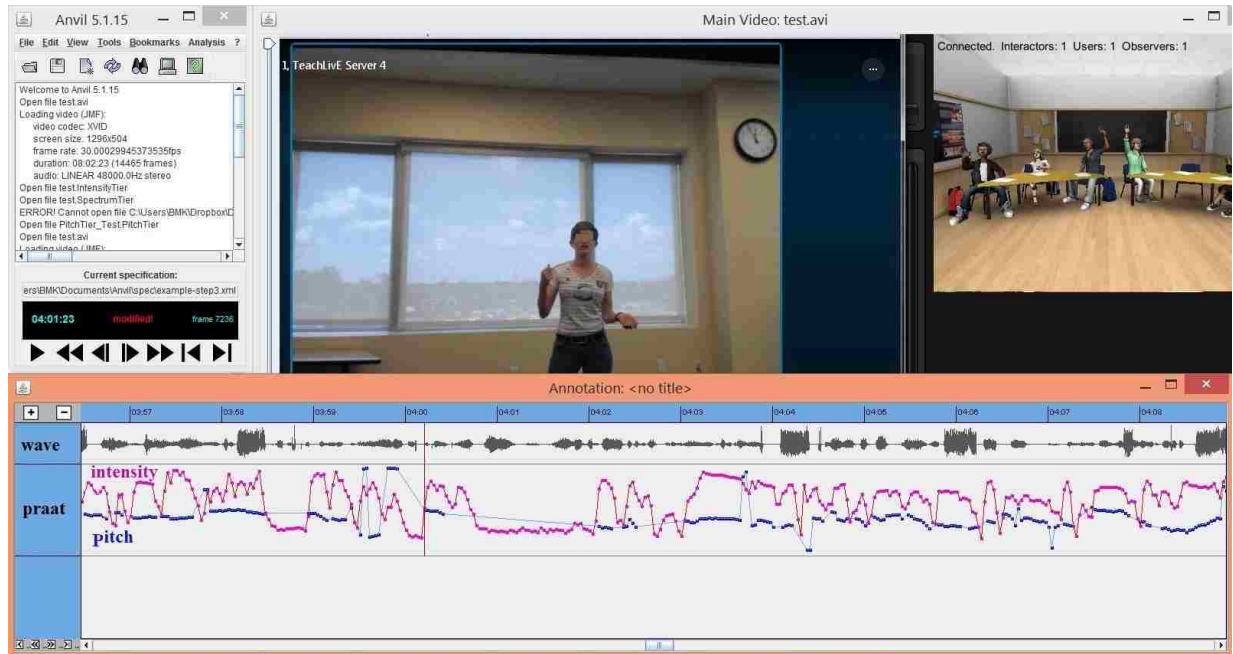
Figure 4.4: A snapshot from one of the recorded video sessions [14] in the ANVIL annotation tool [65]. Three acoustic contours: waveform, pitch and intensity are imported to the annotation project.

The current version of the ANVIL does not support the exported (closed) labels of frames from the Kinect V2 gesture recognition tool as a contour. Our eventual goal is to add this automated gestural information as a new contour to the ANVIL.

ELAN is a professional tool for the creation of complex annotations on video and audio resources [134]. ELAN is written in Java as a local tool and stores the transcription data in a specialized XML format, EAF (ELAN Annotation Format). It is available for Windows, Mac and Linux operating systems. The ELAN window (shown in Figure 4.5) displays a menu bar, media player controls, and a number of viewers pane.
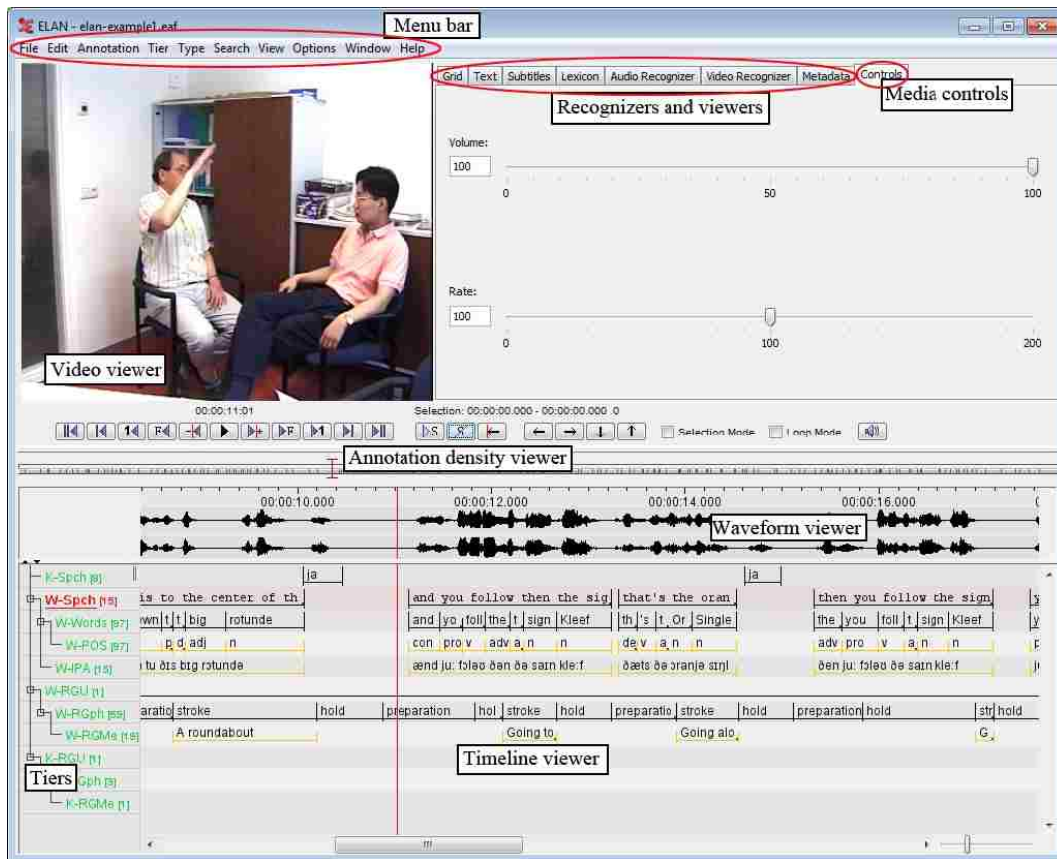
Figure 4.5: The different viewers and controls in the ELAN application's main window [127].

In the following sub-sections, we focus on two important modalities in multimodal machine learning: skeletal tracking for motion recognition, and acoustic signals. We conclude this section with multimodal data fusion definition, categories and challenges.

4.1.1 Skeletal Tracking with the Microsoft Kinect

Nowadays, many virtual, and augmented reality applications rely on low-cost marker-less motion tracking techniques, commonly achieved using depth cameras [70, 132]. A 3D depth sensor is a device that assigns distance from an anchor point in a 3D space to a 2D scene [107].

The Microsoft Kinect, a commonly used 3D depth sensor, is a motion sensing device that

contains three vital pieces that work together to detect motion and create the physical image on the screen: an RGB color VGA video camera, a depth sensor, and a multi-array microphone. Table 4.1 provides a summary of technical specifications of the Microsoft Kinect sensors.

Table 4.1: Technical specifications of the Microsoft Kinect sensors [74].

Sensor \ Properties	Microsoft Kinect for Windows V1	Kinect for Windows V2 or Kinect for Xbox One
Device		
Technology	Structured Light	Time-of-Flight
RGB Camera	640x480 @30fps	1920x1080 @30fps
Depth Sensors	640x480 @30fps	512x424 @30fps
Microphone	Quad-array microphone	Quad-array microphone
Range	0.4 to 4.5 m	0.5 to 4.5 m
# of Skeletons Tracked	2	6
Horizontal Field of View (FOV)	57°	70°
Vertical FOV	43°	60°
Gestures Tracking	Yes	Yes
# of Body Joints	20	25
SDK	Yes	Yes
Portability	No	No
USB Port	2.0	3.0
Minimum Supported Operating System	Windows 7	Windows 8
Minimum Hardware Requirements	32-bit (x86) or 64-bit (x64) processors Dual-core, 2.66-GHz processor 2 GB of RAM Graphics card that supports DirectX 9.0c	64-bit (x64) processor Physical Dual-core 3.1 GHz processor 4 GB of RAM Graphics card that supports DirectX 11

The Microsoft Kinect V1 uses Structured Light technology. It uses a projector to produce an Infrared (IR)beam that passes through a differentiation grating to produce a set of IR dots which are projected onto the scene [74]. On the other hand, Time of Flight (ToF), technology employed by the Kinect V2, uses a continuous wave intensity modulation approach [73]. Time of Flight imaging refers to the process of measuring the depth of a scene by quantifying the changes that an emitted light signal encounters when it bounces back from objects in a scene [25].

The TeachLivE system uses the Microsoft Kinect for Windows V1 for tracking the partici-pant's location in the room, and the Microsoft Kinect for Windows V2 for motion data recording, gesture recognition and immediate feedback provision.

4.1.1.1 *Pose Estimation and Recognition*

This section indicates more details about the Microsoft Kinect V2 Software Development Kit (SDK), its applications for gesture recognition, and two research studies in pose estimation with video (RGB) and RGB-D sensory data.

Visual Gesture Builder (VGB) is an application from the Microsoft SDK. It generates ges-ture databases that are used by applications to perform run-time gesture detection. We used this approach to develop our specific gesture detection procedure (see section 6.1.4 and Figure 6.3 for more details). Figure 4.6 shows the context in which Visual Gesture Builder is used [6].
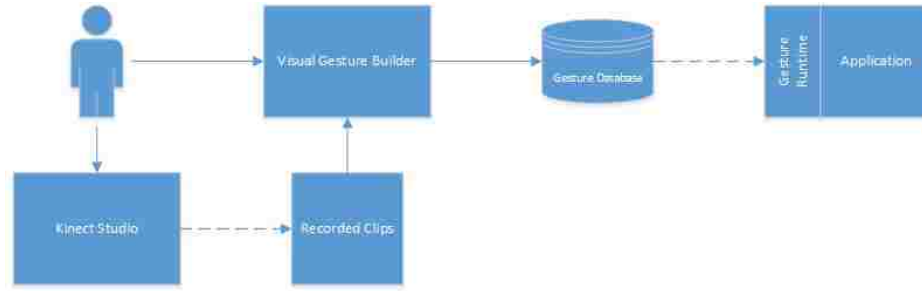
Figure 4.6: Visual Gesture Builder in Context.

We reported some of the following information about VGB from the Microsoft VGB refer-ence document [4]. In this whitepaper, 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). First of all, for each gesture, some clips

[6]https://msdn.microsoft.com/en-us/library/dn785529.aspx

37
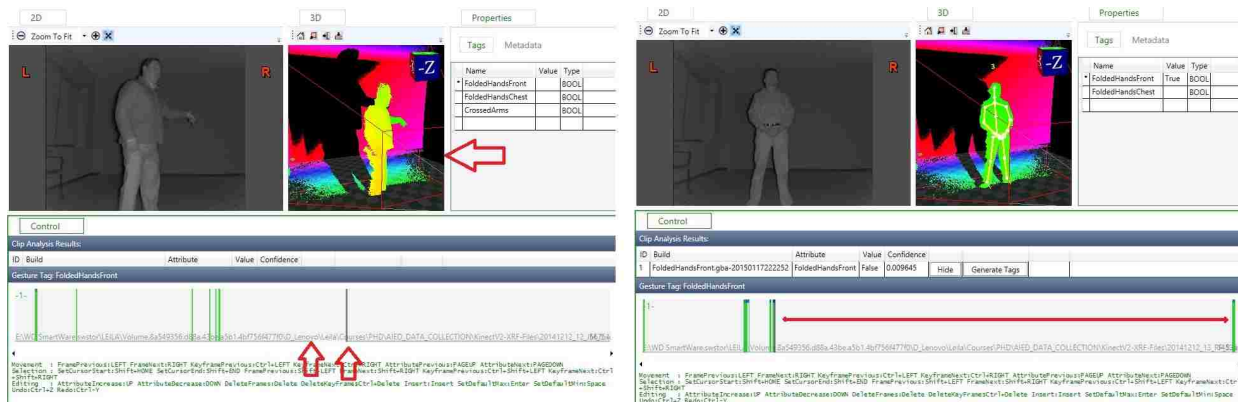
Figure 4.7: Two drawbacks of VGB skeletal data analysis. The right snapshot shows the jumping frames and the left shows lost body-tracking.

from different subjects must be recorded using the Kinect Studio utility. Every clip contains an individual in front of the Kinect (and within its field of view) who exhibits the gesture (with some variations) 5 to 15 times depending on the complexity of the gesture.

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. For each gesture, a project as a binary classifier is trained. All the frames exhibiting the gesture get the label of one (i.e. true) and the remaining frames in the clip get the value of zero (i.e. false). At this point, the problem is similar to a supervised binary classification. The classifier will recognize the frames and the patterns of the skeleton joints in which a gesture is performed. Depending on the number of gestures that we intend to recognize, the gesture recognition engine will have binary classifiers. After the training step, VGB is able to generate the tags automatically in the analysis step from the test clips. As a rule of thumb, the classifier will scan each frame within the test clip and will predict the label for each frame. The label is either true (if the targeted gesture is detected) or false.

VGB has two major drawbacks though. Jumping frames is one of the main issues with the VGB application and also the Kinect Studio utility. VGB has a problem of not tracking the body in some clips as well. Figure 4.7 shows these issues in more detail.

38

Video-based pose estimation. Song et al. [124] present a vision-based approach to gesture recognition that tracks body and hands simultaneously and recognizes gestures continuously from an unsegmented and unbounded camera input stream. Most current systems focus on one source of input, e.g., body pose. Their system tracks both body and hands, allowing a richer gesture vocabulary and more natural interaction. Their system estimates 3D coordinates of upper body joints and classifies the appearance of hands into a set of canonical shapes. A multi-layered filtering technique with a temporal sliding window is developed to enable online sequence labeling and segmentation process [124].

Pose estimation using the Microsoft Kinect. Schwarz et al. [118], present a method for human full-body pose estimation from a sequence of Time of Flight camera images. Their approach consists of detecting anatomical landmarks in the 3D data and fitting a skeleton body model using the constrained inverse kinematics. Based on the depth data, they segment the person in front of the static background and construct a graph-based representation of the 3D points. Using this graph, the method can identify anatomical landmarks in each frame by selecting points with a maximal geodesic distance from the body center of mass.

There is a huge body of literature for pose estimation, and describing them is beyond the scope of this thesis. As previously stated, we used the VGB and the Microsoft Kinect V2 SDK for gesture/pose recognition. In the following section and section 4.2, we describe related work in pose and gesture recognition with specific applications in education.

4.1.1.2   Kinect and Educational Research

After launching the Microsoft Kinect, many researchers explored the potential of this sensor in different disciplines. Some relevant research projects using the Microsoft Kinect and its supporting SDK are presented here.

In [57], the potential of using the Kinect in the classroom and its capabilities to enhance classroom interactions and to ignite student creativity is presented. The author supports that the

39

Kinect is capable of being a tool to enhance teaching and a tool to support learning. At the time that paper was published, Kinect for Windows had just been released (2011) and there were not many applications and software tools to support the use of Kinect. Therefore the author gave very few examples about the creativity aspect of having a Kinect in classrooms.

The work of Won et al. [135] from Stanford University uses the Kinect technology to recognize engagement in interpersonal communication. Previous literature highlights the constructive relationship between gesture and learning during instruction [6, 112]. Therefore, they investigated the gestures of teacher/student dyads to support this correlation. They applied computer vision hardware (Microsoft Kinect sensor) and machine learning algorithms to the gestures of teacher/student dyads (N = 106) during a learning session to automatically distinguish between high and low success learning interactions. Their model predicted learning performance of the dyad with acceptable accuracy (87.5 %).

4.1.2 Audio Streams

Prior work in speech and audio recording and processing has established a minimum frequency of 8Hz for speech recognition, and higher frequencies, between 12 Hz and 24 Hz, for conducting prosodic and spectral analysis [138].

Praat [7] is a free computer software package for the analysis of speech in phonetics [23]. It can operate on Windows, Macintosh and UNIX platforms similar to ANVIL software. Praat also supports speech synthesis, including articulatory synthesis. Praat is one of the most widely used programs within the linguistic research community.

As mentioned in section 4.1, the ANVIL annotation software can import data from the Praat phonetic tool. The most common contours are voice pitch and intensity. Voice pitch is the perceptual correlate of vocal fundamental frequency and voice intensity indicates voice loudness in

---

[7]http://www.fon.hum.uva.nl/Praat/

db. A PitchTier object represents a time-stamped pitch contour, i.e. it contains a number of (time, pitch (Hz)) points, without voiced/unvoiced information. An IntensityTier object represents a time-stamped intensity contour, i.e., it contains a series of (time, intensity) points [23]. In Figure 4.4, pitch and intensity tier associated with one of our recorded sessions was exported to the ANVIL.

4.1.3 Multimodal Data Fusion

The goal of multimodal fusion is to determine the best set of experts (classifiers here for example) in a given problem domain and devise an appropriate expert that can optimally combine the decisions rendered by individual experts [87, 111]. In general, the fusion levels are categorized into two broad categories: pre-classification or fusion before matching, and post-classification or fusion after matching [108]. Table 4.2 shows the categories of multimodal fusion, and different machine learning approaches associated with each category.

Table 4.2: A summary of multimodal fusion methods based on their level of matching [108, 111].

Data Fusion Level	Explanation	Machine Learning Methods
Sensor- **Before matching**	The raw data from multiple sensors are processed and integrated to create new features.	Feature extraction
Feature- **Before matching**	Fusing multiple feature sets from different modalities to create a new feature set.	Dimensionality reduction, feature selection, feature transformation
Match Score- **After matching**	Multiple classifiers output a set of scores that are fused to generate a single scalar score.	Sum rule or the weighted average of the scores, normalization
Rank- **After matching**	Each classifier associates a rank with every enrolled identity. No normalization is required.	
Decision- **After matching**	Each matcher outputs its own class label; a single class label can then be obtained by employing techniques such as AND/OR rules.	AND/OR, majority voting, weighted majority voting, Baysian decision fusion

41

Based on multimodal signal integration, different terms have been defined:

- Simultaneous integration indicates the pattern of two input signals produced in a temporally overlapped manner.

- Sequential integration is a sequence of separated multimodal signals, one presented before the other, with a brief pause intervening.

- Semantic-level or late fusion is a method for integrating semantic information derived from parallel input modes in a multimodal architecture. This approach has been used for processing speech and gesture input.

Our research fits inside the semantic-level category, because we capture information from parallel and separate input modes, including body movements, audio and video and event logs from a multimodal architecture. The input mediums are the AMITIES application (including its own VGB camera, Microphone, speakers) and the Microsoft Kinect V2. As mentioned by Popescu-Belis [105], the main challenge in our multimodal data capture and analysis was also the synchronization of these diverse signal.

4.2 Multimodal Learning Analytics

Multimodal Learning Analytics is the research of processing learning data from dissimilar sources in order to automatically find useful information to give feedback to the learning process [95, 138]. The multimodal approach to classroom analysis has implications for teaching and learning as well as teacher training and development.

The research in [95] presents some multimodal features from the group activity of the students to estimate the level of expertise among students. They processed video, audio and pen strokes information included in the math data corpus [96], a set of multimodal resources provided

to the participants of the second international workshop on multimodal learning analytics. According to their results, very simple multimodal features, such as percentage of time that the student uses the calculator, the speed at which the student writes or draws and the percentage of time that the student mentions numbers or mathematical terms, are good discriminators between expert and non-expert students.

Similar work done by Worsley et al. [137] also focused on understanding and identifying expertise as students engage in a hands-on building activity. Their techniques leverage process-oriented data, and demonstrate how this temporal data (object manipulation and gestures) can be used to identify elements of expertise among students.

Multimodal discourse analysis (MDA) is an emerging paradigm in discourse studies, which extends the study of language per se in combination with other resources, such as images, scientific symbolism, gesture, action, music and sound [98]. In a project done in a general paper classroom in Singapore [97], the meanings made in the multimodal pedagogic discourse, specifically in language, gesture, positioning and movement, have been investigated. Their project particularly focused on proxemics to multimodal classroom discourse analysis as well as material distance socio-semiotic meanings.

The work of Ezen et al. [44] investigates multimodal features related to posture and gesture for the task of classifying students' dialogue acts within tutorial dialogue (called JavaTutor). The results indicate that the accuracy of the existing unsupervised model significantly improved with the addition of automatically extracted posture and gesture information using the Microsoft Kinect sensor.

In summary, we learned about existing literature in nonverbal communication, and multimodal data analysis and its impact in educational context. The following chapters describe our conducted observational, experimental and pilot studies to investigate the impact of informed body language on teaching competency.

CHAPTER 5: VIDEO ANALYSIS OF NONVERBAL BEHAVIORS OF TEACHERS- FORMATIVE STUDIES A & B

In this chapter, we introduce the details of two formative studies that were the knowledge-base for collecting the corpus of closed postures and developing feedback application. The existing video recordings from studies conducted for the TeachLivE research project on effective measures of teaching performance and some online video recordings were reviewed and manually annotated by subject matter experts (UCF Education faculty members and trained PhD candidates) for teaching proficiency and posture analysis in these studies.

5.1   Formative Study A

This study is related to a national research project funded by the Bill & Melinda Gates Foundation. The research focuses on experienced biology high school teachers. High school biology teachers participated in a research project over a five-month period. They were asked to interact with the high school virtual students (three males and two females, each with his or her own predefined profile of knowledge and preferred learning style) to rehearse a sample teaching scenario in nine-minute sessions taking place once a month. A series of teaching plans on technology applications in Biology was used in this project; one of these introductory teaching plans with slight adjustments is provided in Appendix C. All of the sessions of participants were recorded with TeachAARS for later evaluation.

In this formative study, we hypothesized that there is a correlation between teaching performance and having good nonverbal signals. In other words, good speakers/listeners, who can communicate to the student avatars expressively with their body movements, could be indicative of well-prepared and competent teachers. We reviewed the existing video recordings of five teachers (two sessions from each teacher) and extracted their body language properties.

44

Figure 5.1: Some of the observed closed postures in TeachLivE sessions from in-service Biology teachers.

Figure 5.2: Some of the open and ambivalent postures in TeachLivE sessions from in-service Biology teachers.

We manually annotated three types of nonverbal expressions: 1) proximity, 2) open versus closed body postures, and 3) head movements and nodding. Interestingly, the number of head nods were the same for all the reviewed clips, so we decided to exclude this feature from the analysis.

Figure 5.1 and Figure 5.2 show some screen-shots from teachers who taught in the Teach-LivE environment. The difference between postures of participants was very interesting to investigate. Figure 5.1 indicates closed postures, and Figure 5.2 presents open and ambivalent postures

45

that have been discussed earlier in section 3.3.1.1.

To evaluate the teaching skills of the participant teachers, two experts, who were blind to nonverbal assessment analysis, were asked to rate the teaching performance on a 1-10 scale (10 is best) based on an evaluation rubric. We designed this rubric based on Danielson's teaching evaluation framework [31] and the Orange County Florida public high schools instructions [1]. Table 5.1 presents a summary of collected data for five participants. The correlation matrix is shown in Table 5.2.

Table 5.1: Mean, standard deviation and range for nonverbal variables and teaching performance ratings.

Variable	Mean	(SD)	Range
# open posture	15.6	9.17	2 - 29
# closed posture	10.2	3.56	7 - 14
# proximity	15	6.48	5 - 20
total # tags	40.8	15.25	14 - 50
% time open posture	43 %	35%	2% -78 %
% max time non-interrupted		14.9%	0.54% -33%
% max time non-interrupted		28.8%	9.5% -75.9%
teaching performance rating	7	1.41	5 - 9

Table 5.2: Correlations among different nonverbal variables in formative study A, using Pearson coefficients.

	open	closed	Proximity	% open	% max time open	% max time closed
closed	-0.92					
proximity	-0.63	0.26				
% open	0.98	-0.92	-0.57			
% max time open	0.93	-0.81	-0.66	0.94		
% max time closed	-0.84	0.94	0.21	-0.81	-0.62	
teaching performance rating	0.14	-0.33	0.30	0.01	-0.15	-0.55

46

According to Table 5.2, there is a positive correlation between proximity, open body posture and total open posture time with each individual's teaching performance rating. There exists a negative correlation between the maximum of non-interrupted closed posture and closed body posture with lower teaching performance rating as well [9]. However, we did not find any significant correlations in this observational study except a strong negative correlation for open versus closed posture employment, which we expected. This lack of significance might arise from the small number of evaluated clips, or the evaluation of the teaching performance with its broad dimensions and components in a quantified one to ten scale measure.

Following this study, we modeled these postures and extracted the most frequent ones in a corpus. We used this corpus to develop future case studies.

5.2 Formative Study B

This next formative study investigated the way teachers from two different disciplines use gestures in their teaching sessions either in real or virtual classrooms [12]. Here, we analyzed video recordings of Biology and Algebra teachers from TeachLivE and real classrooms. Analysis of video records appear to indicate that Algebra teachers produce more gestures than Biology teachers.

We reviewed the recorded teaching sessions of 17 Biology practicing teachers from Florida public schools. They were asked to teach virtual students about the definition of technology and its applications to Biology in nine-minute sessions (it was from the same corpus of recordings we used in formative study A). We selected their most recent recorded clip for gesture analysis. We also reviewed existing recordings in TeachLivE with Algebra teaching topic (teaching plan is presented in Appendix C). We had access to only two algebra recorded sessions in TeachLivE, so we selected 15 online recorded algebra sessions on teaching Linear Equations. The teaching

47

sessions (or tutorials) were carefully selected from the YouTube or Teaching Channel [1] websites to match the specified teaching plan. The videos from the Teaching Channel were from distinguished teachers in the field. Therefore, we observed some variations in the recordings from the YouTube and Teaching Channel as well. The room in which the TeachLivE interactions was recorded, did not have a setup for a physical board, and the two Algebra teachers were pointing to a "virtual" board while teaching, which impacted the employment of gestures. Hence, we reported the results from Algebra videos in two different groups.

At the next step, the video recordings of teachers were segmented in short 90 second sessions. We randomly selected two segments from each clip and counted the gesture rate for those segments. This segmentation method helped us to compare the gesture employment rate for video recordings with different lengths. We used a coding scheme adapted from the one presented by McNeill [82]. We classified each gesture into one of the following four categories: points, iconic or depictive gestures, tracing gestures, and beat gestures using the work of Alibali et al. as a working example [8]:

- Points are gestures that involve extending a finger (usually the index finger) or the entire hand to indicate an object, a location, or an imaginary object or location in space.

- Iconic gestures are gestures that depict their meaning either via hand-shape or motion trajectory. Such gestures usually depict meanings literally and sometimes depict meanings metaphorically.

- Tracing gestures are gestures that trace a path along an object or inscription (e.g., tracing an axis on a graph on the whiteboard).

- Beats are rhythmic, up-and-down hand movements that are often aligned with the prosody of speech (auditory and acoustic features of speech) and that have no clear semantic meaning.

---

[1]https://www.teachingchannel.org/

48
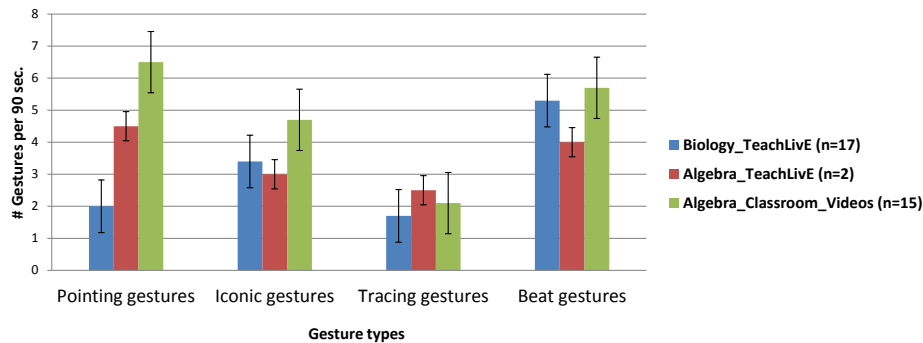
Figure 5.3: Mean rate of gestures for each segment among the Biology and Algebra teachers.

Figure 5.3 presents the mean rate of gestures for 90 second segments of recording clips from Biology and Algebra teachers. We hypothesize that since the Linear Equations topic in Algebra links to some new concepts, teachers may produce more gestures in their sessions [7, 8]. The results from our formative study in Figure 5.3 present some evidence for this hypothesis, but a new study design (with a representative number of participants) is needed to validate the correlations reported here.

5.3 Conclusion and Discussion

The results from formative study A indicated that most of the teachers did not exhibit open postures while teaching. In formative study B, we compared the gestures of Biology and Algebra teachers. We observed that Algebra teachers gesture more in their teaching sessions and this may be because of Algebra teaching content, which involves more abstract topics in comparison to Biology (the impact of technology in Biology in this specific teaching scenario).

These studies provided invaluable clues to us in order to understand the necessity and demand for social skill training in the teaching context. Our observations guided us to do further research on nonverbal behavior analysis and training, which is presented in the following chapters.

CHAPTER 6: MOTION ANALYSIS OF NONVERBAL BEHAVIORS OF TEACHERS - CASE STUDY 1

In this chapter, we describe an automated annotation system for postures. We used the pre-release version of Microsoft Kinect V2 to record the motion of participants in study sessions. For clarification, we need to explain that, in the following chapters, the words pose, posture and gesture are used interchangeably even though in the literature they do not have the same meaning. This arises from the inconsistent definition of these terms across multiple disciplines. As a rule of thumb, we did static body analysis, and we did not consider dynamics and sequence of movements, which is the definition of posture; however, in the Microsoft SDK and its respective references, gesture was used for the same term. Therefore these two terms are being used for the same concept in the following chapters.

This chapter presents more information about case study 1, objective measures for evaluation of the participants, research findings and limitations of the study.

6.1   Methodology

This case study, which we refer to as case study 1, investigated the correlations between objective measures of teaching proficiency (TP) and communication proficiency (CP) with nonverbal behaviors (gestures) of participants. The study purpose is similar to formative study A; however, we semi-automated the closed-gesture annotation procedure in this case study. Since we didn't have representative video recordings from ongoing studies, and using RGBD sensors for gesture recognition became prominent for its effective and affordable technology; we used RGB-D sensor (Microsoft Kinect V2) for human full-body tracking and the gesture recognition process.

50

6.1.1 Participants

The participants were self-selected from a group visiting UCF that was comprised of instructors and university students. The participants ($N = 14, 6F, 8M$) had an average age of 28.1 years and all were Hispanic/Latino with English as their second language. Two of the participants majored in English Language Teaching (ELT). Six of 14 participants had taught for at least five years. The other eight subjects were students with no formal classroom teaching experience. 36% were in advanced level and the remainder in intermediate level according to the UCF TESOL (Teaching English to Speakers of Other Languages) experts. As English was not the first language of members of this group, we need to report on their English language proficiency, a potential confound in our study. Finally, 50% of the participants selected the Spanish Language as a teaching topic and the others picked different topics from science, marketing and algebra. All but one of the participants were novices in the TeachLivE environment.

6.1.2 Study Procedure

The experiment reported here was a between-subjects study, fitting in a single session lasting at most 12 minutes (session range = 8-12 min., mean = 10 min.). Participants were given a consent form and a pre-questionnaire that was used to determine demographic information, teaching and educational background, English speaking level, and topic of interest for the teaching session. Each participant was introduced to the virtual classroom prior to the recording procedure and then was asked to start their teaching session.

6.1.3 Recording

Two types of data were recorded during the study. (I) skeletal data (IR, depth and body frame) and (II) virtual scene paired with user experience from the observer station. Figure 6.1 presents the modified TeachLivE setting for the study and data collection specifications.
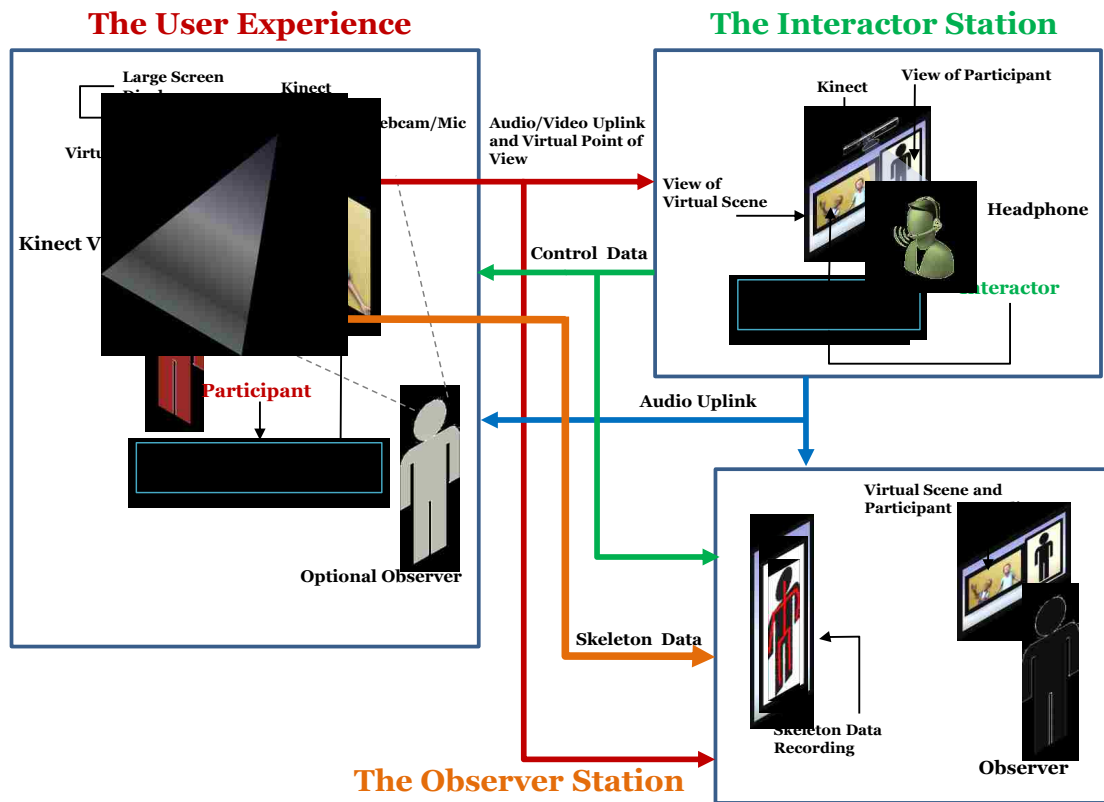
Figure 6.1: TeachLivE setting with some adjustments from the original architecture [92]. Two types of recorded streams are shown with orange and red arrows.

The skeletal data of the participants were recorded using a Kinect V2 sensor, and the Kinect Studio utility application. The recorded clips with the Kinect Studio are required as data input to Visual Gesture Builder (VGB) to do automated tagging of gestures (see Figure 4.6). The second type of collected data was the virtual scene paired with user experience from the classroom recorded with Camtasia Studio[1] in mp4 format from the observer workstation that was used for teaching proficiency analysis by experts.

---

[1]https://www.techsmith.com/camtasia.html

52

We discussed about the Visual Gesture Builder utility and postures in previous chapters. Putting all the findings and observations together, we extracted most frequent closed body postures that we wished to recognize automatically. These five closed gestures and their selected names are shown in Figure 6.2.

Figure 6.2: Reference example of closed postures; left to right: unreceptive (arms folded in front), seductive (hands clasped in back), skeptical (hands placed on hips), protective (hands clasped in lower front), and submissive (hands clasped in upper front).

A corpus of the mentioned closed body postures and some variations of these were recorded from the performances of five male and three female students using the Microsoft Kinect V2. Subjects for the main study were not involved in the corpus data collection. The corpus was used as training clips in the Visual Gesture Builder (VGB) application.

All the recorded corpus clips were hand-coded and tagged in order to train the automated analysis procedure of the participant's clips (test clips). There were five targeted closed gestures to train in this research. For each gesture four 30-sec clips in total were annotated and were added to the gesture database (5×4 short clips were annotated).

The gesture training procedure with the associated database and our created feedback application are shown in Figure 6.3. The online feedback application was developed based on the

trained gesture database (corpus) and was evaluated in case study 2. The details of the posture application are described in Chapter 7.

a) Record Gestures (Training Clips) with Kinect Studio

b) Tag Gestures with VGB (Visual Gesture Builder)

c) Build and Analyze Gestures Using VGB

d) Live Preview for Gestures with VGBView

e) Proposed Real-time Feedback Application
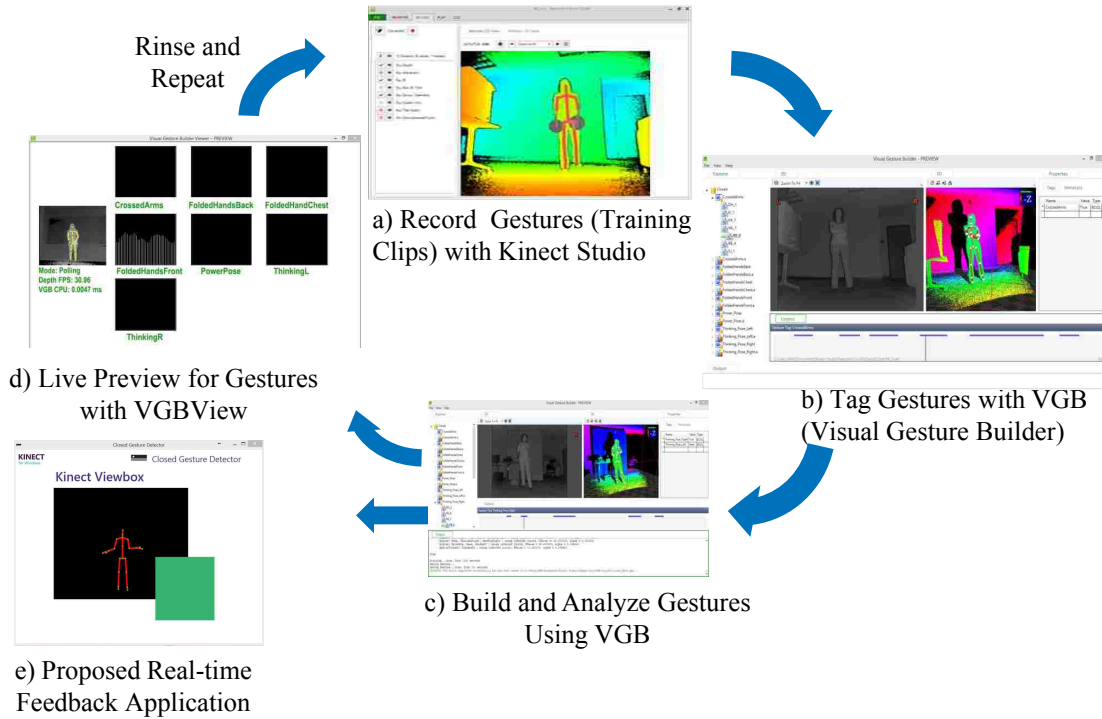
Figure 6.3: The data-driven process of building the gesture database and run-time feedback application.

All recorded valid clips from the case study participants were added to the created VGB solution with different projects trained for each closed gestures, and were automatically annotated. However, there were some problems with automatically tagged clips using the VGB application.

For example, some of the recorded frames were not tagged correctly, even though the preceding and following frames (nearest neighbor frames) were tagged correctly (False Negatives and False Positives). This means that the VGB application did not process sequences of frames, but rather processed each frame individually, independent of its surrounding context.

54
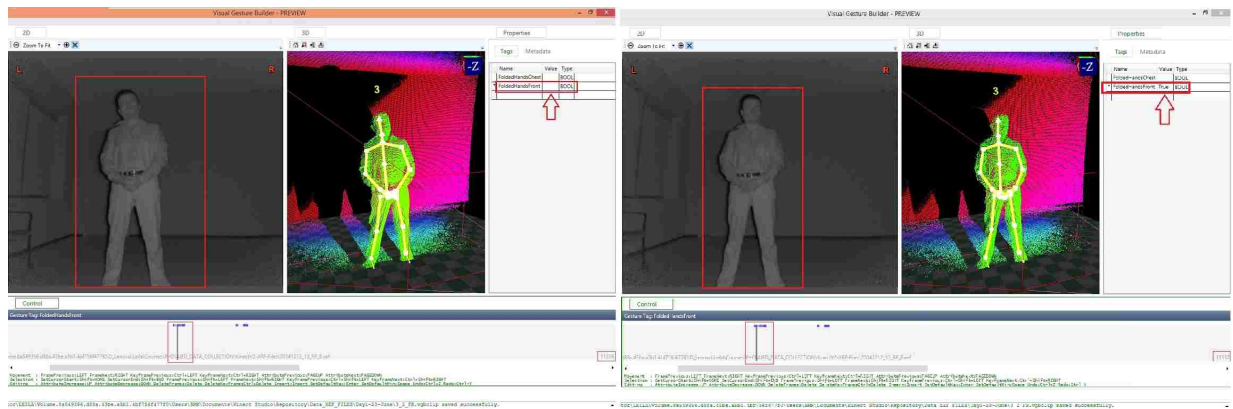
Figure 6.4: The snapshot of the VGB analysis, showing false negative error for protective gesture, or hands folded in front.

Figure 6.4 presents an example of a false negative. The two screen captures present two consecutive frames. The left frame is preceding the right frame, and it is mistakenly tagged as false. Similarly, Figure 6.5 highlights the existence of false positive errors in VGB tagging output. The red rectangle in the tags bar presents the false positive tags, and the lack of temporal considerations of tags for the mentioned classifier. Let's consider the confidence value for the current frame. Even though it has been predicted as true, the confidence level does not support this prediction as it has a very small value close to zero. So, another problem of the VGB is that it ignores or at least does not properly employ the confidence value in its analysis.

To improve the accuracy of the current classifier and reduce FP and FN rates, we sought to use the temporal information (labels of consecutive frames in this case) in the generated tag files (with .vgbclip extension) from the VGB application. We converted these label output files into .XML extension. We then applied a proximity-based outlier-detection algorithm within each window (based on variance from median) in order to appropriately tag the frames [100, 120]. The selected window size was varied based on the dynamics of the subject, i.e. his or her body movements. Figure 6.6 presents the screen capture of the VGB output after the application of the

55

proximity-based technique to one of the analyzed clips.
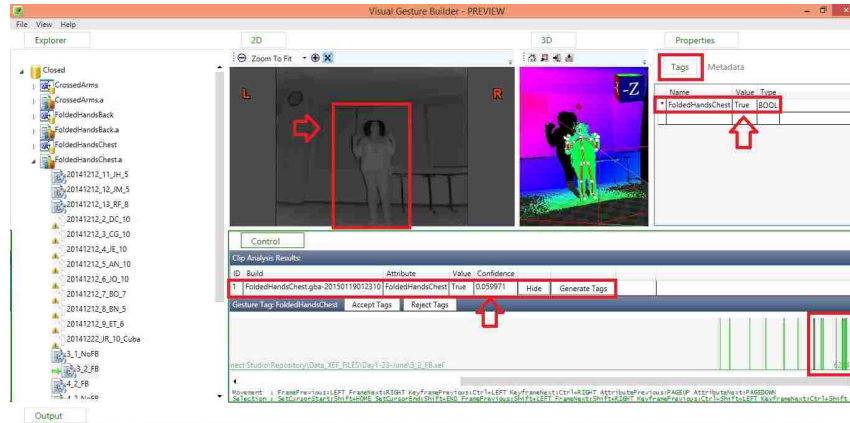
Figure 6.5: The snapshot of the VGB analysis, showing the false positive error for submissive gesture.
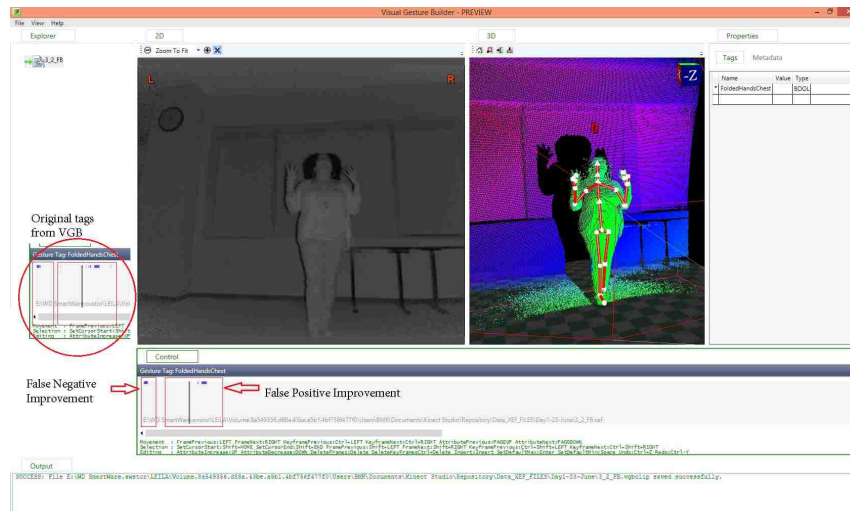
Figure 6.6: The impact of applying proximity-based approach on VGB output. The left red rectangle shows false negative, and the right presents false positive improvements.

Figure 6.7 presents the confusion matrix for the VGB application and Figure 6.8 presents

the confusion matrix for our improved proximity-based method.

56

		Predicted						Accuracy	Precision	Recall	F-score
		not-closed	unreceptive	seductive	submissive	protective	skeptical				
Actual	not-closed	39625	4	15	1055	812	28	0.857	0.876	0.954	0.913
	unreceptive	1	650	0	2	0	0	1.000	0.986	0.995	0.991
	seductive	752	0	1012	0	0	2	0.985	0.959	0.573	0.717
	submissive	847	5	0	935	992	0	0.945	0.469	0.336	0.392
	protective	4030	0	0	2	1199	0	0.889	0.399	0.229	0.291
	skeptical	0	0	28	0	0	803	0.999	0.964	0.966	0.965

F(micro-averaged)= 0.837
F(macro-averaged)= 0.711

Figure 6.7: The confusion matrix for VGB.

		Predicted						Accuracy	Precision	Recall	F-score
		not-closed	unreceptive	seductive	submissive	protective	skeptical				
Actual	not-closed	43540	4	16	1025	809	16	0.930	0.959	0.959	0.959
	unreceptive	0	650	0	0	0	0	1.000	0.986	1.000	0.993
	seductive	11	0	1029	0	0	0	0.999	0.975	0.989	0.982
	submissive	34	5	0	969	992	0	0.961	0.486	0.485	0.485
	protective	1795	0	0	0	1202	0	0.932	0.400	0.401	0.401
	skeptical	0	0	10	0	0	817	1.000	0.981	0.988	0.984

F(micro-averaged)= 0.911
F(macro-averaged)= 0.801

Figure 6.8: The confusion matrix for improved proximity-based outlier detection method.

The tested input data points are sampled from 10 different recorded clips. Five of these clips were from our initial corpus data recordings. This means that similar recordings from some participants had been used for the training classifier and these clips include all of the closed gestures at least once. Furthermore, we observed higher accuracy in classifying those clips as expected. The other five clips were selected from the actual case study (we call them test clips and some closed gestures are not observed in those clips).

Using a simple approach for outlier detection within the sliding windows, is shown in Figure 6.8 to be effective for improving the performance of the VGB classifier. The enhancement in overall F-measures, and precision, recall, and accuracy is considerable for not-closed (those gestures that are the complement of total closed gestures), submissive and protective (hands clasped in lower front) gestures in the proximity-based approach as reflected in Figure 6.8. When comparing the performance of these two approaches, accuracy alone is misleading and not sufficient,

www.manaraa.com

so we have chosen to compare precision and recall performance measures as well. As the confusion matrix indicates, there are an enormous number of data samples in the not-closed class, and the classes are imbalanced. Accuracy is a reliable performance measure when the samples in the classes are equally distributed (i.e. balanced classes).

The precision and recall values for submissive and protective gestures are relatively low in VGB (with significant improvement obtained by our approach). Considering the closeness and similarity of those two gestures, the outcome seems reasonable. This has not negatively affected our follow-up studies at all because our feedback system had only two types of classes at the implementation phase: closed and not-closed. It means that our feedback application design was performing well in recognizing all types of trained closed gestures. Furthermore, the performance of the system is very dependent on the training data size, and the subject's body characteristics (e.g. height, weight, gender, etc.), so we expect to observe different results with a different set of inputs. Cross-validation was not applicable in this case, because dividing the data into test and train clips was not possible. The recorded clips were read-only and we couldn't trim the clips for cross-validation purpose.

After the automated tagging process, all of the tagged clips were reviewed once again to eliminate any remaining error on tagging. Further analysis is reported in section 6.2.

6.1.5 Teaching Performance Measures

The recordings from the user experience sessions (audio-visual clips in mp4 format) were reviewed by two supervisors to measure the teaching proficiency aspect, and the communication aspect (as an interesting sub-domain of the teaching evaluation framework) [10].

The existing teaching frameworks [31, 79] for evaluation instruments are designed for application in real classrooms, but due to the specifications of TeachLivE as a virtual classroom, some domains may not be measurable. Therefore, we extracted the domains that are tractable in the virtual classroom setting (see section 3.2 for more details about using the Delphi method to

58

extract a measurable list of teacher evaluation). The experts were requested to review the video clips and to rate the participant's teaching, and communication performance (TP and CP scores) with a score between one and ten with provided checklist. The reference checklist is shown in Table 6.1.

Table 6.1: Applicable measurements of teaching framework [29, 31, 79] for teaching performance assessment in TeachLivE environment

Teacher Behavior Check-List	Teacher Communication Check-List
1. Communicating with students (providing clear directions, developing procedures and routines) 2. Helping students practice skills, strategies, and processes 3. Using questioning techniques (open vs. closed) 4. Responding when students are not engaged or are displaying inappropriate behavior 5. Creating an environment of respect and rapport (building relationships with students, valuing all students, understanding interests and backgrounds) 6. Managing classroom procedures (transitions, materials) 7. Demonstrating flexibility and responsiveness (lesson adjustment, response to students, persistence) 8. Giving positive feedback to students (celebrating successes, acknowledging adherence to rules and procedures) 9. Using assessment in instruction (formal and informal, formative and summative, monitoring of student learning) 10. Using discussion to help students elaborate on new information	**Nonverbal behaviors** 1. Using affirmative gestures 2. Having expressive facial expression and eye contact (e.g for encouraging students' efforts) 3. Using proximity 4. Having open body language 5. No Self-touching or unsupportive movements **Verbal behaviors** 1. Having respectful talk, active listening and turn-taking interactions with students 2. Dealing with disrespectful behaviors (including both words and actions) of students 3. Encouraging students' respectful efforts 4. Behaving equitably and responding affirmatively to questions

6.2 Observations and Findings

In case study 1, we expected the higher teaching performance scores, obtained from objective measures of teaching (TP and CP scores), to be correlated with closed postural signs (negative correlation) on the virtual classroom experience.

For each clip of the participant, the length of exhibiting closed gestures was reported as a percentage of the total session time. Figure 6.9 presents the average time of exhibiting five different closed gestures for participants with teaching experience (n=6), and students who didn't have formal teaching experience (n=8) in the study.

59
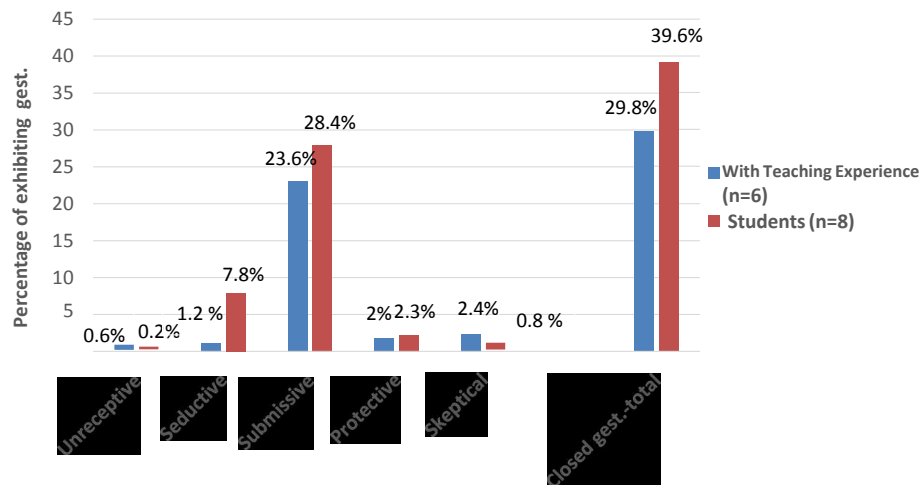
Figure 6.9: The average time of exhibiting closed gesture for two different groups in case study 1.

This bar chart indicates that experienced teachers did not exhibit closed gestures as much as student participants who had little to no teaching experiences, considering the total time of having closed gesture (shown in the last two columns of Figure 6.9). In addition, a submissive gesture (hands folded in upper front) was employed frequently rather than other closed gestures among all participants.

On the other hand, experienced teachers exhibited a skeptical gesture (hands placed on hips) more than students. As Figure 6.9 indicates, the results for all the gestures except for seductive and submissive were close to each other among both the groups of students and those with teaching experience (in-service teachers).

The correlations of TP and CP with demographics and nonverbal indicators is presented in Table 6.2. As one might expect, the number of participants (sample size) is relatively small. Within the group of participants, there is a positive correlation between some demographics information and TP, CP scores. For instance, there is a correlation between English Language Teaching (ELT) as major, having teaching experience, and age with TP score ($p<0.01$). For CP, age and English speaking proficiency level also had positive correlations with communication skills.

60

Table 6.2: Descriptive statistics and correlations using Pearson and Point Biserial (nominal) coefficients [128].

Variable	Mean/Median	SD	TP-Score	CP-Score
Demographic indicators				
Age	28.1	7.0	**0.68 (p < 0.01)**	0.51
Gender	1	N/A	- 0.10	- 0.26
Teaching	0	N/A	**0.87 (p < 0.01)**	**0.62 (p < 0.05)**
Teaching experience (yrs.)	3	4.2	**0.70 (p < 0.01)**	0.50
Grad. degree	1	N/A	**0.63 (p < 0.05)**	0.46
Major	1	N/A	0.40	0.39
English Level	1	N/A	**0.78 (p < 0.01)**	**0.67 (p < 0.05)**
Variable	Mean/Median	SD	TP-Score	CP-Score
NVC indicators (% time exhibiting):				
Unreceptive gesture	0.4	0.8	0.20	0.12
Seductive gesture	5	14.7	- 0.27	- 0.35
Submissive gesture	26.3	22.8	- 0.38	- 0.52
Protective gesture	2.1	3.5	0.13	0
Skeptical gesture	1.5	2.3	0.44	0.19
Closed gesture - total	37.1	26.7	-0.40	**- 0.62 (p < 0.05)**

Note. N = 14. Men coded 1, women, 2; teaching variable coded 1 for teachers, the rest coded 0; the years of teaching experience coded 0 for subjects with no teaching experience; graduate degree coded 2, undergrad coded 1; ELT coded 2, the rest coded 1 in Major variable; and advanced English level coded 2, intermediate, 1.

A negative correlation with the percentages of exhibiting closed gestures (i.e CGP or Closed Gesture Percentage) was reported with CP ($p<0.05$). Furthermore, CGP had negative correlation with teaching proficiency. The negative correlation with some of the closed postures (for submissive and seductive gestures) also has been indicated with weak significance. The other types of gestures had non-negative correlations with TP and CP.

To justify that, we may need to look at the data distribution and the sparsity of other gestures among the subjects. In addition, some closed gestures (such as skeptical) might be interpreted more negatively and defensive by the audience in comparison to other closed gestures. It means that all

61

closed postures may not impact equally in the interpersonal communication.

6.3   Discussion and Limitations

In this case study, we extended the TeachLivE system to include a semi-automated annotation procedure through which posture and performance data are collected. We found that, among the detected closed gestures, submissive and seductive gestures were the most frequent ones exhibited by participants. In addition, since English was not the subjects' first language, they frequently used beat gestures to keep the rhythm and meaning of speech [62, 82]. This observation could be inferred only from exploring the correlation of gesture and auditory features together as multimodal analysis. There was significant correlation between a subject's teaching experience, age and TP/CP.

There may be some confounding factors involved in this case study. First, all of the subjects except one were new to the TeachLivE environment, and the experiment was conducted in a single-session. The second confound was related to their English proficiency that we mentioned earlier in section 6.1.1. The other limitation was the small number of participants because the sample was selected from a group of convenience. The observation also may be affected by personal habits, because people do not have similar body language due to cultural background and other conditions in their interactions. In addition, all closed postures may not impact equally in the interpersonal communication. Finally, some closed postures may correlate with the teaching context.

In summary, despite our accomplishment in semi-automated gesture recognition, this case study had some uncontrolled conditions. It was the first experiment we conducted and we learned a great deal from this experiment for the subsequent steps reported in the next two chapters.

CHAPTER 7: AUTOMATED GESTURE RECOGNITION FOR TEACHING ASSESSMENT - CASE STUDY 2

In this chapter, we describe the details of case study 2. This study was conducted to explore the impact of the proposed real-time feedback application on body-language thoughtfulness [11, 13]. This study was informed by the previous studies with the specific attempt to control confounding conditions of the prior work. For this study, we recruited from a specific group of subjects, with English as their first language, and all of whom had seen the TeachLivE environment before the experiment. We also limited the teaching topic to be one, provided by a teaching plan.

Furthermore, we changed our evaluation methods from subjective and very broad teaching competency measures, reported by subject-matter experts, to self-reported questionnaire responses from participants about usability and system evaluation measures. Details of the study are described in the following sections.

7.1 Methodology

As mentioned earlier, this research evolved based on the existing literature expressing the importance of open body gesturing in successful interactive teaching. We are interested in detecting these closed gestures and reminding the trainees about their closed body language stance. The process of creating the closed gesture recognition database was described in section 6.1.4.

The current developed system is designed to provide feedback either in the form of visual or haptic signals (vibrations) any time that the participant exhibits a closed stance. The case study reported in the following section is based on visual feedback; we tested the vibration feedback method using the Myo armband[1] in a pilot study and the results were consistent with and

---

[1]https://www.myo.com/

as promising as the visual feedback application (See section 8.2 for more details). The following information in this chapter describes the details of the system design and the user study details for the usability evaluation.

7.1.1   *Participants*

In this study, $30(24F, 6M)$ participants were recruited. Only those UCF College of Education and Human Performance (COEHP) students who had experienced the TeachLivE rehearsal environment before the experiment were eligible to be recruited. Participants were ones over age 18 years from the affiliated university community and each participant had a native-level English proficiency, with some teaching experience to be able to teach and interact with virtual students in the simulator.

All the participants were from the students who were registered for a classroom management and strategies undergraduate course in the TESOL (Teaching English to Speakers of Other Languages) program in summer 2015 taught by a faculty member in UCF's COEHP. All the students had an assignment using the TeachLivE system prior to the experiment. The participation in the experiment was mandatory for all of the registered students as a classroom activity because of its perfect overlap with the course objectives. Students did not get compensation or extra credit for their participation.

7.1.2   *Study Procedure*

The case study was a 2×1 counter-balanced within-subjects study, which means that all the participants attended both of the study settings, but the order of their sessions was flipped for half of the study participants.

Individuals were expected to spend approximately 30 minutes for the recruitment (two 7-minute sessions for the teaching plus three 5-minute intervals for questionnaires). We randomly divided the participants into two groups (A, B) based on their scheduled recruitment time. The

64

study design and participant assignment is shown in Figure 7.1 for both of the groups.
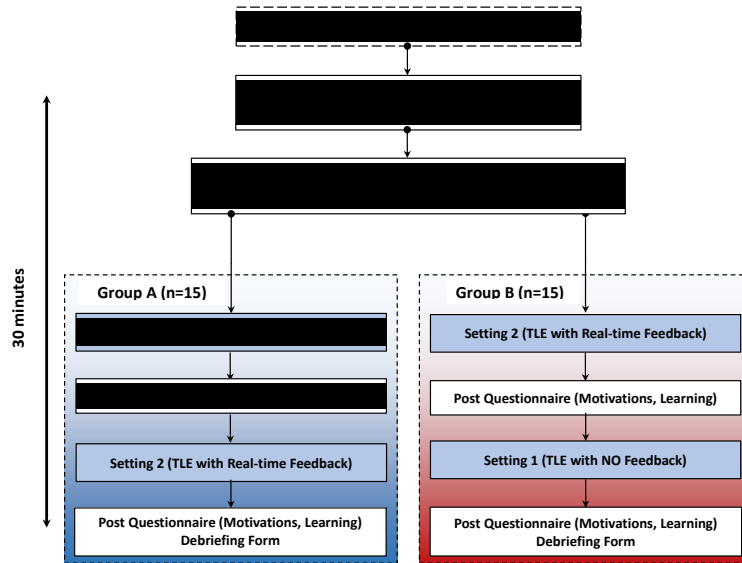
Figure 7.1: Overview and participant assignment of the case study 2.

7.1.3 Apparatus

The selection of hardware and software used in data collection are based on the following technical requirements:

- **Full body tracking.** Full body of the participant (25 joints in skeletal stream from the Kinect V2) teacher is recorded for body language and pose recognition.

- **Teaching session videos.** This is similar to case study 1, and includes the view of participant with the virtual classroom scene during the whole session in avi format. This recording is done in an observer workstation with a window recording application - Bandicam[2]. It is compatible with the ReflectLivE application for both offline and online video assessment.

---

[2]https://www.bandicam.com/

The user experience room and design setting for this case study was similar to Figure 6.1. The user experience had some changes, because the automated feedback was provided. The details of the user experience are shown in Figure 7.2.
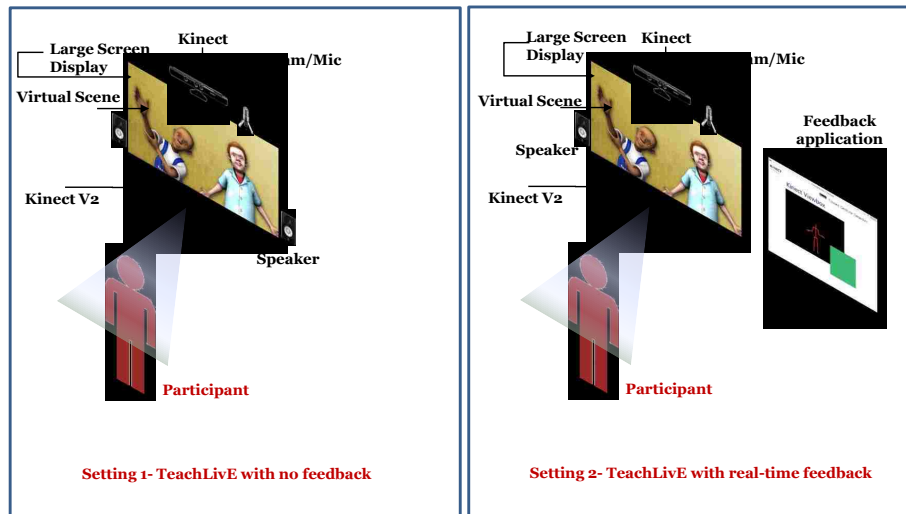
Figure 7.2: The user experience view for case study 2. All the participants experienced both of the settings.

The room was equipped with required laptops and wired network connections for the experiment. Three laptops were used for the experiment. One was for an observer workstation, one for the skeletal recording and visualized feedback, and the last one for the participant's workstation. The virtual classroom's students were high-school avatars. Their level of misbehavior was set to zero, which means no misbehavior in the classroom-side. The interactors (the persons who control the virtual avatars) for all of case study sessions were also asked to give consistent performances in order to minimize the inconsistencies that may occur because of interactor changes.

Next to the TeachLivE classroom projected on a large TV display, we posted main questions and contents of the teaching plan on a board as reference for participants. We asked them to use the reference board instead of their given teaching plan as the gestures of participants were affected while they were holding an object (even a sheet of paper) in their hands.

66
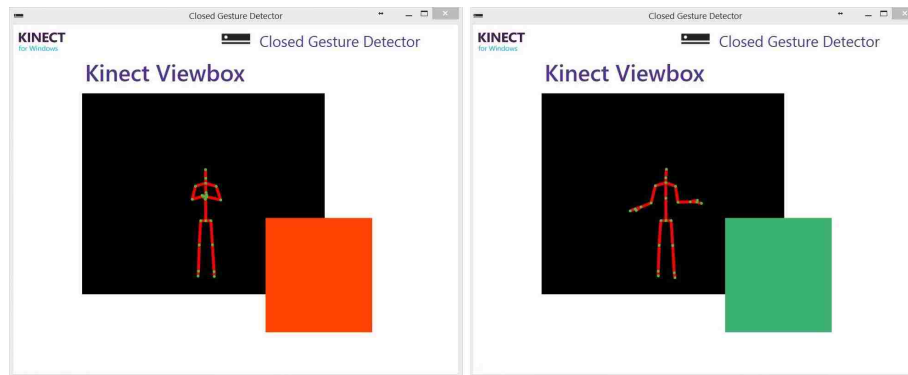
Figure 7.3: The proposed visual feedback application snapshot for two postures.

7.1.4 Feedback Application

The feedback was in the form of a visual prompt each time the participant exhibited a closed, defensive stance. We used the Kinect V2 and the closed-gesture corpus (see Chapter 6 for more information about the corpus) as input to this application. The feedback application was running independently from the TeachLivE program. Figure 7.3 presents the feedback user interface in more details.

In this Figure, Kinect Viewbox shows the skeleton view of the participant in front of the Kinect sensor and the smaller window in the corner is the feedback visualizer. Any time that the participant exhibits one of the closed gestures (submissive gesture, or hands clasped in front of the chest in the left-side snapshot) the color changes from green to red as a visual trigger.

7.1.5 Full-Body Tracking Data

The collected full-body tracking data from the participants was processed to extract a higher level feature. The calculated feature value was the percentage of time that a subject exhibited closed gestures in the recorded clips of the study sessions. We call this variable CGP, standing for Closed-Gesture Percentage exhibition. We use this variable to report the effect of our feedback

67

application on the participant's body language changes. Equation 7.1 denotes the value of CGP for a given clip M.

$$CGP_M = \left(\left. \sum_{t=Start\_Frame_M}^{End\_Frame_M} \bigvee_{i=1}^{\# \, Closed \, Gestures} x_{i_t} \middle/ End\_Frame_M - Start_{Frame_M} + 1 \right) \times 100$$

(7.1)

In this equation, the label of gesture i in frame t is represented as x_{i_t}. If the gesture recognition function for gesture i, in frame t recognizes the gesture from the associated body-joint data, then $x_{i_t} = 1$, otherwise $x_{i_t} = 0$. $\vee$ is the logic or operation. The divisor is the total length of clip M in # of frames and assuming the FPS is fixed for the clip, CGP value represents a percentage of time in which at least one type of closed gesture was identified.

CGP is the dependent variable that we expect to be correlated with a session setting (with or without feedback) as the independent variable. The hypothesis is that regardless of the group assignment (A or B), the effect of the setting is considerable on CGP.

7.1.6 *Questionnaires*

In the pre-questionnaire, participants were asked to indicate their age, gender, ethnicity, major and highest level of education, teaching experience and related teaching topics, and finally their particular experience with the TLE teaching rehearsal environment. The pre-questionnaire is shown in Appendix D.

Participants completed a questionnaire soliciting their evaluation of the user experience and learning perception in each of the sessions. The questionnaire included both 5-point Likert scale and open-ended questions. User experience was evaluated by four questions on experience naturalness, interaction intrusiveness, motivation and likelihood for future participation. Learning

68

perception was evaluated by two questions on improvements in the understanding of teaching strategies and comparison of the proposed application with traditional teacher training approaches. Open-ended questions include the participant's experience in learning, potential suggestions for improvements, and any additional comments. The post-questionnaire form for this user study is shown in Appendix E.

We did not completely inform our subjects in advance about the true nature of our research in order to avoid study bias. Thus, the debriefing form was submitted to the participants after recruitment since our user study involved this mild deception. Appendix F presents the debriefing form for case study 2.

7.1.7 Teaching Plan

The teaching plan is designed to enhance science literacy and is aligned with disciplinary core ideas and cross-cutting concepts from the next generation science standards [125], as well as the common core standards for literacy in science [59]. The lesson has been validated and field-tested in high school Biology classrooms as part of a larger module from the NIH curriculum supplement series "Using Technology to Study Cellular and Molecular Biology" [3]. The title of the teaching plan is: "What is technology?" In this teaching plan, some questions were suggested as topic discussions with virtual members of the classroom. The participant, as the teacher, was expected to ask virtual students about their personal definition of technology, what does technology do for human beings, what technologies are available in the room, and the impact of technology on diseases in humans. The participant was free to discuss what they prefer within the topic of the teaching plan. Appendix C shows the teaching plan for this case study.

7.2   Results

We report the results for case study 2 in two sections, based on closed gesture employment and self-reported usability analysis as follows.

7.2.1   Closed Gesture Evaluation

To evaluate the impact of our proposed feedback application on body language mindfulness, we calculated CGP (see equation 7.1) for 60 recorded clips from 30 participants. The box-plot in Figure 7.4 presents the distribution of closed gesture percentage (CGP) between the two groups during the study.
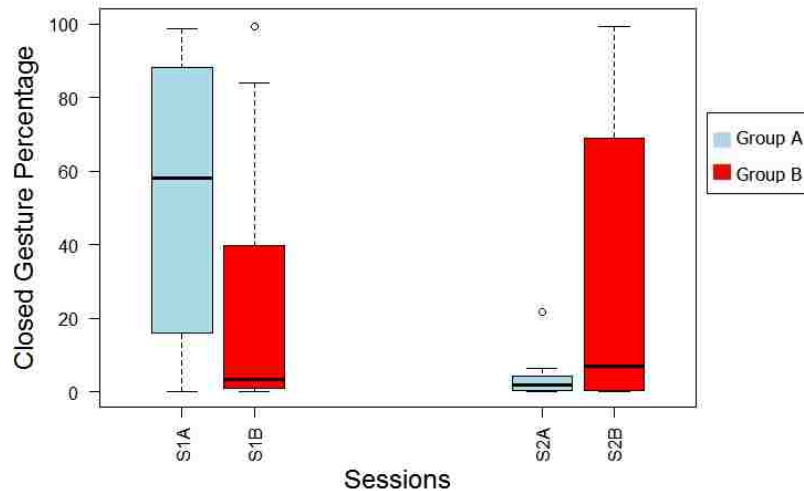
Figure 7.4: Medians and interquartile ranges of CGP exhibition in two sessions among two groups A (n=15), and B (n=15). Circle represent outliers.

Figure 7.4 shows some of key findings from this study. It presents the wide range of closed gesture employment for group A and B in both sessions (excluding session 2 for group A). It also indicates the average of CGP median for group B participants is lower than group A.

70
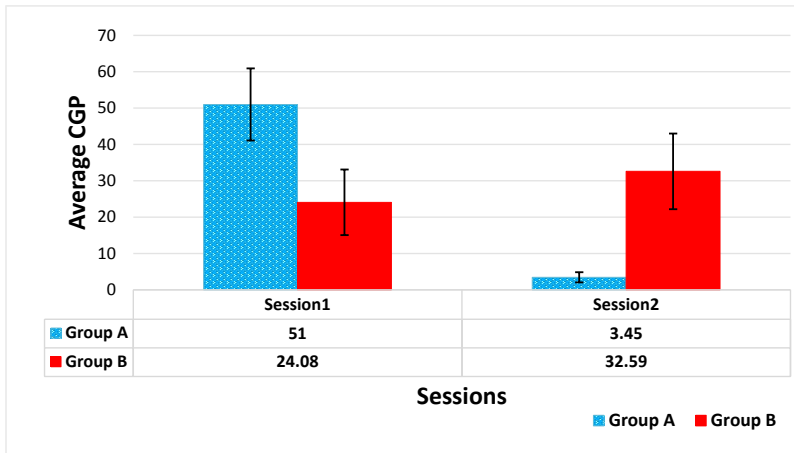
Figure 7.5: Average time of the closed-gesture employment (%) among all of the participants in two sessions.

Considering 3.5% and 6.9% as medians for two sessions for group B and 58%, 1.8% for group A, the average median values were 5.24% for group B and 29.95% for group A. We also calculated the average and standard error for the related closed gesture variable for two groups. Figure 7.5 shows the results.

Figure 7.6: Analysis of variance for case study 2.

Furthermore, according to the analysis of variance shown in Figure 7.6, there was a significant effect of setting (feedback vs. no feedback) on CGP ($F_{(1,28)} = 13.66, p < .001$) in this

study. The group effect was not statistically significant ($F_{(1,28)} = 0.041$, ns). This means counterbalancing was effective; i.e., any learning effect that might have occurred for the B group was effectively offset by a similar and opposing learning effect for group A. The $Setting \times Group$ interaction effect also failed to achieve statistical significance ($F_{(1,28)} = 0.021$, ns) which is a promising finding.

As the results reflected in Figure 7.4 and Figure 7.5, the hypothesized statement is supported for the participants of the study ($F_{(1,28)} = 13.66, p < .001$). The average time that all of the participants in group A exhibited closed gestures reduced significantly from their first session to their second session. As expected, CGP was increased in the second unaided session for group B (median= 6.9%). Most interestingly, the results indicated non-significant correlation between group assignment and CGP, which also supports this successful study design.

7.2.2 *Post-Questionnaire Evaluation*

To report the results from the post-questionnaires, we divided six 5-point Likert scale questions into three groups (two questions each). These groups are naturalness and non-intrusiveness, motivation for future use and learning perception (see section 7.1.6 for more details). We analyzed the 360 responses from 30 participants ($30 \times 2 \times 6 = 360$) in Figure 7.7.

Figure 7.7 presents valuable information about case study 2. In general, we received very positive feedback from the participants (regardless of their assigned groups) as the right-skewness of the stacked bar charts indicates this point. Interestingly, participants' responses indicated that their experience in setting with visual feedback was natural and non-intrusive (62.5% versus 50% for no feedback setting). This is very encouraging to us from a usability aspect in system design. The participants were also very motivated to participate in future studies. For the setting with no feedback, we had more responses as "neutral" or the "same" versus the setting with visual feedback that indicates more positive responses such as "motivated", and "better learning perception".

72

Figure 7.7: Post-questionnaire analysis for case study 2 in three main categories from 30 participants.

7.3 Conclusion and Future Research

In this chapter, we reported a case study to evaluate the performance of developed feedback application for nonverbal communication skill training. We used the Microsoft Kinect V2 sensor and its full-body tracking data stream, as well as ensemble classifiers to develop our real-time gesture feedback application.

73

The results from recorded skeleton data and post-questionnaire responses from the participants indicated the positive impact of informed body language and gesture in communication proficiency. That said, the majority of the participants were very motivated to use the gesture feedback application for their rehearsal sessions. The participants also mentioned their learning experience was improved; some participants reported that their experience with the system was better than the traditional learning resources.

Conducting a new experiment with a representative number of participants to evaluate the haptic (vibration) feedback method versus visual cues is one of the future goals of this research project. Furthermore, we wish to examine changes in communication development over time. We observed the body language changes for three subjects for a four- month period with sessions occurring every two months. The findings from those conducted pilot studies are presented in Chapter 8.

CHAPTER 8: PILOT STUDIES FOR FUTURE RESEARCH

In this chapter, we introduce three pilot studies that we conducted in the 2015-2016 academic year. Data collection and analysis details for all of the pilot studies are similar to case study 2. Three participants from case study 2 $(2F, 1M)$ volunteered to participate in two pilot studies I and II. For the third pilot, we asked TeachLivE collaborators at College of Education $(N = 6, 4F, 2M)$ to participate. To summarize the time-line and study settings, we provide Table 8.1 for more information. The details of each pilot study is presented in the following sections.

Table 8.1: Time-line and settings details for pilot studies.

Semester	Month	Setting			Legend:
Summer 2015	Jul.	**Setting A**	**Setting B**		**Pilot *I***
Fall 2015	Sep.	**Setting B**	**Setting A**		**Pilot *II***
	Nov.	**Setting A**	**Setting B**	**Setting C**	**Pilot *I&II***
Spring 2016	Feb.	**Setting D**	**Setting E**		**Pilot *III***

8.1 Pilot I: Gesture Changes Over the Time

In this phase of data collection, we examine changes in communication development over a semester. The objective is to observe the behavior of participants from the summer 2015 experiment in fall 2015.

8.1.1 Participants

Among the pool of summer 2015 study participants, three undergraduate students $(2F, 1M)$ volunteered to participate in the two follow-up pilot studies. They were compensated with $10 gift certificates for their participation.

8.1.2 Study Procedure

The study settings was similar to case study 2. However, it had a $2 \times 3 \times 1$ design, because we conducted three trials in total.

The participants were all from group A in case study 2, and they had experienced real-time gesture feedback application in their second teaching session (July 2015). Hence, we flipped the settings in their next interactive experience. The experiment had two sessions with two-month intervals and the total period of the pilot was four months and a half, as Table 8.1 presents.

8.1.3 Observations and Discussion

We calculated the mean rate of CGP for each participant on each session of the experiment, and presented the results in Figure 8.1.
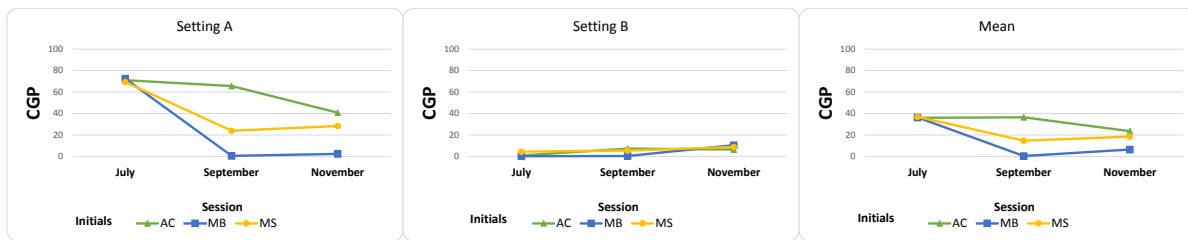
Figure 8.1: closed-gesture employment (%) among the participants in three sessions (for two settings A & B, and their mean rate) over a semester period.

The results show reasonable decrease in closed gesture production among the participants on the pilot study, especially in setting A, and from July to September. The informed body language impact is considerable in this period of time. For setting B, we observed slight variations, but CGP value remained low over a semester.

Our eventual goal is to export this posture information into the ReflectLivE summary reports. As presented in section 2.3 and Figure 2.5, ReflectLivE presents a summary of the trainee's interactions including some automated and hand-coded tags. There are some limitations however

76

that we need to address, including system incompatibility. Further details about our future research on extending the ReflectLivE system is presented in section 9.3.3.

8.2   Pilot II: Haptic Feedback Mechanism

In this pilot, we worked on the form of feedback application. We decided to change the user interface for feedback application, by exchanging visual prompts with haptics feedback (vibration using Myo[1]) in order to reduce the intrusiveness of the application, and to increase the effectiveness and naturalness of the interaction.

8.2.1   Participants

We tested the vibration feedback application with the same participants $(N = 3, 2F, 1M)$ who took part in the pilot study I.

8.2.2   Study Procedure

The study was conducted as an extension of pilot I. It was a single session observation on November. As shown in Table 8.1, the pilot II had a 2×1 design.

In setting C, we asked the participants to wear the Myo armband. Myo uses Bluetooth to send and receive commands, including vibration (short and long) from the connected workstation. We used short vibration commands instead of visual prompts for closed gesture feedback in setting C. Figure 8.2 presents a snapshot of a participant in setting C with a detailed picture of the Myo armband.

We asked participants to evaluate their experience after each session. They completed the same post-questionnaire that was used in case study 2 (shown in Appendix E).

---

[1]https://www.myo.com/
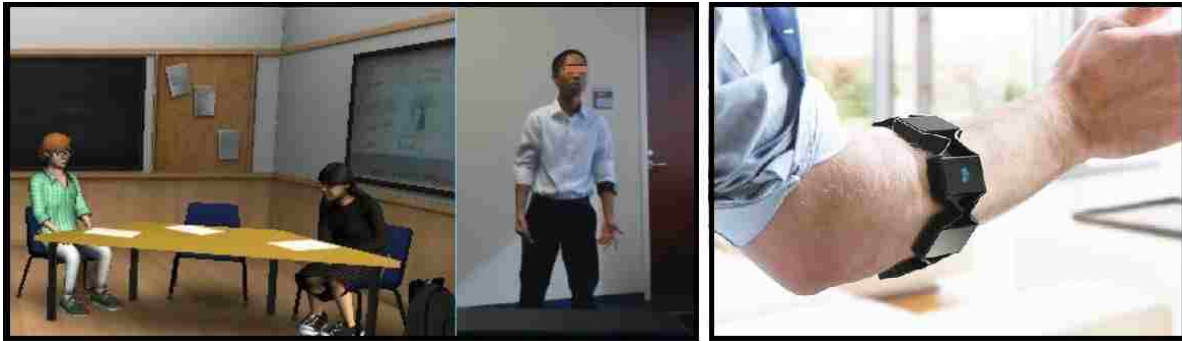
Figure 8.2: A closer look at setting C and the Myo armband.

8.2.3   Observations and Discussion

We analyzed the recorded full-body tracking data from each participant. CGP values for setting B and C are presented in Figure 8.3 for three participants in the pilot.
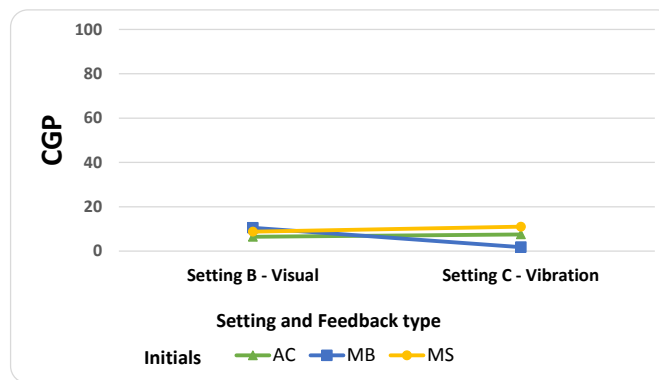
Figure 8.3: Closed-gesture employment (%) among the participants for two types of vibration and visual feedback.

As Figure 8.3 indicates, the closed-gesture employment rate is very low and almost the same in both of the settings among the three participants. There was not a significant difference in closed gesture employment for two settings, as one might expect. This likely arises because of the learning effect over the consecutive sessions of the experiment.

For post-questionnaire analysis, we used the same method as section 7.2.2 to report the participants' evaluation. Figure 8.4 indicates the results from three participants in the pilot study.
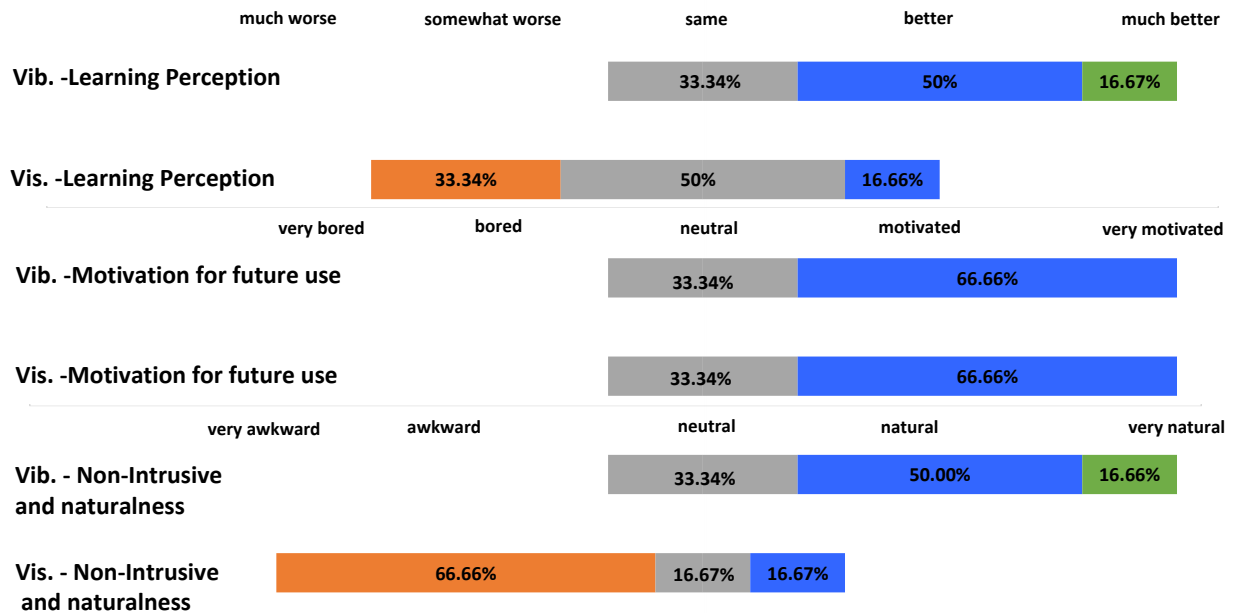
Figure 8.4: Post-questionnaire analysis for pilot II in three main categories from three participants.

As the analysis indicates, all three of the participants were more interested to use the vibration method for gesture feedback rather than visual prompts. One of the participants reflected his evaluation as follows:

"I felt as the arm band was more helpful in providing feedback. It is a lot easier to tell when I am doing something wrong when the arm band vibrates as opposed to trying to look at the Kinect screen while talking".

For future studies, we would like to conduct a counter-balanced study to minimize the learning effect. In these pilots we were not able to do counterbalancing because of the small (and odd) number of participants. We intend to conduct the future studies with a new pool of participants, who haven't experienced the visual feedback application before the experiment.

8.3   Pilot III: Aggressive vs. Non-aggressive Teaching Role

In this pilot, we investigated the impact of a teaching role on body language. We designed a pilot study with two circumstances:

- Effectively getting the virtual students on task.

- Ineffectively attempting to get the virtual students on task (aggressive teaching).

The purpose of this pilot is to understand changes in body language in the mentioned situations and scenarios. For simplicity, we call these settings as non-aggressive and aggressive.

8.3.1   Participants

We had six participants $(4F, 2M)$ from the TeachLivE research team who volunteered to take part in pilot study III. Ages ranged from 30 to 56 years (mean 41.3, SD 8.93). Five of the participants were graduate students in the College of Education and one of them was a visiting faculty member in the Exceptional Education program. All of the participants had more than eight years of professional teaching experience. They all had seen the TeachLivE before conducting the study. Half reported that they also taught in the TeachLivE environment.

8.3.2   Study Procedure

The study was a 2×1 design with no counterbalancing. The participants completed a pre-questionnaire (shown in Appendix D) before their teaching sessions.

We asked the participants to use the technology teaching plan in two four-minute sessions. The first session of the study (setting D) was the similar to setting A in case study 2 (TeachLivE with no feedback) and the purpose of this session was to address effective teaching, i.e. the teachers were asked to play a non-aggressive teacher's role in setting D. In the second session (setting E)

80

they were asked to do role-playing for an aggressive teacher with ineffective attempts to get the virtual students on task.

8.3.3   Observations and Discussion

We calculated CGP and the most frequent gesture type (dominant gesture) for all sessions. Figure 8.5 presents the gesture changes among the participants of pilot study III.
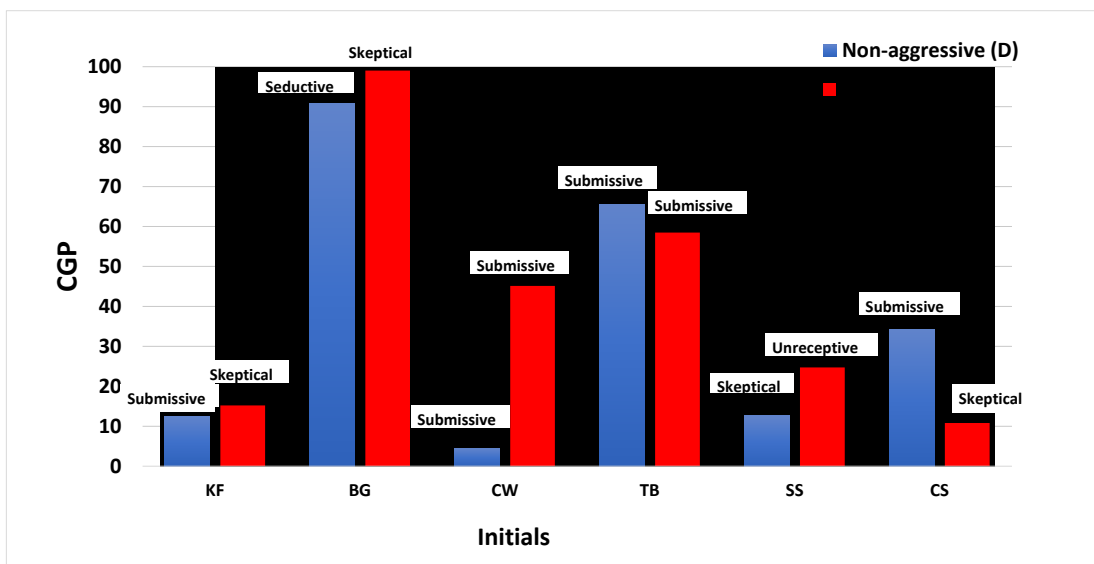
Figure 8.5: Closed-gesture employment (CGP) among six participants for two settings (D & E).

As Figure 8.5 indicates, participants exhibited more skeptical postures in aggressive role scenario (setting E). Interestingly, submissive posture was the most frequent one in non-aggressive settings. These two findings suitably align with our observations from previous studies (case studies 1 & 2).

We observed that two participants (CW and CS) had more changes in their body language and closed gesture employment in two settings. CW exhibited more closed gestures and CS exhibited fewer closed gestures in their aggressive teaching experience.

81

CHAPTER 9: CLOSING REMARKS

In this chapter, we describe research outcomes and conclusions, limitations and the potential directions for future research.

9.1   Conclusion

In this thesis, we described the TeachLivE system as our test-bed simulation environment, the importance of recognizing nonverbal communication behaviors in this context, some means of providing semi- and fully-automated event tagging associated with nonverbal messaging, and a series of preliminary studies developed to inform the research.

We specifically focused on gesture recognition and different types of non-intrusive feedback provision techniques (including visual and haptics) for closed-gesture stances exhibited by trainees.

Although the context for our work was the TeachLivE system, the proposed feedback method reported in this thesis is applicable to many other (rehearsal) applications. For instance, it could be applied to applications associated with public speaking anxiety, social skills training, police de-escalation training, and almost any other human-centered activity. Further information about some of these potential applications is presented in section 9.3.4.

9.2   Limitations

We addressed most of the limitations in each case study. Recruiting the representative number of participants was a difficult process that adversely affected our research agenda. For example, in some studies, we were not able to do counter balancing to reduce the learning effect because we didn't have a sufficient number of participants.

To clarify, it was difficult to find a pool of participants from the College of Education who

were enrolled in a relevant course. Some of this arose from the hesitancy of professors to change their syllabus to include the studies, especially if they were not yet convinced of the advantages to be be gained by their students. We believe that the positive outcomes of the studies reported in this thesis will alleviate that issue in future studies. Additionally, as processes, including question-naires developed here, have already been approved by our IRB office, the approval cycle should be greatly reduced for future experiments. The entire TeachLivE team actively collaborated in the data collection (thanks to all of them), and we are working on encouraging course instructors to use TeachLivE system for research purposes. Therefore we expect to have much greater success in the participant recruitment process for future research studies.

There were also some hardware and software limitations that affected the data collection procedure. Using a high-speed SSD hard-drive could be helpful to store body tracking data and reduce frame loss, but we were not aware of this in advance. We expected to progress in vocaliza-tion analysis as well; however we faced serious technical issues. The recorded audio-visual clips didn't have the minimum quality for speech analysis. Therefore, prior to starting the data collec-tion, we needed to establish a structured, robust and synchronized framework for data capturing and storing. This is a very common problem in multimodal, multi-interface projects though.

9.3   Future Research

Several directions of future work are promising as follows.

9.3.1   Dynamic Nonverbal Components

In this work, we mainly investigated discrete postures of the participants in teaching set-tings. We have done a formative study (section 5.2) for identifying and interpreting the dynamic gestures of the teachers while explaining a topic to virtual or real students [12] based on previ-ous work from Alibali et al. [7]; however, we did not investigate dynamic and continuous gesture

83

recognition. Based on the work of Alibali et al. [7], teachers exhibit more gestures when they teach a new topic, while they introduce an abstract concept, or in response to students' questions. Therefore, exploring the teacher's gestures and her associated conversation while exhibiting the gesture is one of the interesting approaches for future research, especially for beat or iconic gestures as the literature indicates.

We also are looking forward to determine some predictive multimodal features that indicate the nonverbal behaviors of teachers as multimodality is an integral part of the teaching process [138].

9.3.2   Different Forms of Feedback Provision

While this work touched very briefly on the different methods of feedback provision, it did not provide strong evidence about their performance since the size of participants for the conducted pilot study was relatively small. Hence, conducting an experiment with a representative number of participants to evaluate the haptic feedback method versus visual cues is one of the objectives of this research project over time. Another candidate sensor for feedback provision is the Microsoft Hololens[1]. We can investigate the effect of visual (as an augmented display) and auditory feedback for the trainees with the Microsoft Hololens or similar device.

9.3.3   Extending the ReflectLivE System

Automating the process of event tagging to assist in reflection is an eventual goal for this research. While ReflectLivE provides a tool for event annotation and even supports automated objective analysis for proximity (tracked movement to the vicinity of specific avatars) and talk time (relative participant vs avatar talk), the tool does not yet include any recognition of body postures (e.g., open versus closed), nor has it yet to be refined to a state where it fuses utterances with

---

[1]https://www.microsoft.com/microsoft-hololens/en-us

84

gestures or evolves automated or suggested tagging based on observed tags from subject-matter experts (SMEs). We are looking forward to developing a basis for recognizing meaningful events and their associated tags. Such assumed tags would then be used to "semi-automate" tagging by giving raters (typically a blend of SMEs and trained but less sophisticated raters) hints as to when and what events they might tag. The confidence levels of these assumed correlations would rise or fall based on how often the hints are accepted or rejected, with subject-matter experts contributing heavier to the weight shifts at first, and the weightings of others changing based on their performance relative to SMEs (assumed ground truth). As confidence levels rise sufficiently, the goal is to "automate" these highly correlated observations to their associated event tags. Of course, any tagging can be edited or even rejected as part of an overall support system for event-based reflection.

9.3.4 *Practical Examples of the Virtual Rehearsal Environment*

In this section, we describe five applications of TeachLivE and its underlying AMITIES infrastructure for numerous human-centered interactions and training purposes as follows.

- **Training law enforcement personnel in de-escalation skills.** The nature of law enforcement is that its personnel are often placed in intense and dangerous situations, and that they must deal with an extremely diverse public with differing generational, cultural, educational and economic backgrounds. Moreover, officers are under a constant microscope from cameras (cell phones, public and private security, and body-worn). In one sense this impacts public safety but, in another sense, it relates to the officer's health and wellness. High stress situations that are not deescalated cause the officer to enter a stage of hyper-vigilance and retain that state, even when at home. This is clearly a health issue for the officer, but it is also a health and wellness issue for his or her family. Examples of tags specific to this application include events associated with positive/negative facial and body poses, and re-

85

spectful/disrespectful vocal intonation and volume.

- **Health issues focused on youth with intellectual disabilities.** The goal here is to train physicians and other caregivers to deal with the unique issues faced by these young women with intellectual disabilities who are becoming sexually mature. Challenges here start with helping these girls to deal with the confusion and fear that often arises with the onset of menstruation. Even more complex issues surround the potential of their becoming the targets of sexual predators who seek to take advantage of their vulnerability. Our goal is to use our avatar-based environments to provide an opportunity for medically-trained individuals to safely practice their skills at interacting with members of this population and, as appropriate, their guardians (parents or others). Examples of events specific to this application include those associated with caring/indifference through auditory response and/or body/head pose, or interest/disinterest in the patient's needs based on visual and verbal focus.

- **Professional development for teachers of inclusive classrooms.** Children with special needs (cognitive, e.g., autism or intellectual disorders; or physical, e.g., hearing or vision impaired) often are isolated in classrooms, especially where group projects are concerned. The lack of training provided to teachers to support students with special needs is well-documented, as are the barriers to students' full participation, which results in a reduced "opportunity to learn" relative to their peers. Examples of tags specific to this application include events associated with how objects are described, e.g., verbally for visually-impaired students, and by images and pointing for hearing-impaired students.

- **Protective strategies for college students.** Students entering college directly from high school often face many challenges. They are blessed with new opportunities, both academic and social, and with the responsibilities and demands that this freedom of choice entails. The goal is to allow these young adults to understand potential consequences of action and inaction that often arise in a college setting. The hypothesis is that rehearsing and reflecting on

86

their actions in a safe environment will create heightened awareness of risks and responsibilities, and lead to employing more effective protective behaviors for themselves and others. Examples of tags specific to this application include events associated with empathy or lack of it to the needs of a fellow student.

- **Automated behavior descriptors for clinical diagnoses.** Motivated by a large body of research that examined the relationship between nonverbal behavior and clinical conditions [16, 19, 49, 102, 115, 131], we have previously developed automatic behavior descriptors to identify correlates with depression and/or PTSD [45, 115, 116, 126]. The proposed research can provide correlates and tools that expand this work and lead to automated annotations that assist diagnoses for depression and other clinical conditions identifiable through observable but often subtle speech, gestures and facial expressions.

9.4 Final Remarks

The primary technical contributions of this thesis are represented by its employment and adaption of machine learning techniques to automatically recognize a class of gestures, providing real-time feedback to people participating in classroom rehearsal in a virtual environment. While the specific focus here was on closed gestures, the techniques and user interaction paradigms developed apply to feedback associated with any gesture classification. In addition to technical contributions, this thesis presents a road map for future research, ranging from feedback modalities to automated event tagging.

APPENDIX A: OUTCOME LETTER

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
          FWA00000351, IRB00001138

To: **Roghayeh Barmaki, PhD** and Co-PIs: **Lisa A. Dieker, Micha**

Date: **July 28, 2015**

Dear Researcher:

On 07/28/2015 the IRB approved the following modifications / human participant research until 07/07/2016 inclusive:

Type of Review:	IRB Addendum and Modification Request Form
Modification Type:	Some questions in the Pre-Questionnaire have been modified to improve clarity. A revised Pre-Questionnaire document has been approved for use.
Project Title:	Study of the Correlation of Body Gestures to Presence and Learning in a Virtual Learning Environment
Investigator:	Roghayeh Barmaki, PhD
IRB Number:	SBE-15-11413
Funding Agency:	
Grant Title:	
Research ID:	N/A

The scientific merit of the research was considered during the IRB review. **[NOTE: Because this study was not approved before the IRB expiration date, there was a lapse in IRB approval from X/X/20XX to the new approval date above.]** The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study.   All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 07/07/2016, approval of this research expires on that date. Use of the approved, stamped consent document(s) is required. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

[signature: Joanne Muratori]

Signature applied by Joanne Muratori on 07/28/2015 04:58:59 PM EDT

IRB manager

APPENDIX B: INFORMED CONSENT

Permission to Take Part in a Human Research Study

Informed Consent

Experiencing a Virtual Learning Environment

Research Site: University of Central Florida

Principal Investigator(s): Roghayeh Barmaki, Doctoral Student, Computer Science, UCF

Co – Investigators: Michael Hynes, Professor, School of Teaching, Learning, and Leadership, UCF
                    Lisa A. Dieker, Professor, Child, Family, and Community Sciences, UCF

Faculty Supervisor: Charles E. Hughes, Professor, Electrical Engineering and Computer Science, UCF

Dear Participant:
Thanks for your participation. You are being invited to take part in a research study that will include teacher rehearsal because you are currently   participating in a TeachLivE – based study.

The person doing this research is Roghayeh Barmaki of the Computer Science Division, EECS, UCF. Because the researcher is a doctoral student, she is being guided by Dr. Charles E. Hughes, a professor in Computer Science Division and TeachLivE Principal at the University of Central Florida.

What you should know about participating in a research study:
- Someone will explain this research study to you.
- A research study is something you volunteer for.
- Whether or not you take part is up to you.
- You should take part in this study only because you want to.
- You can choose not to take part in the research study.
- You can agree to take part now and later change your mind.
- Whatever you decide it will not be held against you.
- Feel free to ask all the questions you want before you decide.

Purpose of the research study: This study seeks to improve teaching effectiveness in classrooms by evaluating communication skills and related teaching skills of participants in a virtual classroom setting.

What you will be asked to do in the study: During this study, you will be asked to teach an introductory biology topic (the lesson plan will be provided to you prior to the interaction). You will teach this topic to the virtual classroom two times. Before the first interaction, you will be asked to complete a pre-questionnaire form. After each teaching interaction, you will be asked to complete a post questionnaire form.

Location: The study will be conducted at UCF main campus.

Permission to Take Part in a Human Research Study

Time required: This study will take place over the course of two sessions lasting approximately 10–15 minutes each. You will not be evaluated on your performance for employment purposes. The data that is collected on your teaching behaviors while being observed is related to the study.

Videotaping: You will be audio and video recorded during this study. In addition, your body movements and body poses will be recorded. While movement and poses may be shared with others through publication, no identifying information, including names and faces, will be shared beyond the researchers associated with this project. If you do not agree to be recorded, you will not be able to participate in the study. If you have any questions please discuss this with the researcher. Collected data will be kept in a locked, safe place for analysis and only the research team will have access to the recordings. This will be retained for five years in an encrypted form on a secure repository maintained by the Synthetic Reality Laboratory. The collected data will be destroyed within six months after the retaining period (5 years) is over.

Risks: There are no reasonably foreseeable risks or discomforts involved in taking part in this study, other than those normally assumed as part of general teaching responsibilities. This study is voluntary, and at any time you may opt to discontinue participation in this study.

Benefits: We cannot promise any benefits to you or others from taking part in this research. You may benefit from your communication with the virtual classroom to hone your communication and classroom management strategies. You might also benefit by learning more about how research is conducted and acquiring some knowledge about innovative models of professional development in teaching.

Confidentiality: We will anonymize your personal data collected in this study. Efforts will be made to limit your personal information to people who have a need to review this information. We cannot promise complete secrecy. Data will be coded with a personal identification number to keep names confidential.

Study contact for questions about the study or to report a problem: If you have questions, concerns, or complaints, please contact Roghayeh Barmaki, Doctoral Student, CS Division, barmaki@cs.ucf.edu or Dr. Charles E. Hughes, Professor, CS Division, ceh@cs.ucf.edu

IRB contact about your rights in the study or to report a complaint: Research at the University of Central Florida involving human participants is carried out under the oversight of the Institutional Review Board (IRB). This research has been reviewed and approved by the IRB. For information about the rights of people who take part in research, please contact: Institutional Review Board, University of Central Florida, Office of Research & Commercialization, 12201 Research Parkway, Suite 501, Orlando, FL 32826-3246 or by telephone at (407) 823-2901.

Withdrawing from the study: You may decide not to continue the research study at any time. If you choose to withdraw, the research team will destroy the data associated with your participation after initial review. Your feedback, however, may contribute to changes in the interaction paradigm associated with the study.

By participating in the research study, you agree to the following:

□ I have read the procedure described above

□ I voluntarily agree to take part in the research

□ I am at least 18 years of age

□ I agree to be audiotaped and videotaped

□ I agree to have body movements and poses recorded during the study

APPENDIX C: TEACHING PLANS

What is Technology?

First Observation

At a Glance

Overview

This lesson involves classroom discussion and a short scenario to allow students to develop a sense of what technology is, dispel the notion that technology relates mostly to computers, and examine the impact of technology. The lesson is designed to enhance science literacy and is aligned with *Disciplinary Core Ideas* and *Cross-cutting Concepts* from the *Next Generation Science Standards*, as well as the *Common Core Standards for Literacy in Science*. The lesson is based on the *5E Instructional Model* and has been validated and field-tested in high school Biology classrooms as part of a larger module from the NIH Curriculum Supplement Series "Using Technology to Study Cellular and Molecular Biology" which can be found in its entirety online at http://science.education.nih.gov/supplements/nih4/technology/guide/nih_technology_curr-supp.pdf

Major Concepts

Technology is a body of knowledge used to create tools, develop skills, and extract or collect materials. It is also the application of science (the combination of the scientific method and material) to meet an objective or solve a problem.

Standards-based Objectives

- be able to explain what technology is

Part 1: 2 minutes

Begin your session with a brief introduction about yourself.

Ask the students to introduce themselves. Ask students about their most favourite topic in the school/or in general. Try to communicate with all the students in the room.

www.manaraa.com

Part 2: 5 minutes   1.   Begin by asking the class, "How do you define technology?"

Accept all answers and review student responses. Do not attempt to have students refine their definitions of technology at this point. They will revisit their definitions and refine them later.

Students, like older individuals, may harbor the preconception that technology relates mostly to computers. Through advertisements and media articles, they are familiar with the terms *information technology* and *computer technology.*

Teacher note: Asking this question requires students to call on their prior knowledge, and it engages their thinking. As this point, do not critique student responses. Appropriate teacher comments are short and positive, such as "good" and "what else?" Other appropriate teacher responses include, "Why do you believe that?" or "How do you know that?" Questions such as these allow the teacher to assess students current knowledge about the subject and to adjust lessons accordingly. They also provide a springboard to "Let's find out" or "Let's investigate." In general, it is time to move forward when you see that thinking has been engaged.

2.   Ask students, "In general, what does technology do for us?"

This question may help students understand that technology helps us solve problems, makes our lives easier, and extends our abilities to do things. Technology is used to develop skills or tools, both in our daily lives and in our occupations.

Enrichment: If students bring up the term ecosystem, as it may pertain to past biology concepts, it is appropriate to discuss ecosystems.

3.   Focus discussion on technologies that are relevant to each student's life. Ask students to look around the room. What technologies do they see? How do these technologies solve problems and make their lives easier in society, culture, and the environment?

Accept all responses and write them on the board. Students may mention any number of items. Some may be school-related, such as binders, backpacks, pens, pencils, paper, and paper clips. Other items may be more personal, such as water bottles, personal stereos, and hair clips. Students may neglect items such as shoelaces, zippers, buttons, fabric, eyeglasses or contact lenses, makeup, and bandages. Discussion should reinforce the notion that humans develop technology with a specific objective in mind. A related concept is that a give task requires the right tool or tools.

4.   Turn the discussion to how technology has impacted major world problems such as disease. Ask "What diseases have been impacted by technology?" and have students talk in small groups, before discussing as a whole class.
After a brief discussion as a class, tell students they will now use mathematics to support explanations using data from the Bill & Melinda Gates Foundation about technology and the number of cases of polio worldwide.

First Observation

LESSON OUTLINE: Solving Linear Equations in One Variable

Whole-class introduction (9 minutes)

Give each student a mini-whiteboard, a pen, and an eraser. Maximize participation in the discussion by asking all students to show you their solutions on their mini-whiteboards.

This introduction will provide students with a model of how they should justify their solutions in the collaborative activity.

Display Slide P-1 of the projector resource:

True or False?

$$4x + 1 = 3$$

Can you give me a value for x that makes this equation **false**?

Show the calculations that explain your answer.

Students should not have any problems with finding a suitable value for x, but may not be too adventurous in their choices. Spend some time discussing the values given and the reasons for each choice, identifying any common choices, as well as any calculation errors.

Display Slide P-2 of the projector resource:

True or False?

$$4x + 1 = 3$$

Can you give me a value for x that makes this equation **true**?

Show the calculations that explain your answer.

Students may struggle with this at first, especially if their chosen method is substituting values for x. Encourage students to explore fractions, decimals and negative numbers as well as positive whole numbers.

If students find a value for x, challenge them to consider if there are any other values of x. They should be encouraged to justify why this is the only value for x that makes the equation true and how they can be sure of this.

Would we describe this equation as always true, never true or sometimes true? [Sometimes true.]

When is it true? [When $x = \frac{1}{2}$.]

Are there any other values for x that make the equation true? How do you know?

APPENDIX D: PRE QUESTIONNAIRE

Pre-Questionnaire

Dear participant, thanks for your collaboration. Please take a moment to fill out this form. For the following items, please select the one response that is most descriptive of you, or fill in the blank as appropriate. You may also prefer not to answer some of those questions. If so, that will not exclude you from the study.

What is your age? ————— years old

What is your gender? ◯ Female ◯ Male

Please specify your ethnicity.

◯Hispanic or Latino ◯White ◯Black or African American ◯Native Hawaiian or Other Pacific Islander ◯Asian ◯American Indian or Alaska Native ◯Two or More Races (Please specify): —————————————

What is your major? ———————————————————

What is the highest degree you have completed? If currently enrolled, highest degree received.

◯ Less than 4 years of college ◯ Completed 4 years of college ◯ Master's ◯Doctorate

Do you have any teaching experience? ◯Yes ◯No

If **yes**,

a) How long have you been teaching?

◯ Less than a year ◯ 1–3 years ◯ 4–7 years ◯ 8+ years

b) What topics you have been teaching?

◯ Science ◯ Math ◯ English as a Second Language ◯ Social Studies

◯Exceptional Education ◯ Other(Please specify) : —————————————

Have you ever seen TeachLivE™environment? ◯Yes ◯No

Have you ever taught in the TeachLivE virtual classroom? ◯Yes ◯No

If **yes**,

a) Which avatars, and for how long did you teach (Please select all that apply)?
◯ High school avatars for approximately —— minutes
◯ Middle school avatars for approximately —— minutes
◯ Adult avatar for approximately —— minutes
◯ English as a second-language learner avatars for approximately —— minutes

b) Have you ever used a lesson–plan for your TeachLivE interaction? ◯Yes ◯No

Now, please take a moment to review the lesson–plan that is given to you.
Thanks for your participation, enjoy your experience!

APPENDIX E: POST QUESTIONNAIRE

Post-Questionnaire

Instructions: Please answer the following questions as completely as possible. Mark the circle that best represents the rating of your experience related to the TeachLivE™environment

1. How would you rate your experience with TeachLivE environment?

very awkward   awkward   neutral   natural   very natural

2. How motivated would you be to use TeachLivE again?

very bored   bored   neutral   motivated   very motivated

3. How interfering or intrusive was the Kinect sensor being used to collect data from your experience?

very low   low   neutral   high   very high

4. How likely would you be to interact with this application next time?

very likely   likely   neutral   unlikely   very unlikely

5. Do you feel like you learned anything while interacting with the application?

not at all   slightly   somewhat   moderately unlikely   completely

6. How does using this application compare to how you would normally learn the same content (for teaching) in a traditional manner?

much worse   somewhat worse   about the same   somewhat better   much better

100

7. Have you ever used an application that was similar to the one that you just tested (excluding your previous TeachLivE experiences)?

○──────○──────○
Yes   No   Maybe

8. What, if anything, do you feel like you learned from using the application?

9. What improvements would you make to this application to make it better for you and your friends?

10. If you have any additional comments or feedback, please share them here.

APPENDIX F: DEBRIFIENG FORM

Debriefing Statement

For the study entitled:

"Study of the Correlation of Body Gestures to Presence and Learning in a Virtual Learning Environment"

Dear Participant;

During this study, you were asked to teach an introductory biology topic. You were told that the purpose of the study was to improve teaching effectiveness in classrooms by evaluating communication skills and related teaching skills of participants in a virtual classroom setting. While this is true, we did not point out that your body movements and poses are an explicit part of the communication skills we are studying.

We did not tell you everything about the purpose of the study because we were concerned that this knowledge may affect your body movements.

You are reminded that your original consent document included the following information:
"You may decide not to continue the research study at any time. If you choose to withdraw, the research team will destroy the data associated with your participation after initial review. Your feedback, however, may contribute to changes in the interaction paradigm associated with the study."
If you have any concerns about your participation or the data you provided in light of this disclosure, please discuss this with us. We will be happy to provide any information we can to help answer questions you have about this study.

Now that you know the true nature of the study, you have the option of having your data removed from the study. Please contact the PI if you do not want your data to be used in this research and it will be withdrawn.

Study contact for questions about the study or to report a problem: If you have questions, concerns, or complaints or think the research has hurt you, please contact Roghayeh Barmaki, Doctoral Student, Computer Science Program (321) 800-8383 or by email at barmaki@cs.ucf.edu or Dr. Charles E. Hughes, Faculty Supervisor, CS Division Professor at ceh@cs.ucf.edu.

IRB contact about your rights in the study or to report a complaint: Research at the University of Central Florida involving human participants is carried out under the oversight of the Institutional Review Board (UCF IRB). This research has been reviewed and approved by the IRB. For information about the rights of people who take part in research, please contact: Institutional Review Board, University of Central Florida, Office of Research & Commercialization, 12201 Research Parkway, Suite 501, Orlando, FL 32826-3246 or by telephone at (407) 823-2901.

Please again accept our appreciation for your participation in this study.

LIST OF REFERENCES

[1] Instructional evaluation manual and protocols. `https://www.ocps.net/cs/pds/assessment/Pages/default.aspx`. Retrieved April 2015.

[2] Interstate new teacher assessment and support consortium. `http://www.bsu.edu/cte/rpm/examples/julie/intasc/`. Retrieved: April 2015.

[3] Using technology to study cellular and molecular biology. `http://science.education.nih.gov/supplements/nih4/technology/default.html`, 2005. Retrieved: August 2015.

[4] Visual gesture builder: A data-driven solution to gesture detection. http://aka.ms/k4wv2vgb, July 2014. Retrieved March 2016.

[5] D. Abercrombie. Paralanguage. *International Journal of Language & Communication Disorders*, 3(1):55–59, 1968.

[6] 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.

[7] 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.

[8] M. W. Alibali, A. G. Young, N. M. Crooks, A. Yeo, M. S. Wolfgram, I. M. Ledesma, M. J. Nathan, R. Breckinridge Church, and E. J. Knuth. Students learn more when their teacher has learned to gesture effectively. *Gesture*, 13(2):210–233, 2013.

[9] R. Barmaki. Nonverbal communication and teaching performance. In *Proceedings of the 8th International Conference on Educational Data Mining*, EDM '14, pages 441–443, 2014.

[10] R. Barmaki. Multimodal assessment of teaching behavior in immersive rehearsal environment-teachlive. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ICMI '15, pages 651–655, New York, NY, USA, 2015. ACM.

[11] R. Barmaki. Improving social communication skills using kinesics feedback. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 86–91, New York, NY, USA, 2016. ACM.

[12] 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.

[13] R. Barmaki and C. E. Hughes. Providing real-time feedback for student teachers in a virtual rehearsal environment. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ICMI '15, pages 531–537, New York, NY, USA, 2015. ACM.

[14] R. Barmaki and C. E. Hughes. Towards the understanding of gestures and vocalization coordination in teaching context. In *Proceedings of the 9th International Conference on Educational Data Mining*, EDM '16, pages 663–665, 2016.

[15] B. Barry, J. Bodenhamer, and J. J. O'Brien Jr. Student nonverbal communication in the classroom. In *American Society for Engineering Education*. American Society for Engineering Education, 2011.

[16] R. Beheshti, A. M. Ali, and G. Sukthankar. Cognitive social learners: an architecture for modeling normative behavior. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, pages 2017–2023. AAAI Press, 2015.

[17] R. Beheshti and G. Sukthankar. A normative agent-based model for predicting smoking cessation trends. In *Proceedings of the International Conference on Autonomous Agents and Multi-agent Systems*, pages 557–564, 2014.

[18] R. Beheshti and G. Sukthankar. A hybrid modeling approach for parking and traffic prediction in urban simulations. *AI & SOCIETY*, 30(3):333–344, 2015.

[19] R. Beheshti and G. Sukthankar. Modeling tipping point theory using normative multi-agent systems. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pages 1731–1732. International Foundation for Autonomous Agents and Multiagent Systems, 2015.

[20] F. Biocca and C. Harms. Defining and measuring social presence: Contribution to the networked minds theory and measure. *Proceedings of PRESENCE*, 2002:1–36, 2002.

[21] P. Blikstein. Unraveling students' interaction around a tangible interface using gesture recognition. In *Educational Data Mining 2014*, 2014.

[22] T. Blum, V. Kleeberger, C. Bichlmeier, and N. Navab. mirracle: An augmented reality magic mirror system for anatomy education. In *2012 IEEE Virtual Reality Workshops (VRW)*, pages 115–116. IEEE, 2012.

[23] P. Boersma and D. Weenink. Praat: doing phonetics by computer [computer program] version. 6.0.17, 2016. Retrieved March 2016 from http://www.praat.org/.

[24] B. B. Brown. Delphi process: A methodology used for the elicitation of opinions of experts. Technical report, DTIC Document, 1968.

[25] V. Castaneda and N. Navab. Time-of-flight and kinect imaging. *Kinect Programming for Computer Vision*, 2011.

[26] C. Caswell and S. Neill. *Body language for competent teachers*. Routledge, 2003.

[27] M. Chollet, G. Sratou, A. Shapiro, L.-P. Morency, and S. Scherer. An interactive virtual audience platform for public speaking training. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*, AAMAS '14, pages 1657–

1658, Richland, SC, 2014. International Foundation for Autonomous Agents and Multiagent Systems.

[28] M. Chu and S. Kita. The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, 140(1):102, 2011.

[29] 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.

[30] S. D. Craig, J. Twyford, N. Irigoyen, and S. A. Zipp. A test of spatial contiguity for virtual humans gestures in multimedia learning environments. *Journal of Educational Computing Research*, page 0735633115585927, 2015.

[31] C. Danielson. *The framework for teaching evaluation instrument*. Danielson Group, 2011.

[32] J. Davenport, A. Rafferty, M. Timms, D. Yaron, and M. Karabinos. Chemvlab+: evaluating a virtual lab tutor for high school chemistry. In *Inter. Conf. of the Learning Sciences (ICLS)*, 2012.

[33] F. Dermody and A. Sutherland. A multimodal system for public speaking with real time feedback. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ICMI '15, pages 369–370, New York, NY, USA, 2015. ACM.

[34] L. Dieker, M. Hynes, C. Hughes, and E. Smith. Implications of mixed reality and simulation technologies on special education and teacher preparation. *Focus on Exceptional Children*, 40(6):1–20, 2008.

[35] 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.

[36] M. R. DiMatteo, R. D. Hays, and L. M. Prince. Relationship of physicians' nonverbal communication skill to patient satisfaction, appointment noncompliance, and physician workload. *Health Psychology*, 5(6):581, 1986.

[37] A. T. Dittmann and L. G. Llewellyn. Relationship between vocalizations and head nods as listener responses. *Journal of personality and social psychology*, 9(1):79, 1968.

[38] W. Doyle. The uses of nonverbal behaviors: Toward an ecological model of classrooms. *Merrill-Palmer Quarterly of Behavior and Development*, pages 179–192, 1977.

[39] P. Ekman and W. V. Friesen. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1(1):49–98, 1969.

[40] P. Ekman and W. V. Friesen. Unmasking the face: A guide to recognizing emotions from facial cues, 1975.

[41] P. Ekman and W. V. Friesen. Facial action coding system. 1977.

[42] P. Ekman and E. L. Rosenberg. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, 1997.

[43] P. C. Ellsworth and J. M. Carlsmith. Effects of eye contact and verbal content on affective response to a dyadic interaction. *Journal of Personality and Social Psychology*, 10(1):15, 1968.

[44] A. Ezen-Can, J. F. Grafsgaard, J. C. Lester, and K. E. Boyer. Classifying student dialogue acts with multimodal learning analytics. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, pages 280–289. ACM, 2015.

[45] A. J. Flint, S. E. Black, I. Campbell-Taylor, G. F. Gailey, and C. Levinton. Abnormal speech articulation, psychomotor retardation, and subcortical dysfunction in major depression. *Journal of psychiatric research*, 27(3):309–319, 1993.

[46] C. Fröhlich, P. Biermann, M. E. Latoschik, and I. Wachsmuth. Processing iconic gestures in a multimodal virtual construction environment. In *Gesture-Based Human-Computer Interaction and Simulation*, pages 187–192. Springer, 2009.

[47] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 529–536. ACM, 2003.

[48] D. Gentner and S. Goldin-Meadow. Language in mind, 2003.

[49] J. Gratch, G. M. Lucas, A. A. King, and L.-P. Morency. It's only a computer: the impact of human-agent interaction in clinical interviews. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 85–92. International Foundation for Autonomous Agents and Multiagent Systems, 2014.

[50] S. Gregory and Y. Masters. Real thinking with virtual hats: A role-playing activity for pre-service teachers in second life. *Australasian Journal of Educational Technology*, 28(3):420–440, 2012.

[51] L. K. Guerrero and K. Floyd. *Nonverbal communication in close relationships*. Routledge, 2006.

[52] P. L. Gunter et al. On the move: Using teacher/student proximity to improve students' behavior. *Teaching Exceptional Children*, 28(1):12–14, 1995.

[53] J. Hattie and H. Timperley. The power of feedback. *Review of educational research*, 77(1):81–112, 2007.

[54] A. T. Hayes. *The experience of physical and social presence in a virtual learning environment as impacted by the affordance of movement enabled by motion tracking*. PhD thesis, 2015.

[55] A. T. Hayes, C. L. Straub, L. A. Dieker, C. E. Hughes, and M. C. Hynes. Ludic learning: Exploration of tle teachlive and effective teacher training. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, 5(2):20–33, 2013.

[56] 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.

[57] H.-m. J. Hsu. The potential of kinect in education. *International Journal of Information and Education Technology*, 1(5):365–370, 2011.

[58] A. S. Imada and M. D. Hakel. Influence of nonverbal communication and rater proximity on impressions and decisions in simulated employment interviews. *Journal of Applied Psychology*, 62(3):295, 1977.

[59] C. C. S. S. Initiative et al. *Common core state standards for English language arts & literacy in history/social studies, science, and technical subjects*. Common Core Standards Initiative, 2012.

[60] U. Kale. Levels of interaction and proximity: Content analysis of video-based classroom cases. *The Internet and Higher Education*, 11(2):119–128, 2008.

[61] L. T. Keith, L. G. Tornatzky, and L. E. Pettigrew. An analysis of verbal and nonverbal classroom teaching behaviors. *The Journal of Experimental Education*, 42(4):30–38, 1974.

[62] 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.

[63] S. D. Kelly, T. McDevitt, and M. Esch. Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, 24(2):313–334, 2009.

[64] A. Kendon. *Gesture: Visible action as utterance*. Cambridge University Press, 2004.

[65] M. Kipp. Anvil: The video annotation research tool. In J. Durand, U. Gut, and G. Kristof-fersen, editors, *The Oxford Handbook of Corpus Phonology*. Oxford University Press, 2014.

[66] H. G. Klinzing and B. Gerada-Aloisio. Intensity, variety, and accuracy in nonverbal cues and de-/encoding: Two experimental investigations. *Online Submission*, 2004.

[67] M. Knapp, J. Hall, and T. Horgan. *Nonverbal communication in human interaction*. Cengage Learning, 2013.

[68] T. Koumoutsakis, R. B. Church, M. W. Alibali, M. Singer, and S. Ayman-Nolley. Gesture in instruction: Evidence from live and video lessons. *Journal of Nonverbal Behavior*, pages 1–15, 2016.

[69] G. Lakoff and R. E. Núñez. *Where mathematics comes from: How the embodied mind brings mathematics into being*. Basic books, 2000.

[70] B. Lange, C.-Y. Chang, E. Suma, B. Newman, A. S. Rizzo, and M. Bolas. Development and evaluation of low cost game-based balance rehabilitation tool using the microsoft kinect sensor. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 1831–1834. IEEE, 2011.

[71] R. Laslett and C. Smith. *Effective classroom management: a teacher's guide*. Routledge, 2002.

[72] D. J. Leach and H. Conto. The additional effects of process and outcome feedback following brief inservice teacher training. *Educational Psychology*, 19(4):441–462, 1999.

[73] D. Lefloch, R. Nair, F. Lenzen, H. Schäfer, L. Streeter, M. J. Cree, R. Koch, and A. Kolb. Technical foundation and calibration methods for time-of-flight cameras. In *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*, pages 3–24. Springer, 2013.

111

[74] C. W. Leong, L. Chen, G. Feng, C. M. Lee, and M. Mulholland. Utilizing depth sensors for analyzing multimodal presentations: Hardware, software and toolkits. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ICMI '15, pages 547–556, New York, NY, USA, 2015. ACM.

[75] W. A. Long Jr. Personality and learning. 1988 john wilson memorial address. *Focus on Learning Problems in Mathematics*, 11(4):1–16, 1989.

[76] D. Luciew, J. Mulkern, and R. Punako. Finding the truth: Interview and interrogation training simulations. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*, volume 2011, 2011.

[77] M. Macedonia, K. Müller, and A. D. Friederici. The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*, 32(6):982–998, 2011.

[78] P. Maragos, A. Potamianos, and P. Gros. Multimodal processing and interaction: Audio, video. *Text. Springer Science Media, Heidelberg*, 2008.

[79] R. J. Marzano and M. D. Toth. *Teacher evaluation that makes a difference: A new model for teacher growth and student achievement*. ASCD, 2013.

[80] E. Z. McClave. Linguistic functions of head movements in the context of speech. *Journal of pragmatics*, 32(7):855–878, 2000.

[81] H. McGinley, R. LeFevre, and P. McGinley. The influence of a communicator's body position on opinion change in others. *Journal of Personality and Social Psychology*, 31(4):686, 1975.

[82] D. McNeill. *Hand and mind: What gestures reveal about thought*. University of Chicago Press, 1992.

[83] A. Mehrabian. Communication without words. *Psychological today*, 2:53–55, 1968.

[84] A. Mehrabian. *Nonverbal communication*. Transaction Publishers, 1977.

[85] A. Mehrabian. Silent messages - a wealth of information about nonverbal communication (body language). *Personality & Emotion Tests & Software: Psychological Books & Articles of Popular Interest. Los Angeles, CA*, 7(31):2011, 2009.

[86] P. W. Miller. Body language in the classroom. *Techniques: Connecting Education and Careers*, 8:28–30, 2005.

[87] B. Minaei-Bidgoli, R. Barmaki, and M. Nasiri. Mining numerical association rules via multi-objective genetic algorithms. *Information Sciences*, 233:15–24, 2013.

[88] M. Minsky. Telepresence. 1980.

[89] J. Morkes, H. K. Kernal, and C. Nass. Effects of humor in task-oriented human-computer interaction and computer-mediated communication: A direct test of srct theory. *Human-Computer Interaction*, 14(4):395–435, 1999.

[90] E. H. Mory. Feedback research revisited. *Handbook of research on educational communications and technology*, 2:745–783, 2004.

[91] T. P. Mottet, V. P. Richmond, and J. C. McCroskey. *Handbook of instructional communication: Rhetorical and relational perspectives*. Allyn & Bacon, 2006.

[92] 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.

[93] R. Niewiadomski, M. Mancini, and S. Piana. Human and virtual agent expressive gesture quality analysis and synthesis. *Coverbal Synchrony in Human-Machine Interaction*, pages 269–292, 2013.

[94] K. L. Nowak and F. Biocca. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence*, 12(5):481–494, 2003.

[95] X. Ochoa, K. Chiluiza, G. Méndez, G. Luzardo, B. Guamán, and J. Castells. Expertise estimation based on simple multimodal features. In *Proceedings of the 15th ACM on International conference on multimodal interaction*, pages 583–590. ACM, 2013.

[96] X. Ochoa, M. Worsley, N. Weibel, and S. Oviatt. Multimodal learning analytics data challenges. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, LAK '16, pages 498–499, New York, NY, USA, 2016. ACM.

[97] K. L. O'Halloran. Multimodal analysis and digital technology. In *Interdisciplinary Perspectives on Multimodality: Theory and Practice, Proceedings of the Third International Conference on Multimodality, Palladino, Campobasso*, 2009.

[98] K. L. OHalloran. Multimodal discourse analysis. *Companion to discourse*, pages 120–137, 2011.

[99] S. Y. Okita, J. Bailenson, and D. L. Schwartz. Mere belief in social action improves complex learning. In *Proceedings of the 8th International Conference on International Conference for the Learning Sciences*, ICLS'08, pages 132–139. International Society of the Learning Sciences, 2008.

[100] K. Ord. Outliers in statistical data. *International Journal of Forecasting*, 1(12):175–176, 1996.

[101] R. D. Pea. Video-as-data and digital video manipulation techniques for transforming learning sciences research, education, and other cultural practices. In *The international handbook of virtual learning environments*, pages 1321–1393. Springer, 2006.

[102] J. E. Perez and R. E. Riggio. Nonverbal social skills and psychopathology. *Nonverbal behavior in clinical settings*, pages 17–44, 2003.

114

[103] D.-P. Pertaub, M. Slater, and C. Barker. An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators and virtual environments*, 11(1):68–78, 2002.

[104] 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.

[105] A. Popescu-Belis. Managing multimodal data, metadata and annotations: Challenges and solutions. *Multimodal Signal Processing: Theory and applications for human-computer interaction*, page 207, 2009.

[106] L. Pozzer-Ardenghi and W.-M. Roth. On performing concepts during science lectures. *Science Education*, 91(1):96–114, 2007.

[107] V. Ramanarayanan, C. W. Leong, L. Chen, G. Feng, and D. Suendermann-Oeft. Evaluating speech, face, emotion and body movement time-series features for automated multimodal presentation scoring. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 23–30. ACM, 2015.

[108] S. Renals. *Multimodal signal processing : human interactions in meetings.* Cambridge ; New York : Cambridge University Press, 2012., 2012.

[109] P. J. Rich and M. Hannafin. Video annotation tools technologies to scaffold, structure, and transform teacher reflection. *Journal of Teacher Education*, 60(1):52–67, 2009.

[110] V. P. Richmond, J. C. McCroskey, and M. L. Hickson III. Nonverbal behavior in interpersonal relations. 2014.

[111] A. A. Ross, K. Nandakumar, and A. K. Jain. *Handbook of multibiometrics*, volume 6. Springer Science & Business Media, 2006.

[112] W.-M. Roth. Gestures: Their role in teaching and learning. *Review of Educational Research*, 71(3):365–392, 2001.

[113] S. Saleh, M. Sahu, Z. Zafar, K. Berns, R. Rama, S. Skatulla, C. Sansour, W. Auccahuasi, J. B. Marcatoma, S. Takayama, et al. A multimodal nonverbal human-robot communication system. *Proceedings of Sixth International Conference on Computational Bioengineering, ICCB*, 2015.

[114] T. L. Sawyer and S. Deering. Adaptation of the us armys after-action review for simulation debriefing in healthcare. *Simulation in Healthcare*, 8(6):388–397, 2013.

[115] J. T. M. Schelde. Major depression: Behavioral markers of depression and recovery. *The Journal of nervous and mental disease*, 186(3):133–140, 1998.

[116] S. Scherer, G. Stratou, J. Gratch, and L.-P. Morency. Investigating voice quality as a speaker-independent indicator of depression and ptsd. In *Interspeech*, pages 847–851, 2013.

[117] J. Schneider, D. Börner, P. van Rosmalen, and M. Specht. Presentation trainer, your public speaking multimodal coach. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ICMI '15, pages 539–546, New York, NY, USA, 2015. ACM.

[118] L. A. Schwarz, A. Mkhitaryan, D. Mateus, and N. Navab. Estimating human 3d pose from time-of-flight images based on geodesic distances and optical flow. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 700–706. IEEE, 2011.

[119] J. Scoresby and B. E. Shelton. Reflective redo from the point of error implications for after action review. *Simulation & Gaming*, page 1046878114549426, 2014.

[120] S. Seo. *A review and comparison of methods for detecting outliers in univariate data sets*. PhD thesis, University of Pittsburgh, 2006.

[121] L. Shapiro. *The Routledge handbook of embodied cognition*. Routledge, 2014.

[122] M. Slater, J. Howell, A. Steed, D.-P. Pertaub, and M. Garau. Acting in virtual reality. In *Proceedings of the third international conference on Collaborative virtual environments*, pages 103–110. ACM, 2000.

[123] H. A. Smith. Nonverbal communication in teaching. *Review of Educational Research*, 49(4):631–672, 1979.

[124] Y. Song and R. Davis. Continuous body and hand gesture recognition for natural human-computer interaction. In *Proceedings of the 24th International Conference on Artificial Intelligence*, pages 4212–4216. AAAI Press, 2015.

[125] N. G. S. S. N. L. States. Next generation science standards: For states, by states, 2013.

[126] G. Stratou, S. Scherer, J. Gratch, and L.-P. Morency. Automatic nonverbal behavior indicators of depression and ptsd: Exploring gender differences. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 147–152. IEEE, 2013.

[127] M. Tacchetti. User guide for elan linguistic annotator version 4.1.0. *Nijmegen: The Language Archive*, 2013.

[128] R. F. Tate. Correlation between a discrete and a continuous variable. point-biserial correlation. *The Annals of mathematical statistics*, pages 603–607, 1954.

[129] F. Vanni, C. Conversano, A. Del Debbio, P. Landi, M. Carlini, C. Fanciullacci, and L. Dell'Osso. A survey on virtual environment applications to fear of public speaking. *Eur Rev Med Pharmacol Sci*, 17(12):1561–1568, 2013.

[130] W. Wang and S. Loewen. Nonverbal behavior and corrective feedback in nine esl university-level classrooms. *Language Teaching Research*, page 1362168815577239, 2015.

[131] P. Waxer. Nonverbal cues for depression. *Journal of Abnormal Psychology*, 83(3):319, 1974.

117

[132] D. Wiebusch, M. Fischbach, F. Niebling, and M. E. Latoschik. Low-cost raycast-based coordinate system registration for consumer depth cameras. In *Proceedings of the 23rd IEEE Virtual Reality (IEEE VR) conference*, 2016.

[133] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and virtual environments*, 7(3):225–240, 1998.

[134] P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes. Elan: a professional framework for multimodality research. In *Proceedings of LREC*, volume 2006, page 5th, 2006.

[135] A. S. Won, J. N. Bailenson, and J. H. Janssen. Automatic detection of nonverbal behavior predicts learning in dyadic interactions. *Affective Computing, IEEE Transactions on*, 5(2):112–125, 2014.

[136] A. E. Woolfolk and D. M. Brooks. The influence of teachers' nonverbal behaviors on students' perceptions and performance. *The Elementary School Journal*, pages 513–528, 1985.

[137] M. Worsley and P. Blikstein. Towards the development of multimodal action based assessment. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, LAK '13, pages 94–101, New York, NY, USA, 2013. ACM.

[138] M. Worsley, K. Chiluiza, J. F. Grafsgaard, and X. Ochoa. 2015 multimodal learning and analytics grand challenge. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ICMI '15, pages 525–529, New York, NY, USA, 2015. ACM.