IMPROVING MOBILE MOOC LEARNING VIA IMPLICIT PHYSIOLOGICAL SIGNAL SENSING

by

Xiang Xiao

Bachelor of Engineering, Wuhan University, 2010

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of Doctor of Philosophy

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES,
DEPARTMENT OF COMPUTER SCIENCE

This dissertation was presented

by

Xiang Xiao

It was defended on

September 26, 2016

and approved by

Dr. Jingtao Wang, Department of Computer Science, University of Pittsburgh

Dr. Milos Hauskrecht, Department of Computer Science, University of Pittsburgh

Dr. Adam J. Lee, Department of Computer Science, University of Pittsburgh

Dr. Peter Brusilovsky, School of Information Science, University of Pittsburgh

Dissertation Advisor: Dr. Jingtao Wang, Department of Computer Science, University of Pittsburgh

IMPROVING MOBILE MOOC LEARNING VIA IMPLICIT PHYSIOLOGICAL SIGNAL SENSING

Xiang Xiao, Ph.D.

University of Pittsburgh, 2017

Massive Open Online Courses (MOOCs) are becoming a promising solution for delivering high-quality education on a large scale at low cost in recent years. Despite the great potential, today's MOOCs also suffer from challenges such as low student engagement, lack of personalization, and most importantly, lack of direct, immediate feedback channels from students to instructors. This dissertation explores the use of physiological signals implicitly collected via a "sensorless" approach as a rich feedback channel to understand, model, and improve learning in mobile MOOC contexts.

I first demonstrate AttentiveLearner, a mobile MOOC system which captures learners' physiological signals implicitly during learning on unmodified mobile phones. AttentiveLearner uses on-lens finger gestures for video control and monitors learners' photoplethysmography (PPG) signals based on the fingertip transparency change captured by the back camera. Through series of usability studies and follow-up analyses, I show that the tangible video control interface of AttentiveLearner is intuitive to use and easy to operate, and the PPG signals implicitly captured by AttentiveLearner can be used to infer both learners' cognitive states (boredom and confusion levels) and divided attention (multitasking and external auditory distractions).

Building on top of AttentiveLearner, I design, implement, and evaluate a novel intervention technology, Context and Cognitive State triggered Feed-Forward (C2F2), which infers and responds to learners' boredom and disengagement events in real time via a combination of PPG-based cognitive state inference and learning topic importance monitoring. C2F2 proactively reminds a student of important upcoming content (feed-forward interventions) when disengagement is detected. A 48-participant user study shows that C2F2 on average improves learning gains by 20.2% compared with a non-interactive baseline system and is especially effective for bottom performers (improving their learning gains by 41.6%).

Finally, to gain a holistic understanding of the dynamics of MOOC learning, I investigate the temporal dynamics of affective states of MOOC learners in a 22-participant study. Through both a quantitative analysis of the temporal transitions of affective states and a qualitative analysis of subjective feedback, I investigate differences between mobile MOOC learning and complex learning activities in terms of affect dynamics, and discuss pedagogical implications in detail.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to express my sincere thanks to my advisor, Dr. Jingtao Wang, for the support, guidance, and encouragement throughout these years. For me, he not only taught me research methods and skills, but also taught me the correct attitude I should have towards research, work, and life. I remembered when I first started as a PhD student, I made many "rookie" mistakes. Jingtao remained very patient with me, spent hours talking with me about his experience, and offered a lot of valuable guidance. His mentoring is greatly valued and I could not have accomplished this dissertation without his dedication.

I would like to thank my PhD committee: Dr. Peter Brusilovsky, Dr. Milos Hauskrecht, and Dr. Adam J. Lee, for their time spent on reviewing thesis draft and attending my meetings. Their advices helped me improve my research and present it the best way I can.

I would like to thank my labmates and collaborators in the MIPS research group at Pitt: Xiangmin Fan, Phuong Pham, Lanfei Shi, Teng Han, Yuxin Liu, Wei Guo, Carrie Demmans Epp, Andrew Head, and others. They have provided me valuable feedbacks and constructive suggestions regarding my projects, papers, and presentations. I also want to thank them for their company throughout these years. They made this experience much more fun and enjoyable.

I would like to thank all my friends in the USA and China, for their company and help. They gave me a lot of support and encouragement when I had doubt about myself. I also want to thank the people who actively participated in my user studies and offered me valuable feedbacks. Without their participation, none of the studies in this dissertation would be possible.

I am deeply grateful to Xiaoyu Xiao, a special “mentor” of mine, who gave me many valuable suggestions of how to deal with life’s challenges and remain calm and patient.

I would like to express my utmost thanks to my parents. They let me follow my desire to come to the US and chase my dream. I could never have done it without them. Their belief in me is my biggest motivation to succeed.

1.0 INTRODUCTION

With the increasing ubiquity of Internet access, Massive Open Online Courses (MOOCs) are experiencing rapid growth. MOOCs are becoming a common form of online course delivery with more than 35 million registered students by December 2015 [109]. Although polarized opinions still persist, many experts believe that MOOCs and flipped courses, when properly designed and executed, could serve as an excellent complement to traditional education by delivering high-quality education on a large scale at a lower cost [46]. MOOCs allow learners "*to control where, what, how and with whom they learn*" [68]. Researchers also found MOOCs could help students develop new skills such as autonomous learning, and improve the visibility of universities and courses [62].

In current MOOCs, the courses are mostly organized as sequences of pre-recorded lecture videos, split into 3-15 minute pieces for better engagement [51]. Such small video clips are easy to consume on mobile devices during learners' fragmented time. Indeed, major MOOC providers, such as Coursera, edX, and Udacity, have released their mobile apps to support "*learning on the go*". Compared with traditional MOOCs, mobile MOOCs provide unique advantages such as "*ubiquitous, respond to urgent learning need, and flexibility of location and time to learn*" [129].

Despite the great potential, MOOCs today still have much room for improvement. Studies on MOOCs have shown problems such as low completion rate (10% in [46], 7% in [87]), high in-session distractions [94], and lack of interactions among students and instructors [48, 116]. I summarize three essential challenges for MOOCs. First, learners are more likely to become disengaged and "*mind wander*" in MOOC contexts than in classrooms [46, 100]. This is because of external distractions in the non-classroom environment and a lack of sustained motivation when studying alone with technology. Second, there is a lack of personalization for individual learners. Given the large pool of students, it is hard for MOOC instructors to cater instructions and learning materials for individual learner's need and learning process. Third, the current design of MOOCs is primarily unidirectional, i.e., from instructors to students. There is a lack of direct, immediate feedback channels from students to instructors. Unlike in traditional classrooms, MOOC instructors no longer have access to important cues, such as facial expressions, raised hands, or oral questions to infer students' mental and learning state. Although questionnaires, post-lecture reflections [45, 48], and browser log analysis (including both activities in learning sessions [66] and follow-up discussion forums [136, 137]) can be used to infer the quality of learning, such post-hoc analysis techniques are usually coarse-grained, highly delayed, and indirect measurements of the actual learning process.

To directly measure learners' actual learning process, researchers have explored using physiological signals to infer learners' cognitive and affective states in educational systems. AutoTutor [36, 55], Wayang intelligent tutor [128], GazeTutor [42] and Artful [116], are a few notable examples. These systems collect physiological signals, such as heart rates [52, 55, 59], facial features [36, 61, 128], galvanic skin responses (GSR) [13, 17, 52, 55, 59] and Electroencephalography (EEG) [52, 59, 116], and use machine learning algorithms to predict students' affective and cognitive states (e.g., attention, mind wandering, and confusion, etc.) during learning. Although these systems have been shown to successfully detect and respond to learners' affective and cognitive states, most of them require dedicated sensors, such as cameras

[36, 128], eye trackers [42], and EEG headsets [116] for signal collection. The cost, availability, and portability of such equipment have prevented the wide adoption of such systems beyond lab settings.

1.1 ATTENTIVELEARNER

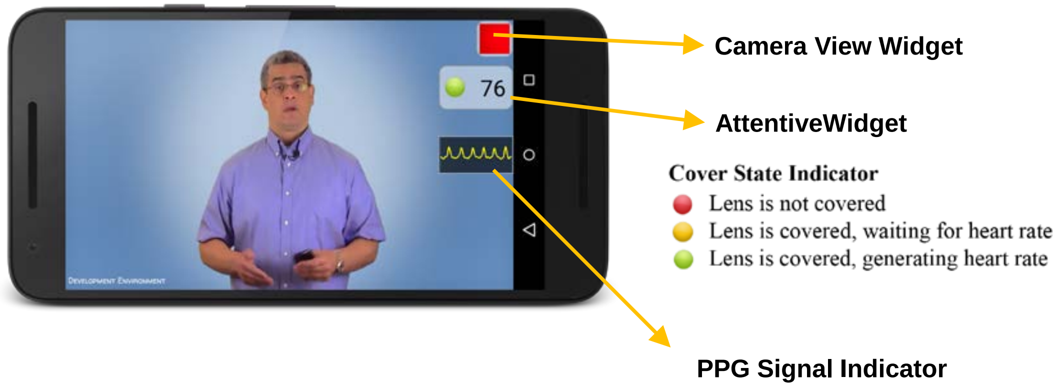


Figure 1. The video play interface of AttentiveLearner.

This video play interface is similar to that of a mobile video player, with additional widgets for visualizing the camera preview window, the attention indicator and the PPG preview window. The camera preview window shows a live video stream from the back camera. The AttentiveWidget shows the learner's instant heart rate as well as the covering state of the lens. The PPG indicator shows the real-time waveform of the learner's PPG signals.

Motivated by previous research which use physiological signals to understand and enhance learning in computer-based educational systems, I develop AttentiveLearner (Figure 1), a mobile learning system which captures and uses learners' physiological signals *implicitly* to improve mobile MOOC learning *without* leveraging any dedicated sensors. AttentiveLearner uses on-lens finger gestures to control the MOOC video playback (i.e., covering and holding the back camera lens to play a lecture video, uncovering the lens to pause the video). Moreover, it monitors learners' photoplethysmography (PPG) signals *implicitly* based on the fingertip transparency changes captured by the back camera on *unmodified* mobile phones. Compared with previous cognitive-aware educational systems which utilize dedicated devices to track physiological signals, AttentiveLearner uses the built-in camera of mobile phones to monitor physiological signals. More importantly, in addition to traditional activity logs, questionnaires, and quiz performance, AttentiveLearner provides an informative and orthogonal feedback channel for MOOC instructors.

I use the term "implicit physiological signal sensing" to describe AttentiveLearner to differentiate between the usage scenario of AttentiveLearner and traditional methods when tracking physiological signals. The "implicitness" of AttentiveLearner has two aspects: *Implicit Hardware* and *Implicit Software*. Implicit Hardware means that AttentiveLearner does not require users to purchase and mount extra sensing equipment to collect physiological signals. Implicit Software suggests that users do not need to "explicitly" launch monitoring apps and spend an uninterrupted amount of time in data collection. Instead, extraction of PPG signals has been integrated into the process of video control, thus AttentiveLearner can infer learners' physiological signals as a *side effect* while they are watching the lecture videos.

By predicting learners' cognitive and affective states (e.g., attention, "mind wandering" events, or confusion) using the PPG signals implicitly captured by the built-in camera of unmodified mobile phones, AttentiveLearner has the potential to address the three MOOC challenges mentioned earlier by: 1) adaptive interventions (e.g., integrated exercises, feed-forward reminders) when "mind wandering" events or learner disengagement are detected, thus making the learning process more attentive and enjoyable; 2) providing learners with personalized learning materials and instructional paradigms (e.g., adjusting material difficulty, switching learning tasks, etc.) based on their real-time cognitive state inference; and 3) providing instructors with fine-grained, real-time feedback on learners' actual cognitive states synchronized with the learning materials, thus facilitate bi-directional learning.

1.2 RESEARCH OVERVIEW

In this dissertation, I systematically explore this “sensorless” approach adopted by AttentiveLearner (Figure 1) which implicitly captures learners’ physiological signals and infers their cognitive states during mobile MOOC learning. Figure 2 shows the three major components of AttentiveLearner: 1) a *tangible video control channel* (Chapter 3); 2) an *implicit PPG sensing module* (Chapter 4); and 3) *cognitive state inference algorithms* (Chapter 5, 6). These three components together allow instructors to gain a deeper understanding of MOOC learners’ cognitive and affective states, making AttentiveLearner a rich, fine-grained feedback channel from learners to instructors.

Moreover, AttentiveLearner can benefit learners by providing personalized and adaptive learning interventions based on the detected cognitive states. To justify this assumption, I designed and implemented a novel intervention technology, context and cognitive state triggered adaptive feed-forward (C2F2), within AttentiveLearner. This technology uses cognitive state triggered proactive reminders (feed-forward) as an intervention to recognize and alleviate disengagement in mobile MOOC learning. C2F2 shows the feasibility and effectiveness of using PPG signals implicitly recorded by mobile cameras to improve mobile MOOC learning (Chapter 7).

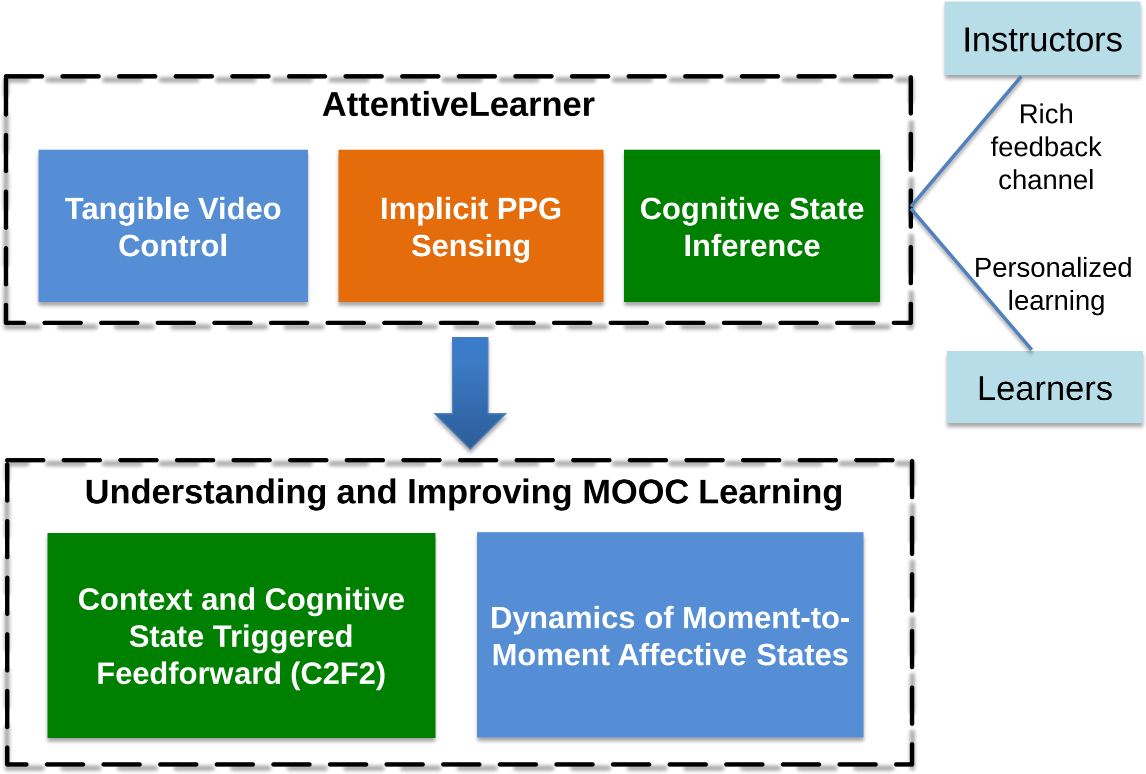

Instructors
AttentiveLearner
Tangible Video Control
Implicit PPG Sensing
Cognitive State Inference
Rich feedback channel
Personalized learning
Learners
Understanding and Improving MOOC Learning
Context and Cognitive State Triggered Feedforward (C2F2)
Dynamics of Moment-to-Moment Affective States

Figure 2. Main components of this research.

Moreover, I investigated the dynamics of learners' moment-to-moment affective state transitions during mobile MOOC learning (Chapter 8). I extended the model of affect dynamics in complex learning environments [33] to MOOC contexts. This research promotes a better understanding of the dynamic learning process and provides important pedagogical implications of how and when to regulate the learners' affective states in MOOC contexts.

1.3 STATEMENT

Given the above research overview, this dissertation has the following thesis statement:

By proposing a "sensorless" approach to collect photoplethysmography (PPG) signals implicitly from users on unmodified mobile devices, this dissertation explores novel technologies to monitor learners' cognitive and affective states, and provide cognitive state triggered adaptive interventions, which can effectively improve learning in mobile MOOC contexts.

1.4 HYPOTHESES

To support this thesis, I generate a list of hypotheses grouped by the different components presented in Figure 2:

A. The tangible video control channel:

A1. The AttentiveLearner tangible video control channel is accurate and responsive enough to support smooth user interactions.

A2. The tangible video control channel is user-friendly. This requires that it is easy to operate and comfortable to use for the learners. The battery life should also be enough to support most learners' daily MOOC learning requirement.

B. The implicit PPG sensing module:

B1. With only the built-in camera of commodity mobile phones, I can collect high quality PPG signals (comparable to signals collected by a standard pulse oximeter) from the user. With the PPG signals captured by the camera, I can accurately measure the user's heart rates.

B2. I can collect reliable, high quality PPG signals during actual mobile MOOC learning sessions with the implicit PPG sensing module enabled by the built-in camera.

C. Cognitive state inference:

C1. Using the PPG signals captured by AttentiveLearner to predict learners' cognitive states, such as boredom and confusion during learning, I can achieve comparable predication performance as existing systems [32, 55] which rely on dedicated physiological sensors to predict cognitive and affective states.

D. Context and cognitive state triggered feed-forward:

D1. Cognitive-aware interactive systems built on top of AttentiveLearner will benefit learners and improve their learning outcomes by providing in-situ adaptation and feedback.

D2. Providing adaptive cognitive state triggered proactive reminders before important topics (i.e., C2F2) during mobile MOOC learning will improve learning performance.

E. Dynamics of moment-to-moment affective states in MOOC contexts:

E1. The dynamics of moment-to-moment affective states in MOOCs is different from that in complex learning [33] due to the different characteristics of these two learning contexts.

E2. Using the PPG signals captured by AttentiveLearner to predict the fine-grained moment-to-moment affect dynamics in mobile MOOC learning, I can achieve significantly better performance than a random classifier (Kappa = 0), suggesting the feasibility of using AttentiveLearner to detect moment-to-moment affect dynamics in MOOC contexts.

1.5 CONTRIBUTIONS

To summarize, this research has three major contributions:

- It promotes a better understanding of MOOC learners by providing instructors with a direct feedback channel of learners' actual learning states, specifically their cognitive and affective states. On the contrary, current feedback mechanisms in MOOCs (e.g., questionnaires, post-lecture discussions and reflections, browser-log analysis, etc.) are mostly used to infer the quality of learning, and cannot provide direct, fine-grained information of learners' actual learning states during each learning session.

- It presents AttentiveLearner, a cognitive-aware mobile learning system, as well as C2F2, an intelligent intervention technique. Through a 48-participant user study, this dissertation shows that the cognitive-aware interactive system can effectively improve learners' performance in mobile MOOC learning.
- It provides valuable insights and pedagogical implications from the PPG signal analysis and the investigation of affect dynamics. These insights and implications could help instructors improve the design of MOOC courses.

Because of this research, I also developed the following algorithms, techniques, or systems working on the Android platform:

- **LensGesture** [133]: a mobile interaction technique that augments mobile interactions via finger gestures (e.g., covering the lens fully or partially, swiping across the lens) on the back camera of mobile devices. I implemented the recognition algorithms for different types of LensGestures as well as multiple LensGesture applications (Chapter 3.1.3) which demonstrate various usage scenarios of LensGesture as a new input channel. The source code of these applications can be downloaded from http://mips.lrdc.pitt.edu/lensgesture.
- **LivePulse** [53]: A commodity-camera-based photoplethysmography (PPG) sensing and heuristic-based heart rate measurement algorithm working on unmodified smartphones. I have developed both a mobile application running LivePulse algorithm to measure heart rate in real-time and a PC application for debugging the LivePulse algorithm. Both are open source software released under BSD license and the implementation can be downloaded from

http://mips.lrdc.pitt.edu/bayesheart. In order to make the heart rate monitoring task more engaging, my colleagues in the University of Pittsburgh and I have also designed serious mobile games (LivePulse Games) that leverage the LivePulse algorithm and integrate heart beat measurement implicitly in the game play [53]. The implementation of LivePulse Games can be downloaded from http://mips.lrdc.pitt.edu/livepulsegames/.

- **AttentiveLearner** [130, 134]**:** a mobile MOOC application which collects and analyzes learners' PPG signals to infer their cognitive states while they watch lecture videos. I developed the AttentiveLearner mobile client based on the open source edX Android application (https://github.com/edx/edx-app-android). Similar to existing MOOC mobile clients by Coursera, edX, and Udacity, the AttentiveLearner mobile application allows learners to browse, stream, and watch lecture videos on their mobile phones. Moreover, I integrated the tangible video control channel (Chapter 3) and implicit PPG sensing module (Chapter 4) in the mobile client. The AttentiveLearner mobile client should be available for everyone to download at http://www.attentivelearner.com in the near future.
- **C2F2** [131]**:** an adaptive intervention technique which monitors learners' engagement while they watch lecture videos and adaptively reminds learners of important upcoming content when they are disengaged. I implemented the C2F2 technique and integrated it into the AttentiveLearner mobile client. AttentiveLearner loads a pre-built engagement prediction classifier (Chapter 7.3.1), uses it to predict learners' engagement states after they watch each topic, and triggers C2F2 when the conditions are met. The AttentiveLearner mobile client with C2F2 adaptive intervention should be available to download at http://www.attentivelearner.com in the near future.

1.6 OUTLINE

Chapter 2 presents relevant technologies for improving MOOCs, provides a background on the role of affective-cognitive states in learning, and introduces related work on using physiological signals to detect and react to students' affective-cognitive states in educational systems.

Chapter 3 and **Chapter 4** focus on the two enabling technologies of AttentiveLearner. Chapter 3 presents *LensGesture* [133], a general interaction technique which leverages on-lens finger gestures to interact with mobile devices. I optimize LensGesture specifically for the tangible video control channel of AttentiveLearner. Through off-line benchmarking and an 18-subject user study, I verified that the tangible video control channel was both accurate and responsive (Hypothesis A.1). Chapter 4 demonstrates *LivePulse*, a real-time heart rate measurement algorithm based on commodity-camera-based photoplethysmography (PPG) sensing. LivePulse enables the implicit PPG sensing module of AttentiveLearner on unmodified smartphones. I conducted a *12-participant* study to test the accuracy and reliability of LivePulse, which proved Hypothesis B.1. Furthermore, through another *18-participant* user study and follow-up analyses, I showed that the tangible video control interface in AttentiveLearner was intuitive, and comfortable to use (Hypothesis A.2). AttentiveLearner could also collect reliable PPG signals during actual MOOC learning sessions (Hypothesis B.2).

Chapter 5 and **Chapter 6** focus on cognitive state inference via implicit PPG sensing on unmodified mobile phones [130, 132, 134]. Chapter 5 presents an *18-participant* study which demonstrates the feasibility of detecting boring and confusing topics in MOOC videos for individual learners, while Chapter 6 investigates the impact and detection of divided attention in the context of mobile MOOC learning through a second *18-participant* study. These two chapters support Hypothesis C.1.

Chapter 7 proposes an intervention technology which builds upon AttentiveLearner. This new technology, Context and Cognitive State triggered Feed-Forward (*C2F2*), proactively reminds a learner of important upcoming content (the feed-forward intervention) when disengagement is detected. A *48-participant* user study was performed, and I found that C2F2 yielded superior learning performance compared with a standard non-interactive learning system [131]. This result verified Hypothesis D.2. The effectiveness of C2F2 also validated Hypothesis D.1, showing the feasibility and efficacy of building end-to-end, affect-aware mobile MOOC systems on top of AttentiveLearner.

Chapter 8 presents a *22-participant* study to understand the dynamic transitions of affective states during a MOOC learning session. I show that MOOC learning has a different model of affect dynamics from complex learning (Hypothesis E.1). Also, I demonstrate the feasibility of using implicit PPG sensing to detect moment-to-moment affective states (Hypothesis E.2).

Chapter 9 summarizes the contributions of this dissertation and presents a discussion of future work.

2.0 BACKGROUND

2.1 TECHNOLOGIES FOR IMPROVING MOOCS

The passive, asynchronous, distributed viewing experiences of MOOCs present unique challenges to quality education. To address these challenges, researchers have proposed many new techniques, which can be grouped into three categories based on the goal of the technique (Table 1).

Table 1. Existing techniques for improving MOOCs.

Technique Category	Examples	Benefits	Limitations
Improving video quality and interactivity	Video annotation (overlay video content [26], sub-goal labels [122], digital footnote [71])	Augment the video viewing experience; improve student engagement	No personalization; require extra video production effort
	Video navigation (learner activity augmented timeline [65], NoteVideo [82], Video Digest [88], QuizCram [69])		
	Video interactivity (in-video exercises [64], embedded comments [81, 83])		

Enhancing communication	Asynchronous communication (discussion forums [4, 20]; post-lecture reflections [45, 47, 48])	Promote social interactions and create a feeling of community; generate feedback for instructors; promote in-depth and thoughtful discussion and reflections	Passive student “lurkers” do not participate in the activities (low forum participation rate in MOOC [16]); asynchronous methods lack immediate feedback; synchronous methods may not allow deep reflections
	Synchronous communication (chat systems [21])		
	Hybrid (time-anchored commenting [72])		
Post-hoc clickstream and major video event analysis	Activities within learning sessions (video event analysis [51, 66, 67])	Reveal insightful information to understand MOOC learners	Log analysis reveal learners' actions rather than actual cognitive states in learning.
	Activities in the follow-up discussion forums ([136, 137]) and exercises		
	Activities in the “course-level” ([22, 123])		

The first group of techniques aims to enhance learner engagement by improving the quality and interactivity of MOOC videos. Researchers have used video annotations to augment interactions with the video [26, 122] and enhance the video viewing experience [71]. Furthermore, various navigation controls have been proposed to help users browse and skim videos [65, 69, 82, 88]. For example, Kim et al. [65] designed a learner activity augmented timeline to facilitate video navigation. The timeline visualizes learner interaction peaks based on historical interactions and enables non-linear scrubbing through friction. Kovacs [69] proposes the QuizCram interface which enables question-directed video viewing. Users navigate through the video segments by answering questions. Moreover, there are techniques which add interactive elements in the video, such as interactive exercises [64] and embedded comment threads/assessment [81, 83]. For example, L.IVE by Monserrat et al. [81] provides in-situ learning, discussion, and assessment via an interactive overlay directly on top of the lecture video. RIMES [64] supports interactive, multimedia responses (handwriting, audio, and video) in lecture videos. However, one problem of these interaction techniques is a lack of personalization for individual learners. Some techniques (e.g., video annotation, interactive exercises, etc.) also require extra video production effort.

The second group of techniques aims to promote communications among students and instructors in the system. These include asynchronous communication techniques, such as discussion forums [4, 20] and post-lecture reflections [45, 48]; synchronous communication techniques, such as chat-room systems [21]; and hybrid techniques combining elements of both synchronous and asynchronous communication, such as time-anchored commenting [72]. These techniques can either promote student-student interactions [21, 72], or support student-instructor feedback [45, 48]. For example, Glassman and colleagues [48] invented the Muddy Card technique to allow students to indicate confusing concepts (muddy points) on the corresponding lecture slide. Such a technique provides a way for students to efficiently and specifically express their confusion to instructors. However, most of these interactive techniques rely heavily on learners' active participation while prior research indicated low participation rates of activities or class discussions in MOOC contexts [16].

The third group of techniques is post-hoc analysis of clickstream data and major video interaction events. Researchers have analyzed both activities within the learning session (e.g., click-level interactions [51, 66, 118], visual attention [67], and exercise events [73]), activities in the follow-up discussion forums [136, 137], and “course-level” activities and events [22, 123]. For example, Kim et al. [66] applied temporal pattern analysis techniques on video play activities and found that students tended to selectively pick parts of the video to watch and 61% of the interaction peaks involved a visual transition that proceeds or after the peak. Yang et al. [137] used learners’ discussion forum behavior and clickstream data to identify posts that express confusion. These analyses can reveal insightful information about MOOC learning, such as the correlation of video production and student engagement [51, 67], and factors that contribute to course dropout [51, 136, 137]. Although server-side activity logs are easy to collect and can reveal insightful information, mouse clicks and keystrokes are relatively sparse in a single learning session. More importantly, they reveal learners' *actions* rather than actual *cognitive states* in learning. There is still little *direct* measurement of learners’ actual learning process in MOOCs.

This dissertation explores the continual collection and use of spontaneous PPG signals and heart rate as a fine-grained feedback channel in MOOC learning. Such physiological signals correlate directly with learners' physiological states and cognitive states [96] and can complement today's log analysis techniques.

2.2 AFFECTIVE-COGNITIVE STATE AND LEARNING

It has been widely acknowledged that cognition, motivation, and emotion are the key components of learning [2, 55, 128]. Previous studies have shown that learners experience a rich diversity of learning-centered affective states, including *Engagement/Flow*, *Boredom*, *Confusion*, *Curiosity*, *Happiness*, and *Frustration* in the process of learning [2, 7, 29]. Affective-cognitive states are highly relevant and influential to both the processes and outcomes of learning. Experimental mood studies have found that affect influences a broad variety of cognitive process that contributes to learning, such as perception, attention, cognitive-problem solving, decision making, and memory processes [90]. Positive affect (e.g., flow and confusion) could promote rational processing of information and induce relational processing [92], while negative affect (e.g., *Boredom* and *Frustration*) "*produce task-irrelative thinking, thus reducing cognitive resources available for task purposes, and undermine students' intrinsic motivation*" [90]. For example, *Boredom*, a commonly observed affective state during learning, could disengage learners from educational activities and seriously decrease learners' abilities to acquire knowledge [119]. *Boredom* was also found to be an antecedent to gaming the educational system [7] and negatively correlated with learning performance [7, 33].

2.3 DETECTING AND RESPONDING TO AFFECTIVE-COGNITIVE STATE IN EDUCATION

Because of the important correlation between affective-cognitive states and learning, researchers have built various learning environments which detect and respond to learners' affective and cognitive states [23, 36, 42, 43, 58, 98, 102, 115, 116, 128, 140]. These systems collect physiological signals, such as heart rates [52, 55, 59, 127], galvanic skin responses (GSR) [12, 127, 128], facial expressions [61, 128, 135], eye gazes [12, 42, 58, 98], and EEG signals [98, 115, 116, 127, 140], etc., and use machine learning algorithms to predict students' affective and cognitive states (e.g., attention, boredom, confusion, mind wandering, etc.) while they are interacting with the system. These affect-sensitive systems then dynamically respond to the sensed affective and cognitive states using pedagogical strategies, such as direct feedback [34, 35, 42], pedagogical agent [6, 128, 135], and adaptive activities [116].

AutoTutor [35, 36, 55] is a pioneer in detecting and adapting to learners' affect from multi-channel physiological signals in an intelligent tutoring system. The authors used supervised machine learning algorithms to achieve satisfactory performance on affect classification (Kappa = 0.35 for predicting valence, Kappa = 0.23 for predicting arousal). A set of production rules were designed to dynamically map students' cognitive and affective states with appropriate tutor actions. Through an 84 participant between-subject study, the authors found the affect-sensitive AutoTutor more effective for low-domain knowledge students. Low domain knowledge students learned significantly more from the affect-sensitive AutoTutor (54.9% learning gains) than the regular tutor (38.2% learning gains); while the students with more knowledge didn't benefit from the affect-sensitive AutoTutor (19.8% learning gains vs. 37% learning gains).

To predict students' affective states, the Wayang intelligent tutor [6, 128] used four types of sensor data: facial expressions, seat pressure, mouse pressure and GSR. Empirical studies showed that the animated affect-aware agents improved average learning gains by 12% after only two classes and had benefits such as improving learners’ self-concept and making them more engaged.

Gaze Tutor [42] monitors a student’s gaze patterns to identify when the student is disengaged or zoning out. The tutor attempts to re-engage the student with direct gaze-reactive statements. A 48 participant evaluation study showed that students provided more accurate responses to deep reasoning questions when they interacted with the gaze-reactive tutor (31.3% learning gains) than the non-gaze-reactive tutor (0% learning gains). Furthermore, the gaze-reactive statement associated with a significant improvement for students with high aptitude, while not as effective for students with average aptitude.

Szafir et al. [116] developed an adaptive-review technology which monitored learners' attention using their EEG signals and adaptively provided reviews on topics with low-attention levels. Results of a 48-participant user study showed that the adaptive review technology significantly increased learning gains compared with the no review baseline (57.35% vs. 39.71%), while no difference were found in learning between the adaptive and full conditions (57.35% vs. 58.33%).

One common problem with most of these systems [36, 42, 116, 128] is the requirement of dedicated sensors, such as cameras, chest bands or EEG headsets to collect physiological signals. The cost, availability, and portability of such equipment have prevented the wide adoption of such systems beyond lab settings. A “sensorless” approach to collect physiological signals is necessary in order to make the cognitive-aware learning systems widely adopted beyond lab settings.

3.0 LENSGESTURE AND THE TANGIBLE VIDEO CONTROL CHANNEL

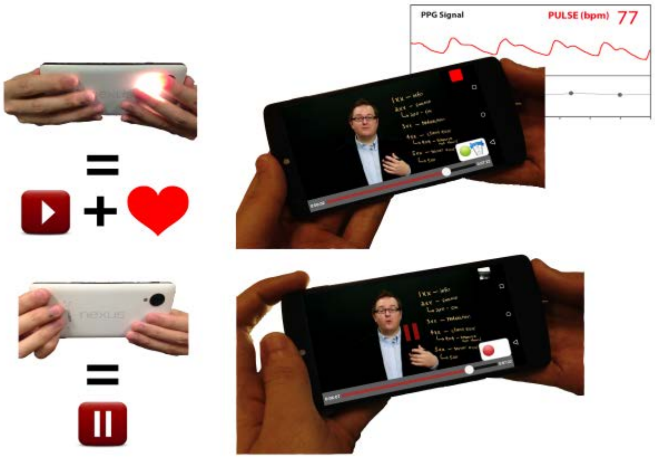


Figure 3. AttentiveLearner uses the back camera as both a tangible video control channel and an implicit heart rate sensing channel in MOOC learning.

In AttentiveLearner, on-lens finger gestures are used as an intuitive control mechanism for video playback (Figure 3). A learner plays the instructional video by covering and holding the back camera lens with her fingertip, pausing the video by uncovering the lens. Such a tangible video control mechanism has two advantages when compared with traditional touchscreen widgets: 1) the "cover-and-hold-to- play" mechanism enables more flexible video control; 2) this approach allows the implicit extraction of PPG signals via commodity-camera-based photoplethysmography (PPG) [53] in MOOC learning.

To implement the tangible video control interface, I begin with a more general question, that is, “*can we use finger gestures on the back camera lens to support general mobile* *interactions?"* I did a systematic exploration of the design space of on-lens finger gestures and proposed a general interaction technique, LensGesture, which leverages finger gestures on the back camera lens to support mobile interactions. In this section, I first describe the design, implementation, and evaluation of LensGesture. Then, I optimize and evaluate LensGesture specifically for controlling video playback in the context of AttentiveLearner. Implementation of LensGesture can be downloaded from http://mips.lrdc.pitt.edu/lensgesture. The contents of this chapter can be found in the published papers [130] and [133].

3.1 LENSGESTURE

LensGesture is initially motivated by the challenges faced by one-handed mobile interaction. When a user interacts with her phone with one hand, the user's thumb, which is neither accurate nor dexterous, becomes the only channel of input for mobile devices, leading to the notorious "fat finger problem" [10, 138], the "occlusion problem" [10, 124], and the "reachability problem" [126]. In contrast, the more responsive, precise index finger remains idle on the back of mobile devices throughout the interactions. Because of this, many compelling techniques for mobile devices, such as multi-touch, became challenging to perform in such a "situational impairment" [107] setting.

Many new techniques have been proposed to address these challenges, from adding new hardware and new input modality [10, 113, 124, 125], to changing the default behavior of applications for certain tasks [138]. Due to challenges in backward software compatibility, availability of new sensors, and social acceptability [99], most of the solutions are not immediately accessible to users of existing mobile devices.

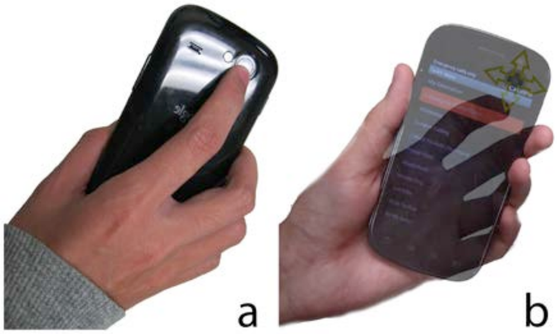


Figure 4. LensGesture in use for menu navigation.

Motivated by these challenges, I propose LensGesture (Figure 4), a mobile interaction technique which provides a new input channel via finger gestures on the back camera of mobile devices.

3.1.1 The LensGesture Taxonomy

I propose two groups of interaction techniques, *Static LensGesture* and *Dynamic LensGesture*, for finger initiated direct touch interactions with mobile cameras (Figure 5).

Static LensGesture (Figure 5, top row) is performed by covering the camera lens either fully or partially. Supported gestures include covering the camera lens in full (i.e., full covering gesture) and covering the camera lens partially (e.g., partially covering the left, right, and bottom region of the lens[1]). *Static LensGesture* converts the built-in camera into a multi-state push

[1] According to informal tests, I found the top-covering gesture both hard to perform and hard to distinguish (when compared with left-covering gestures, Figure 6, third row, first and last images). So I intentionally removed the top-covering gesture as a supported *Static LensGesture*. Please also note that the definition of "top", "left", "right" and "bottom" depends on the holding position (e.g., portrait mode or landscape mode) of the phone.

button set. Interestingly, the edge/bezel of the camera optical assembly can provide natural tactile feedback to the user's index finger when performing static gestures.

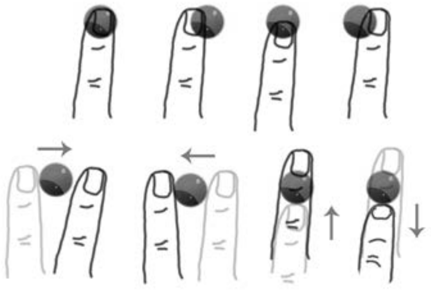


Figure 5. The LensGestures Taxonomy.

Top: Static LensGestures; Bottom: Dynamic LensGestures.

A user can also perform a *Dynamic LensGesture* (Figure 5, bottom row) by swiping her finger horizontally (left and right) or vertically (up and down) across the camera lens. *Dynamic LensGestures* convert the back camera into a four-way, analog pointing device based on relative movement sensing.

3.1.2 The LensGesture Algorithm

I designed a set of three algorithms to detect full coverage, partial coverage and dynamic swiping of fingers on the lens. Depending on usage scenarios, these three algorithms can be cascaded together to support all or part of the LensGesture set. In all LensGesture detection algorithms, the camera is set in preview mode, capturing 144x176 pixel color images at a rate of 30 frames per second. I disable the automatic focus function and the automatic white balance function to avoid interference with the algorithms.

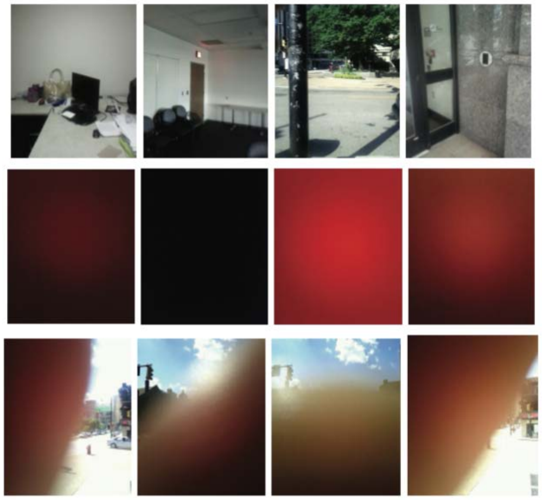


Figure 6. Sample images of *Static LensGesture*.

First row: no gesture. Second row: full covering gestures. Third row: partial-covering gestures. Left to right: left-covering, right-covering, bottom-covering, and top-covering (not supported).

Static LensGesture - Full covering: The full covering gesture (Figure 6, second row) can be detected quickly and reliably via a linear classification model on the global mean and standard deviation of all the pixels in an incoming image frame in the 8-bit grayscale space. The intuition behind the underlining detection algorithm is that when a user covers the camera's lens completely, the average illumination of images drops, while the illumination among pixels in the image will become homogeneous (i.e., smaller standard deviations).

Figure 7 shows a scatter plot of global mean vs. global standard deviation of 791 test images (131 contained no LensGesture; 127 contained *full-covering gestures*; 533 contained *partial covering gestures*). I collected test images from 9 subjects and in four different environments: 1) indoor bright lighting, 2) indoor poor lighting, 3) outdoor direct sunshine, and 4) outdoor in the shadow. All the subjects in the data collection stage were undergraduate and graduate students at the University of Pittsburgh, recruited through school mailing lists. The number of samples in each environment condition is evenly distributed. When I choose mean <=

100, stdev <=30 as the linear decision boundaries for detecting full-covering gestures (highlighted in Figure 7), I can achieve an accuracy of 97.9%, at a speed of 2.7ms per estimate. While more advanced detection algorithms could improve the accuracy, an accuracy of 97.9% is sufficient in interactive applicants where users can adapt their behaviors via real-time feedback.

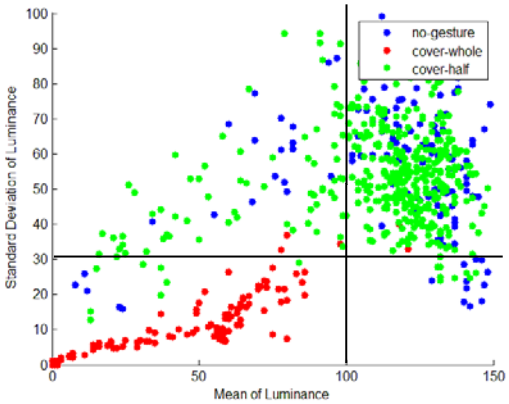


Figure 7. The full-covering LensGesture classification.

Global mean vs. standard deviation of all the pixels in images with (full-covering: red dots, partial covering: green dots) and without (blue dots) Static LensGestures. Each dot represents one sample image.

Static LensGesture - Partial covering: To detect partial covering gestures in real time, I designed three serial cascaded binary kNN (k=5) classifiers to detect covering-left, covering-bottom, and covering-right gestures. After deciding that the current frame does not contain a full covering gesture, the image will be fed to the covering-left, the covering-bottom, and the covering-right classifier one after the other. If a partial covering gesture is detected, the algorithm will stop immediately, if not, the result will be forwarded to the next binary classifier. If no partial-covering gesture is detected, the image will be labeled as "no gesture". I adopted this cascading approach and the kNN classifier primarily for speed concerns.

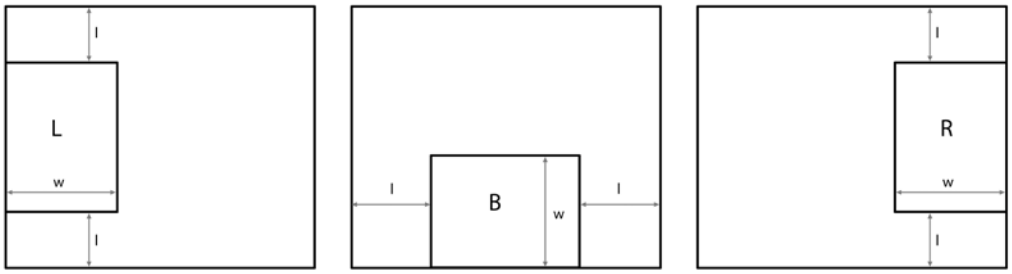


Figure 8. Extracting local features for the partial-covering LensGestures.

From left to right, extracting local features from Region L (*covering-left* classifier), Region B (*covering-bottom* classifier), and Region R (*covering-right* classifier).

The features I used in the kNN classifiers include both global features (mean, standard deviation, maximal and minimal illuminations in the image histogram) and local features (same features in a local bounding box, defined in Figure 8). There are two parameters (w, l) that control the size and location of the local bounding boxes. The (w, l) values (unit=pixels) should be converted to a relative ratio when used in different preview resolutions.

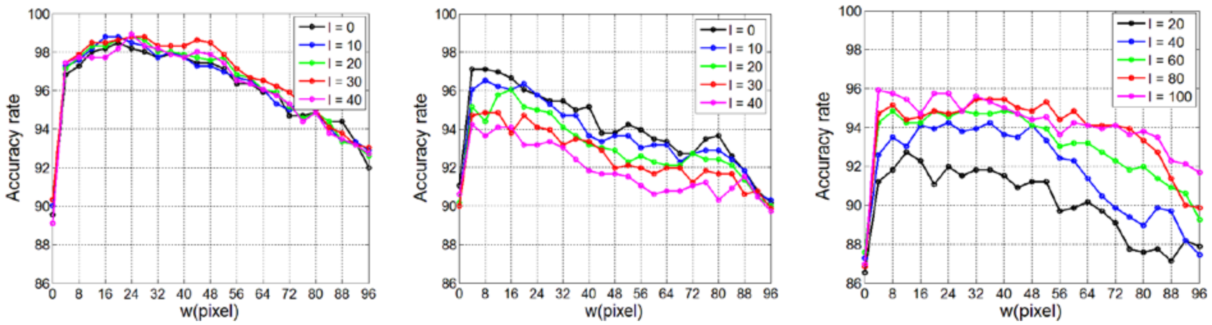


Figure 9. Classification accuracies of partial-covering classifiers.

Left to right: covering-left, covering-bottom, covering-right.

I use the data set described in the previous section, and ten-fold classification to determine the optimal values (w and l) for each classifier (Figure 9). As shown in Figure 9, I found that for the *covering-left* classifier, w = 24, l = 40 will give us the highest binary classification accuracy at 98.9%. For the *cover-bottom* classifier, w = 4, l = 0, gives the highest accuracy at 97.1%, for the *covering-right* classifier, w = 4, l = 100, gives the highest accuracy at 95.9%. The overall accuracy of the cascaded classification is 93.2%. The speed for detecting partial covering ranges from 16 – 42 ms.

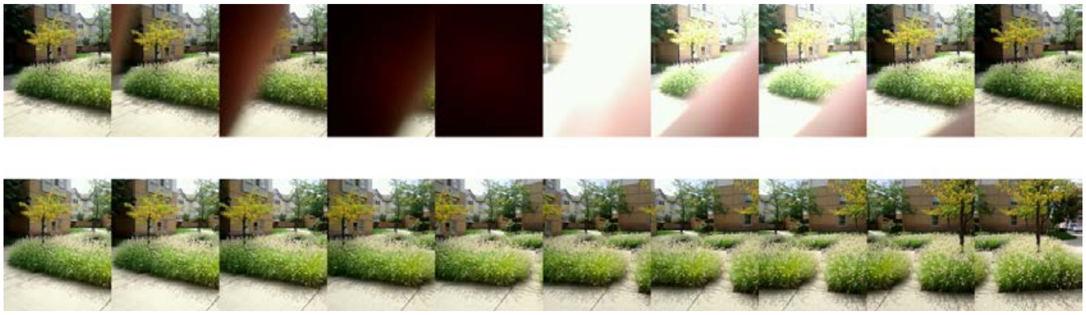


Figure 10. The difference between image sequences captured by LensGesture (up) and TinyMotion (down) in the same scene.

Dynamic LensGesture: The *Dynamic LensGesture* algorithm is based on the TinyMotion algorithm [121] with minor changes and additional post-processing heuristics. TinyMotion [121] detects the movement of a cell phone in real time by analyzing image sequences captured by its built-in camera. The TinyMotion algorithm detects motion by calculating the block shifting distance between two temporally adjacent frames. It applies a Full-search Block Matching algorithm (FBMA) [5] to compare the pixel blocks shifted in the current frame with corresponding pixels in the previous frame. A motion vector $\overrightarrow{MV} = (MV_x, MV_y)$ is calculated to represent the relative distance changes in the x and y directions from the previous frame. The source code of TinyMotion can be downloaded at http://people.cs.pitt.edu/~jingtaow/tinymotion/download.html. As reported by Wang, Zhai, and

Canny in [121], TinyMotion users discovered that it was possible to put one's other hand in front of the mobile camera and control motion sensing games by moving that hand (Dynamic LensGestures) rather than moving the mobile phone (TinyMotion). However, as shown in Figure 10, the fundamental causes of image change are quite different in TinyMotion and LensGesture. In TinyMotion (Figure 10, bottom row), the algorithm detects the background shifting caused by lateral movement of mobile devices. When performing Dynamic LensGestures (Figure 10, top row), the background keeps almost still while the fingertip moves across the lens. Another important observation is that in Dynamic LensGesture, a user's finger will completely cover the lens in one or two frames, making brute force motion estimation results noisy. Therefore, I need to modify the TinyMotion algorithm in order to detect *Dynamic LensGestures*.

Figure 11 shows the relative movements from the TinyMotion algorithm, as well as the actual images captured when a left-to-right *Dynamic LensGesture* was performed. In Figure 11, I see that although the TinyMotion algorithm successfully captured the strong movements in the x-axis (frames 3, 4, 5, 7, 8, 10, 11), estimations became less reliable (frame 6) when a major portion of the lens was covered. To address this issue, I use a variable weight-moving window to process the raw output from the TinyMotion algorithm. I give the output of the current frame a low weight when a full covering action is detected. The Dynamic LensGesture detection algorithm works as follows: After the algorithm first detects a relative movement in consecutive image frames captured by the camera, it calculates and accumulates the weighted relative movements in the x and y axes on the following image frames until no movement is detected, meaning that the Dynamic Gesture has been completed. The algorithm then determines the direction of the Dynamic Gesture based on which direction has the overall largest relative movement.

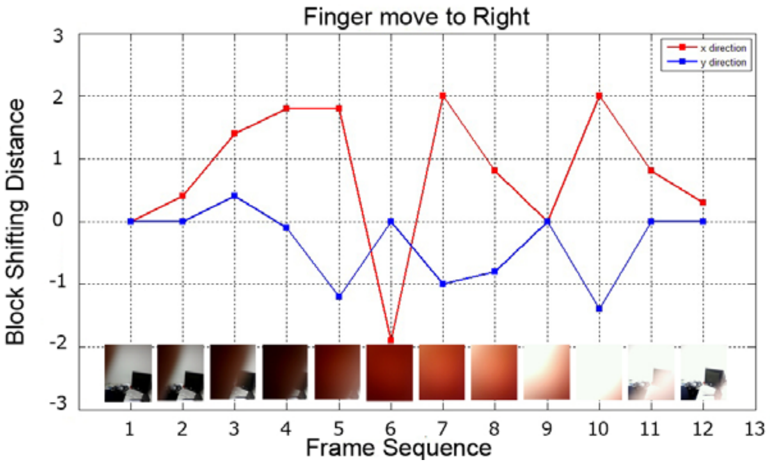


Figure 11. The plot of the distant changes in both x and y directions for 20 image samples of a Dynamic LensGesture.

I collected 957 sets of *Dynamic LensGesture* sample from 12 subjects (241 *move-to-right* gesture samples, 240 *move-to-left* gesture samples, 238 *move-up* gesture samples, and 238 *move-down* gesture samples). The subjects were asked to hold the mobile phone with either their left hand or right hand, and make each type of *Dynamic Gestures* on the back camera lens for 20 times while the camera captured the image frames at 30Hz. All subjects were undergraduate or graduate students at the University of Pittsburgh, recruited through school mailing lists. Two *move-up* gesture samples and *move-down* gesture samples were removed due to the poor image quality. There were more than 30000 images in this data set. For each Dynamic LensGesture, depending on the finger movement speed, 10-20 consecutive images were usually captured. I achieve an accuracy of 91.3% for detecting *Dynamic LensGestures* on this dataset, at a speed of 3.9 ms per estimate. I looked deeper into the misclassified sample sequences and found that most errors were caused by the confusion between the swiping down and the swiping left gestures. Most of the misclassified sequences looked confusing even to human eyes because the actual swiping actions were diagonal rather than vertical or horizontal. I attribute this issue to the relative positioning of the finger and the lens, as well as the lack of visual feedback during data collection.

3.1.3 The LensGesture Applications

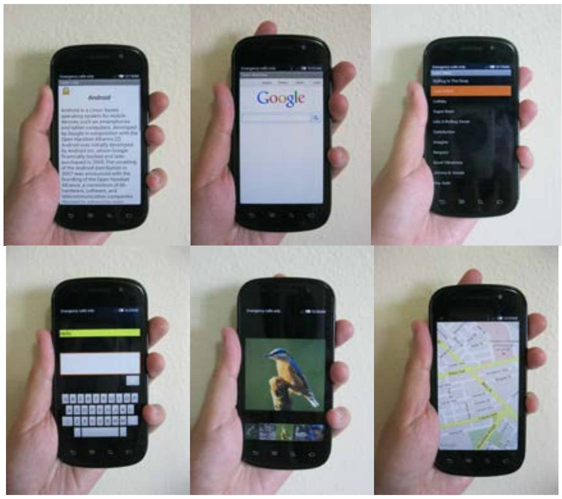


Figure 12. Sample LensGesture applications. From left to right, top to bottom - LensLock, LensCapture, LensMenu, LensQWERTY, LensAlbum, and LensMap.

To explore the efficacy of LensGesture as a new input channel, I wrote six applications (LensLock, LensCapture, LensMenu, LensQWERTY, LensAlbum, and LensMap). All these prototypes can be operated by Static or Dynamic LensGestures (Figure 12). All but one application (LensQWERTY) can be operated with one hand.

LensLock leverages the Static LensGesture and converts the camera into a "clutch" for automatic view orientation changes. When a user covers the lens, LensLock locks the screen at the current landscape/portrait format until the user's finger releases from the lens. LensLock can achieve the same "pivot-to-lock" technique proposed by Hinckley [54] without using the thumb finger to touch the front screen, which may lead to unexpected state changes.

LensQWERTY uses Static LensGesture to control the SHIFT state of a traditional on screen QWERTY keyboard. The user can use the hand holding the phone to toggle the SHIFT state when the other index finger is being used for typing.

LensAlbum and LensMap are two applications that leverage Dynamic LensGestures for one-handed photo album/map navigation. These two application shows that LensGesture can alleviate “fat finger problem” and the “occlusion problem” by avoiding direct thumb interaction on the touch screen. The LensMenu also illustrates a feasible solution to the "reachability problem" via a supplemental back-of-device input channel enabled by LensGestures.

3.1.4 Implementation

I implemented LensGesture on a Google Nexus S smartphone. I wrote the LensGesture algorithms and all the LensGesture applications in Java. The LensGesture algorithm can be implemented in C/C++ and compiled to native code via Android NDK if higher performance is needed. Source code of LensGesture can be downloaded from http://mips.lrdc.pitt.edu/lensgesture.

3.1.5 Evaluation

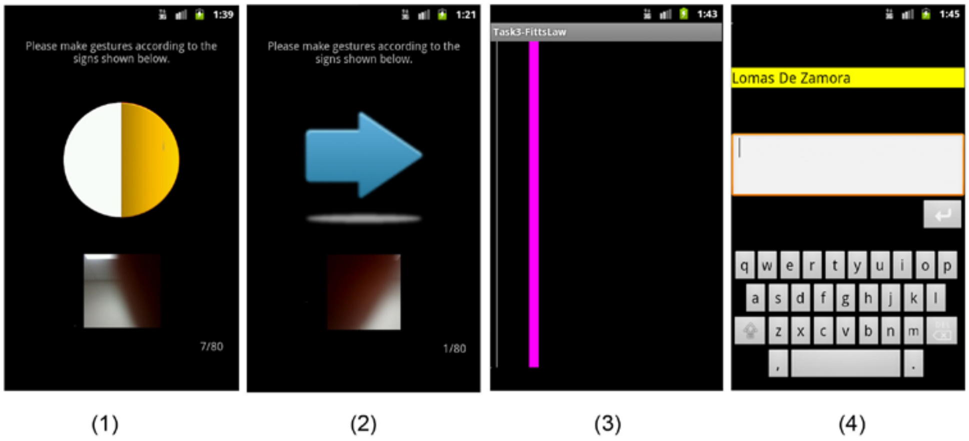
Although the results of our LensGesture algorithm on pre-collected data sets were very encouraging, a formal study was necessary to understand the capabilities and limitations of LensGesture as a new input channel.

3.1.5.1 Experiment Design

The whole study took less than one hour, and each participant was compensated with a $10 gift card after completing all tasks. The study consisted of six parts:

Overview. I first gave participants a brief introduction to the LensGesture project. I explained each task to them, and answered their questions.

Reproducing LensGestures. This session was designed to test whether users could learn and comfortably use the LensGestures I designed, and how accurate/responsive the gesture detection algorithm was in a real-world setting. A symbol representing either a *Static LensGesture* or a *Dynamic LensGesture* was shown on the screen (Figure 13, (1) (2)). Participants were required to perform the corresponding LensGesture with their index fingers as fast and as accurately as possible. The application would still move to the next stimulus if a user could not perform the expected gesture within the timeout threshold (5 seconds). A user completed 20 trials for each supported gesture. The order of the gestures was randomized.


Please make gestures according to the signs shown below.

Please make gestures according to the signs shown below.

Task3-FittsLaw

Lomas De Zamora

Figure 13. Screen shots of applications in the user study.

Target Acquisition/Pointing. The goal of this session was to quantify the human performance of using LensGesture to perform target acquisition tasks. For each trial, participants needed to use *Dynamic LensGestures* to drive an on-screen cursor from its initial position to the target (Figure 13, (3)). After the cursor hit the target, participants were required to tap the screen to complete the trial. Regardless of whether participants hit the target or not, the target acquisition screen disappeared and an information screen indicating the number of remaining trials in the current block would show up. I encouraged participants to hit the target as fast as possible and as accurately as possible. Each participant completed 160 randomized trials.

Text Input. In this task, I compared the performance of standard Android virtual keyboard with the LensQWERTY keyboard (Figure 13, (4)).

Each participant entered 13 short phrases in each condition. The 13 test sentences were: “Hello”, “USA”, “World”, “Today”, “John Smith”, “Green Rd”, “North BLVD”, “Lomas De Zamora”, “The Great Wall”, “John H. Bush”, “Sun MicroSystem”, “Mon Tue Wed Thu”, and “An Instant In The Wind”. These test sentences were intended to maximize the usage of LensGesture based shifting feature and simulate commonly used words in a mobile environment (person names, place names, etc.).

Other Applications. In this session, Participants were presented with five LensGesture applications I created (LensLock, LensCapture, LensMenu, LensAlbum, and LensMap, Figure 12). After a brief demonstration session, I encouraged the participants to play with these applications as long as they wanted.

Collect Qualitative Feedback. After a participant completed all tasks, I asked him or her to complete a questionnaire (B.1). I also asked the participant to comment on each task, and describe one’s general feeling towards LensGesture.

3.1.5.2 Participants and Apparatus

16 subjects (4 females) between 22 and 30 years of age participated in our study. 15 of the participants owned a smartphone. The user study was conducted in a lab with abundant light. All of the participants completed all tasks.

Our experiments were completed on a Google Nexus S smartphone with a 480 x 800 pixel display, a 1GHz ARM Cortex-A8 processor, running Android 4.0.3. It has a built-in 5.0 mega-pixel back camera located at the upper right region.

3.1.5.3 Evaluation Results

Reproducing LensGestures

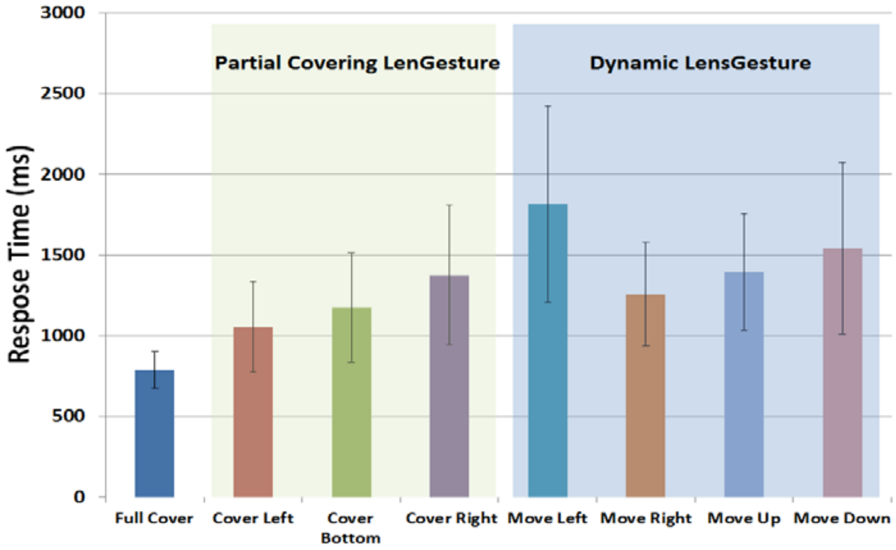


Figure 14. Average response time of Static and Dynamic LensGestures with one standard deviation error bars.

As shown in Figure 14, the time needed to perform a static gesture varied on gesture type. Repeated measure variance analysis showed significant difference due to gesture type: F(7, 120) = 9.7, p < .0001. Fisher’s post hoc tests showed that the response time of full-occlusion gesture

(787 ms) was significantly shorter than any of the partial occlusion gestures (left = 1054 ms, $p < 0.01$; right = 1374 ms, $p < 0.0001$; bottom= 1175 ms, $p < 0.0001$) and dynamic gestures. The left partial occlusion gesture is significantly faster than right partial occlusion, $p < 0.01$, the speed differences between other partial occlusion gestures are not significant. For Dynamic Gestures, the move-right gesture (1258.6 ms) was significantly faster than move-left (1815.2 ms, $p < 0.01$) and move-down (1540.6 ms, $p < 0.05$) gestures, but there was no significant time difference between move-right and move-up (1395.7 ms, p= 0.15). The move-up gesture was also significantly faster than move-left ($p < 0.01$). The differences in detection time of Dynamic LensGestures might be caused by the location of the camera. The camera was located on the upper right region of the experiment device, making it easier to make the move-right and move-up gesture.

Target Acquisition/Pointing

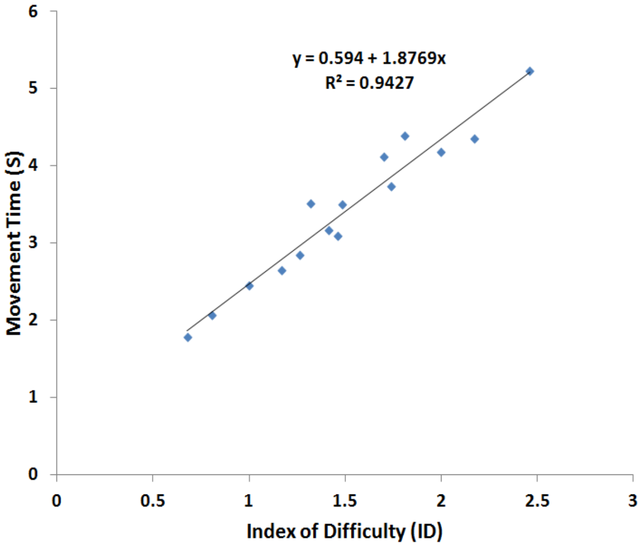


Figure 15. Scatter-plot of the Movement Time (MT) vs. the Fitts' Law Index of Difficulty (ID) for the overall target acquisition task controlled by Dynamic LensGestures.

2560 target acquisition trials were recorded. 2298 pointing trials were successful, resulting in an error rate of 10.2%. This error rate is about twice as that of popular pointing devices in Fitts' law studies. After adjusting target width W for the percentage errors, linear regression between movement time (MT) and Fitts' index of difficulty (ID) is shown in Figure 15:

$$MT = 0.594 + 1.8769 \log_2(A/W_e+1) \quad \text{(sec)}$$

In the equation above, A is the target distance and W_e is the effective target size. While the empirical relationship between movement time (MT) and index of difficulty (ID = log (A/W_e + 1)) followed Fitts' law quite well (with R^2 = 0.9427, see Figure 15), the information transmission rate 1/b = 1/1. 1.8769 = 0.53 bits/sec) indicated a relatively low performance for pointing. In comparison, Wang, Zhai and Canny [121] reported a 0.9 bits/sec information transmission rate for device motion based target acquisition on camera phones. I attribute the performance difference to the usage patterns of *Dynamic LensGestures* - due to the relatively small touch area of the built-in camera, *repeated* finger swiping actions are needed to drive the on-screen cursor for a long distance. I believe that the performance of LensGesture could be improved with better algorithms and faster camera frame rates in the future. More importantly, since LensGesture can be performed *in parallel* with interaction on the front touch screen, I believe that there are opportunities to use LensGesture as a supplemental input channel and even use LensGesture as a primary input channel when the primary channel is not available.

Text Input

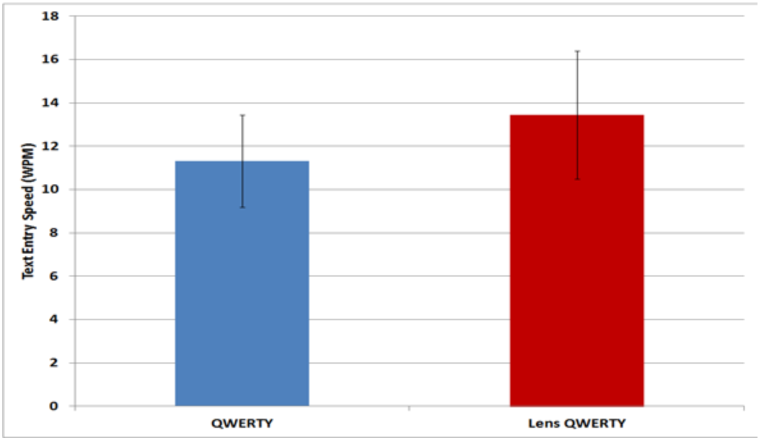


Figure 16. Text entry speed from the experiment with one standard deviation error bars.

As shown in Figure 16, the overall speed of LensGesture enabled virtual keyboard, i.e., LensQWERTY (13.4 wpm), was higher than that of the standard virtual keyboard (11.7 wpm). The speed difference between these two keyboards was significant F(1, 15) = 4.17, p < 0.005. The uncorrected error rate was less than 0.5% for each condition. The average error rates for the standard keyboard and LensQWERTY were 2.1% and 1.9% respectively. The error rate difference between the standard keyboard and LensQWERTY was not significant (p = 0.51).

Other Applications and Subjective Feedback

All participants can learn to use the LensGestures applications I provided with minimal practice (< 2 min). Almost all participants commented that the portrait/landscape lock feature in LensLock was very intuitive and much more convenient than alterative solutions available on their own smartphones. Participants also indicated that changing the "shift" state of a virtual keyboard via LensGesture was both easy to learn and time saving.

The participants reported positive experiences with LensGesture. All participants consistently rated LensGesture as “useful” on the closing questionnaire using a five-point Likert scale. When asked about how easy it was to learn and use LensGesture, 13 participants selected “easy”. 9 participants commented explicitly that they would use LensGesture on their smartphones. 4 of them expressed a very strong desire to use LensGesture applications every day.

3.1.6 Summary

LensGesture is a pure software approach for augmenting mobile interactions with back-of-device finger gestures. LensGesture detects full and partial occlusion as well as the dynamic swiping of fingers on the camera lens by analyzing image sequences captured by the built-in camera in real time. I report the feasibility and implementation of LensGesture as well as newly supported interactions. Both offline benchmarking results and a 16-subject user study show that LensGestures are easy to learn, intuitive to use, and can complement existing interaction paradigms used in today's smartphones.

3.2 OPTIMIZING LENSGESTURE FOR TANGIBLE VIDEO CONTROL

LensGesture is a general interaction technique and it could be used in various usage scenarios. For example, since the *Static LensGesture* basically converts the built-in camera into a multi-state push button set, it could be used as a switch to “*turn on/off*” various functions, such as lock/unlock the screen orientation, open/close applications. One specific use case for the *Static*

LensGesture is to control video play, i.e., covering the camera lens to play video, and uncovering the lens to pause the video. This LensGesture-based video control mechanism is adopted in AttentiveLearner.

There are two major differences between usage scenarios of LensGesture in Chapter 3.1 and those in AttentiveLearner. First, the flashlight of the mobile camera is on by default in AttentiveLearner to improve heart rate measurements in low illumination conditions[2]; second, the original *Static LensGesture* algorithm only determines the coverage of camera lens without differentiating whether the coverage was by a finger or inorganic surfaces (e.g., putting the phone on a desk).

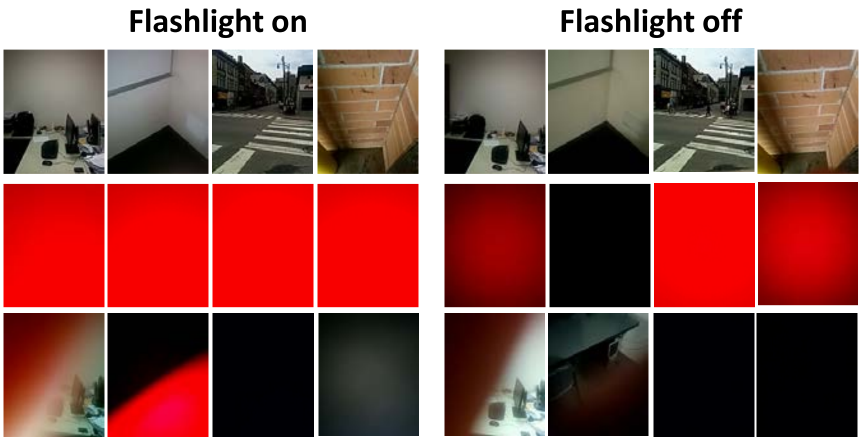


Figure 17. Samples images collected during the study.

First row: the lens is not covered. Second row: the lens is covered. Third row: the lens is partially covered (first two images) or blocked by surface (last two images).

To enhance the lens covering detection algorithm based on AttentiveLearner's unique requirements, I rebuilt the linear classification model of *Static LensGesture* detection. Again, I

[2] The flashlight in AttentiveLearner can be turned off if the environmental illumination is sufficient.

collected 483 representative test images in 4 (lens fully covered by finger, lens partially covered by finger, lens blocked by opaque surface, lens uncovered) x 2 (flashlight on, flashlight off) x 4 (indoor high illumination, indoor low illumination, outdoor direct sunshine, and outdoor in the shade) conditions from 10 subjects. All the subjects were graduate students at the University of Pittsburgh, recruited through school mailing lists. Figure 17 shows some sample images collected.

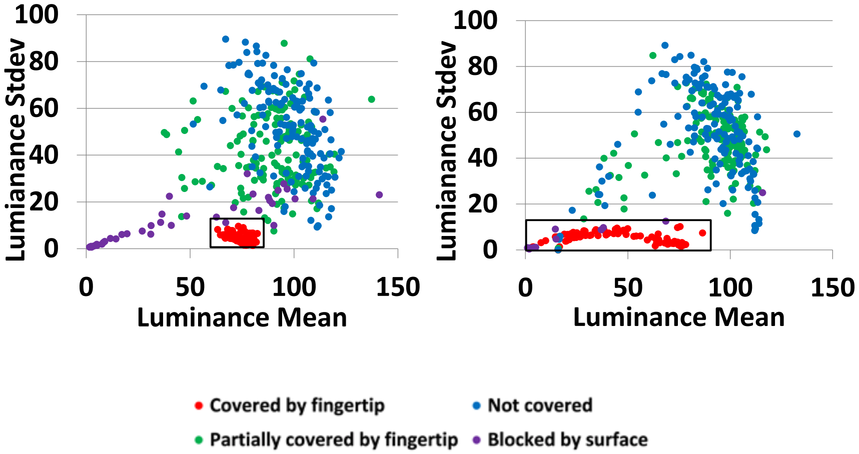


Figure 18. Global mean vs. standard deviation of pixels in sample images.

No-covering: blue dot, full-covering-by-fingertip: red dot, partial-covering-by-fingertip: green dot, blocked-by-surface: purple dot. Left: samples when the flashlight was on. Right: samples when the flashlight was off.

Figure 18 shows scatter plots of global mean vs. global standard deviation of all test images when the flashlight was on (left) and off (right) respectively. I found that when the flashlight was turned on, the full-covering-by-fingertip samples (inside the black rectangle) are more aggregated than when the flashlight was off. I also observed turning on the flashlight can significantly reduce the variations caused by environmental illumination (Figure 17's left second row vs. Figure 17's right second row). I achieved an accuracy of 99.59% when using 60 ≤ mean ≤ 90, stdev ≤ 15 as the decision boundaries for detecting lens-covering gestures. The only two misclassifications were false positives when the mobile phone was on a red semi-transparent plastic surface (of a plastic flying disc). AttentiveLearner can reject such false positives by incorporating output from the heart rate sensing module since non-body parts cannot generate periodic transparency changes expected by the PPG detection algorithm.

3.3 USABILITY

Other than accuracy, I also systematically investigated the usability of the lens-covering based video control interface from other aspects, such as speed and battery life.

3.3.1 Speed

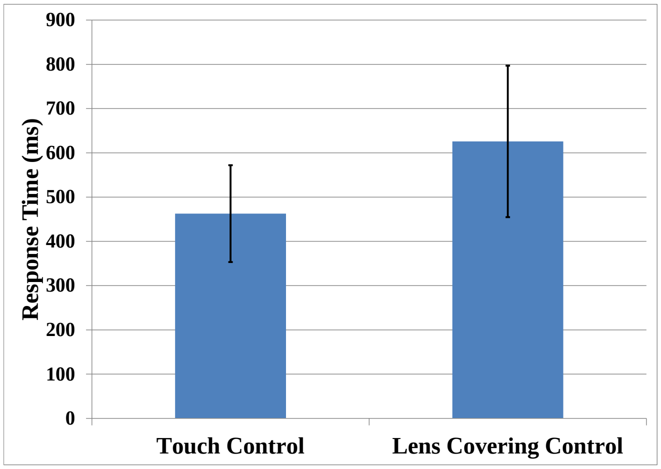


Figure 19. Participants' average response time using the TouchWidget vs. LensGesture to control video play

I quantitatively studied the responsiveness of the AttentiveLearner tangible video control channel and a traditional touchscreen widget. I ran an 18-participant (7 females) study to measure the response time of both interfaces. In the experiment, in response to a randomly appearing visual stimulus (a 200dp x 200dp "Play" icon in the center of the screen), participants were instructed to play the video by either touching the on-screen “Play” button (traditional, widget interface) or by covering the camera lens (the AttentiveLearner tangible video control channel) as fast as possible. Each participant completed 20 trials in each condition, and the order of conditions was counterbalanced. The participants could choose their preferred hands to complete all the tasks. I used a Nexus 5 smartphone with a 5 inch, 1080 x 1920 pixel display for the experiment.

I collected 360 successful inputs from the traditional interface and 359 successful inputs from the tangible video control interface (the only invalid input happened when the subject covered the lens before the “Play” icon appeared). Figure 19 shows the results. The average response time of the traditional interface was 462.6ms (σ = 109.3); the average response time of the tangible video control interface was 625.9ms (σ = 171.1). Although the tangible interface in AttentiveLearner was 160ms slower, it is acceptable for interactive tasks such as playing and pausing a video. I attribute the current delay to two reasons. First, there is a 30ms latency caused by the 30fps camera frame sampling rate and follow-up image processing. Second, the new tangible video control channel was less familiar to the participants. It is expected that the users' response time will decrease when high frame rate cameras become popular, and the learners have more practice with the tangible interface.

3.3.2 Comfort

One major usability concern is whether it is comfortable and natural to cover and hold the camera lens during video watching. To address this issue, I reviewed representative smartphones on the market. I found that the touchscreens of most phones were 4 to 5.7 inches, and the back cameras were located in the upper region of the device. When holding the mobile phone in landscape mode, index or middle fingers of the same hand can naturally reach and cover the lens under normal grip in both one-handed and two-handed holding postures. Figure 20 shows various lens-covering postures I observed during the MOOC learning study presented in Chapter 5. Despite the various fingers and postures used, my algorithm can detect lens-covering actions accurately. The only problematic posture I found was to cover the lens with a whole palm. In this situation, the collected heart rate signals were weak and unreliable. I also collected the users' subjective preference on the tangible video control channel in an 18-subject MOOC learning study to be detailed in Chapter 5. Results showed that participants could comfortably use AttentiveLearner to watch lecture videos without pausing for at least 8 minutes. Eight minutes is longer than the recommended 6-min duration of MOOC videos [51]. For longer video clips, the learners can pause the video by uncovering the lens at any time.

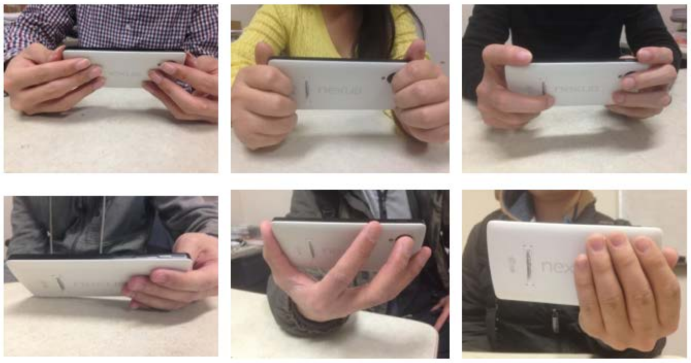


Figure 20. Users cover the lens using various hand postures (back camera is in the top right corner).

3.3.3 Lens Damage

I consulted design experts in leading mobile phone manufacturers to see if covering and swiping directly on the surface of the lens scratch or damage the lens. According to them, mainstream optical assemblies in the mobile phones have been carefully designed to avoid damages from accidental drops, scratches, and collisions. The external unit in an optical unit is usually made of crystal glass, cyclic olefin copolymer, or sapphire. While they are not scratch free, these materials are strong enough to resist frictions caused by finger touch. Interestingly, the surrounding bezel of the camera is usually made of a different material, slightly higher than the external lens surface. Such material and height difference provide excellent tactile feedback for both locating the lens and performing different LensGestures (especially partial occlusion gestures and dynamic gestures).

3.3.4 Battery Life

I ran three mini-experiments to test the impact of AttentiveLearner on battery life. I used a Nexus 5 smartphone running Android 5.0.1 for the mini-experiments. I compared the battery life with both the built-in video player in Android and AttentiveLearner. I tested battery life with 50% backlight of the screen.

As shown in Table 2, AttentiveLearner can run 2.5 hours after a full charge, which is a 60% playtime reduction when compared with the built-in video player. The playtime can be significantly improved considering that 1) Nexus 5 has a below average battery life when compared with existing smartphones on the market; 2) Hardware-accelerated video decoding was used by the built-in video player but not AttentiveLearner in the experiment. Implementing hardware-accelerated decoding could significantly improve the battery life of AttentiveLearner. Last but not least, considering that the average time spent in lecture videos is 2 to 3 hours per week for devoted certificate earners [108] and people usually charge their smartphones daily, I believe that the 2.5-hour battery life is enough to support most learners in MOOCs or flipped courses.

Table 2. Battery life in video playback

Condition	**Duration**
Built-in video player	6 hours 19 minutes
AttentiveLearner, flashlight off	3 hours 57 minutes
AttentiveLearner, flashlight on	2 hours 31 minutes

4.0 LIVEPULSE AND THE IMPLICIT PPG SENSING MODULE

AttentiveLearner captures each learner’s physiological signals implicitly during learning via commodity-camera-based photoplethysmography (PPG). The underlying theory is: in every cardiac circle, the heart pumps blood to the capillary vessels of a human body, including fingertips. The arrival of fresh blood changes the transparency of fingertips. Such transparency changes correlate directly with heart rates, and can be detected by the built-in camera when the user covers the camera lens with her finger tip.

This chapter presents an algorithm I have developed, LivePulse, which analyzes changes in the transparency of learners’ fingertip to get real-time PPG signals and heart rate while learners cover the back camera lens (to watch the MOOC videos). I also present an 18-participant user study which demonstrates the feasibility of implicit physiological signal sensing via AttentiveLearner during actual MOOC learning sessions. The contents of this chapter can be found in the published papers [53] and [130]. Implementation of LivePulse algorithm can be downloaded from http://mips.lrdc.pitt.edu/bayesheart.

4.1 LIVEPULSE

LivePulse is essentially a camera based heart rate detection algorithm. Although camera based heart rate detection algorithms have been reported previously [9, 93], the previous algorithms were not optimized for extracting heart rates from noisy and intermittent covering actions and both algorithms require a long bootstrap time before generating the first estimate (20 seconds in [9], 9 seconds in [93]) in clean and continual measurements. For commercial products, such as Instant Heart Rate [56], and Cardiograph [18], their algorithms were never disclosed, and neither app could extract heart rates if there are finger movements during the measurement. To achieve better robustness and shorter bootstrap time for noisy, intermittent signals from implicit user interactions, I designed my own algorithm that could meet the speed, accuracy and robustness requirement of this project.

4.1.1 Algorithm Design

Instead of using component analysis and then transforming signals to the frequency domain [8, 93, 97], the LivePulse algorithm extracts PPG signals and heart rate information directly from the relatively noisy temporal signal. I made this decision because I expect the LivePulse algorithm could run efficiently on mobile devices in real-time and leave enough CPU power to handle the media player. Matrix factorization operations and frequency domain transformations are still expensive on mobile devices even with support from mobile GPUs.

In the LivePulse algorithm, the camera is set in preview mode, capturing 144x176 pixel color images at a rate of 30 frames per second. I disable the automatic focus function and the automatic white balance function to avoid interference with the algorithm. The built-in flashlight is turned on to improve performance in low illumination conditions.

When a lens covering action is detected (*Static LensGesture* detection algorithm), the LivePulse algorithm extracts heart rates via a 6-step, heuristic based process detailed below.

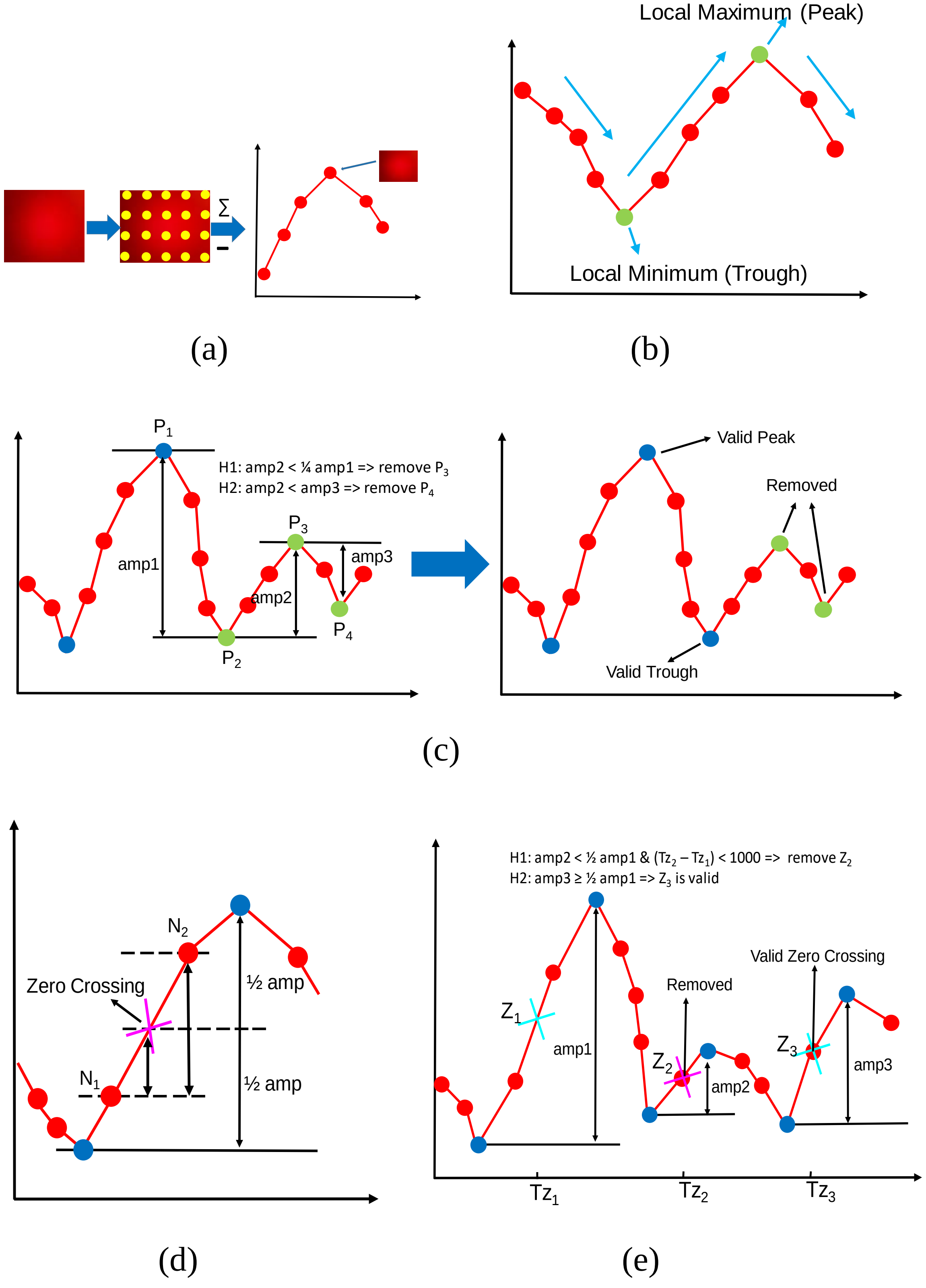


Figure 21. Major steps of the LivePulse algorithm

(a. Signal preprocessing; b. Locating raw local peaks & valleys; c. Locating valid peaks & valleys; d. Locating raw zero-crossings; e. Locating valid zero-crossings).

Step 1: The LivePulse algorithm first converts each image frame captured by the built-in back camera into one time-stamped heart beat sample point via equal distance sampling and the summation of 800 pixels on the Y (luminance) channel in each frame (Figure 21.a). The signal is flipped in y-axis to match waveform output tradition of existing PPG techniques. In this way, I derive a set of time-stamped signal vectors. This step generates the PPG signals, from which I extract heart rate information.

Step 2: The algorithm then detects all the local peaks (local maximum points) and valleys (local minimum points) in the converted temporal sequence signal by measuring the local curvature changes (Figure 21.b).

Step 3: Adjustable threshold based heuristics are applied to eliminate small and noisy local minima/maxima points (Figure 21.c). After the algorithm locates a new peak or valley (e.g., P4 in Figure 21.c), it checks the two most recent peaks (P1, P3) and valleys (P2, P4) to decide if the newest peak and valley (P3, P4) are indeed valid. For P3, the algorithm compares the amplitude of P3-to-P2 (*amp2*) with the amplitude of P1-to-P2 (*amp1*). If P3 has a small amplitude (less than a quarter of P1), P3 is considered as a spurious peak and removed from the peak list. If P3 is an invalid peak, the algorithm further checks which one of P2 and P4 is a valid valley for P1. The algorithm compares the amplitude of P2-to-P3 (*amp2*) and P4-to-P3 (*amp3*) to see which one of them has a lower value. The one with a lower value is marked as a valid valley for P1. In Figure 21.c, P2 is lower than P4, so P2 is the valid valley. Based on offline benchmarking using the data collected in the study in Section 4.1.2, this step can remove up to 70% invalid peaks while keeping all valid peaks in the PPG signal.

Step 4: Instead of estimating heart rates from two adjacent peaks/valleys directly, I found that it is more reliable to estimate and interpolate the time stamps of zero-crossing points between peaks and valleys and then derive heart rates from the time differences between two adjacent time stamps of zero-crossing points. The zero crossing line is dynamic, defined as the line with the mean Y values between two adjacent valley and peak (Figure 21.d). The time stamps of a zero crossing point (i.e., the zero crossing point in Figure 21.d) are linear interpolated values between two adjacent time stamps.

Step 5: Similar to Step 3, the algorithm applies adjustable threshold based heuristics to eliminate invalid, spurious zero crossing points (Figure 21.e). In the algorithm, a valid zero crossing should satisfy either of the following two conditions: 1) Amplitude of the corresponding peak must be greater than half of the amplitude of the previous peak; 2) The interval between the new zero crossing and the previous zero crossing must be greater than 1000ms. For example, in Figure 21.e, since *amp2* is less than half of *amp1* and the interval between Z1 and Z2 is less than 1000ms, Z2 is an invalid zero crossing. On the other hand, Z3 is a valid zero crossing because *amp3* is greater than half of *amp1*.

Through Step 2 to 5, the LivePulse algorithm identifies each heart beat in the PPG signals. The interval between two consecutive (valid) zero crossing points corresponds to one heart beat cycle. This interval is often referred to as RR or NN interval [3]. Using this formula: $HR = 60000(ms) / RR(ms),$ I can get an instant heart rate estimate for each heartbeat.

Step 6: However, instant heart rates are unstable and could be noisy due to finger movement. To further improve the robustness of LivePulse, I save and sort the time stamp distances (RR-intervals) between adjacent zero-crossing points during the past five seconds. Real time heart rate is reported when at least 80% of the time-stamp distances are within a given variation range (200ms) (Figure 22).

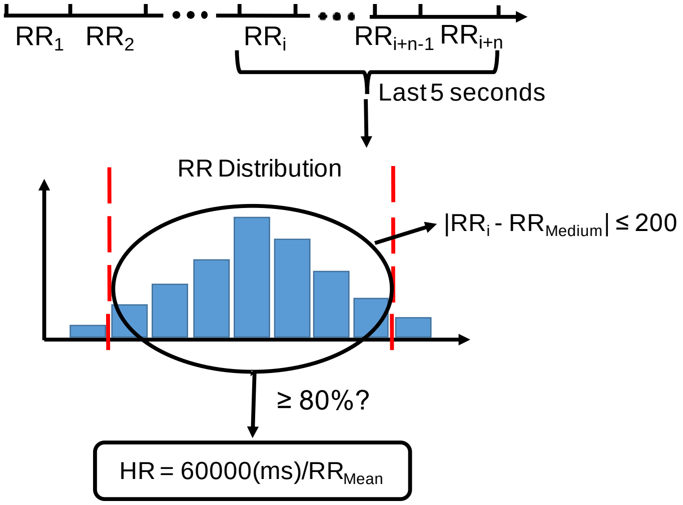


Figure 22. Calculating heart rate**.**

4.1.2 Evaluation

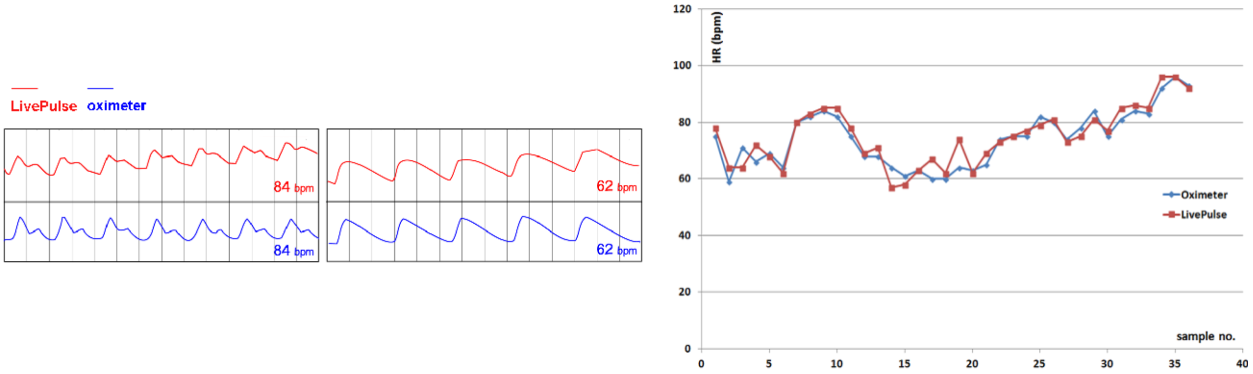


Figure 23. Comparison between LivePulse and pulse oximeter.

Left: 14 seconds of sample heart beat signals from LivePulse (red) and Oximeter (blue); Right: Heart rates (bpm) of 12 subjects from LivePulse (red) and Oximeter (blue).

I performed a formal user study to quantify the accuracy of LivePulse. 12 participants (3 female) between 19 and 27 years of age participated in my study. All the participants were undergraduate or graduate students at the University of Pittsburgh. One thing to notice is that the goal of this lab study is to explore the feasibility of using LivePulse to estimate heart rate. I am interested in knowing whether the performance of LivePulse is good-enough to meet the requirement of AttentiveLearner. The number of participants in this study is also on par with many previous studies on similar topics, e.g., Nacke et al. [84] (10 Subjects), Nenonen et al. [85] (8 subjects), and Row et al. [105] (13 Subjects). Clinical studies with more participants would be conducted if I were to evaluate the efficacy of LivePulse for commercial or medical use.

The whole experiments took about 10 minutes. The experiments were completed on a Google Galaxy Nexus Smartphone with a 4.65 inch, 720 x 1280 pixels display, 1.2 GHZ dual core ARM Cortex-a9 processor, running Android 4.1. It has a 5 mega-pixel back camera and a LED flash. I also used a CMS 50D pulse oximeter with a USB interface to measure heart rate. This pulse oximeter is an FDA approved, medical grade device. The accuracy of CMS 50D for pulse ratio is +/-2BPM. I measured participants' heart rates in resting condition using both the LivePulse application running on a mobile phone and the pulse oximeter for two minutes. Participants sat comfortably in a chair, holding the phone with their left hands and using the index finger of the left hand to cover the camera lens. I attached the pulse oximeter on the index fingers of the participants' right hands.

Overall, raw PPG signals from LivePulse and the oximeter were highly consistent in both beat-to-beat interval and the actual wave shape (Figure 23, left). To compare the accuracy of LivePulse quantitatively, I aligned and re-sampled readings from both LivePulse and the oximeter to 20 HZ. When treating the readings from the oximeter as the gold standard, the Mean Error Rate (MER) of LivePulse was 3.9% (-7 ~ +5 bpm, Figure 23, right). Analysis of variance results showed that the differences in heart rate readings from LivePulse and the oximeter were not statistically significant (F(1, 11) = 1.64, p = 0.13). Pearson product-moment correlation coefficient and there was a significant positive correlation between readings from LivePulse and oximeter, $r = 0.98$, $t = 15.5$, $p < 10^{-7}$.

Despite the relatively low MER, the maximal absolute difference could be up to 7 bpm between LivePulse and pulse oximeter. I attribute these outliers to noise caused by finger position/posture changes and background illumination changes. These outliers can be further eliminated by applying a low pass filtering algorithm or heuristics [8, 97] in situations where accuracy is more important than responsiveness.

4.1.3 Application

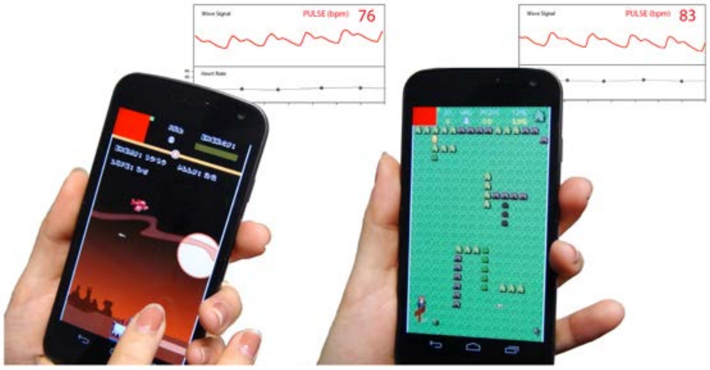


Figure 24. Real-time heart rate measurement via LivePulse Games (left: City Defender, right: Gold Miner).

LivePulse based PPG sensing and heart rate measurement on mobile devices opens opportunities for collecting, interpreting and using physiological signals on smartphones. When integrated with LensGesture, LivePulse can extract heart rate from everyday mobile interactions implicitly, and this provides interesting research opportunities for healthcare, personal well-being [19], and adaptive learning in the future. For example, mobile intelligent tutoring systems could adapt question difficulty based on learners' stress levels inferred from heart rates measured by LivePulse.

One specific application of LivePulse I have explored along with some of my colleagues is LivePulse Games (LPG, Figure 24) [53]. LPG integrate users' camera lens covering actions as essential parts of the game so as to detect heart rate implicitly during game play. LPG make the heart rate monitoring task more engaging and have the potential to measure heart rate longitudinally in a natural and enjoyable way. LivePulse Games are open source software released under BSD license. The implementation can be downloaded from http://mips.lrdc.pitt.edu/livepulsegames.

4.2 IMPLICIT PPG SENSING IN ATTENTIVELEARNER

AttentiveLearner is an application of the combination of LensGesture and LivePulse used in educational contexts. While a learner watches MOOC videos with the tangible video control, her real-time PPG signals are implicitly captured using the LivePulse algorithm. I use this commodity camera based PPG sensing, instead of a dedicated heart rate monitor, to capture PPG signals for two reasons: 1) it works directly on unmodified smartphones and requires no extra devices, and 2) forcing learners to cover the camera lens to watch the videos could potentially make them pay more attention to the lecture.

4.3 USABILITY OF ATTENTIVELEARNER

Even though I have systematically investigated the usability of the tangible video control regarding its accuracy, speed and battery life, it is still unknown whether this interface is comfortable to use and provides valuable physiological feedback in MOOC contexts. Therefore, I conducted a lab-based study to evaluate the usability of AttentiveLearner during actual MOOC learning sessions.

4.3.1 Experiment Design

The whole study took about one hour, and each participant was compensated with a $10 gift card after watching all MOOC videos. The study consisted of three parts:

Overview. I first gave participants a brief introduction to the AttentiveLearner project and then collected background information. I demonstrated the AttentiveLearner mobile app to the participants and answered their questions.

MOOC Learning. Participants studied the introductory chapter of a MOOC course (*Game Theory*) with AttentiveLearner. The course was offered by Stanford and was available on Coursera (https://www.coursera.org/learn/game-theory-1). The topic was selected because participants were unlikely to have prior knowledge of it, while it was "representative" as a real-world STEM learning topic. The chapter ("*Informal Analysis and Definitions*") had four video lectures named "*Introduction to Game Theory and the Predator Prey Example*", "*Normal Form Definitions*", "*Dominance*", and "*Nash Equilibrium*". The durations of the four lectures were *14m47s*, *16m54s*, *8m48s*, and *8m24s* respectively. The duration of the whole chapter was

48m53s. I intentionally selected longer videos so that I could study whether using the tangible video control to consume lecture videos could lead to fatigue when the video was long.

For each of the four video lecture clips, participants first watched the video using AttentiveLearner in landscape mode. Participants could pause the video at any time. Immediately after finishing each video lecture, participants were instructed to rate the interest levels and confusion levels of each topic in the lecture on a 5-point Likert scale. There were 7, 9, 5, and 7 topics in the four video lectures respectively. The duration for each topic ranges between 33s to 2m46s (average: 1m41s). Participants could take a short break between two video lectures. The self-reported ratings on learning topics were used as the ground truth for cognitive state inference, which will be detailed in the next chapter.

Qualitative Feedback. Each participant completed a closing questionnaire (B.2) after finishing the lesson.

4.3.2 Participants and Apparatus

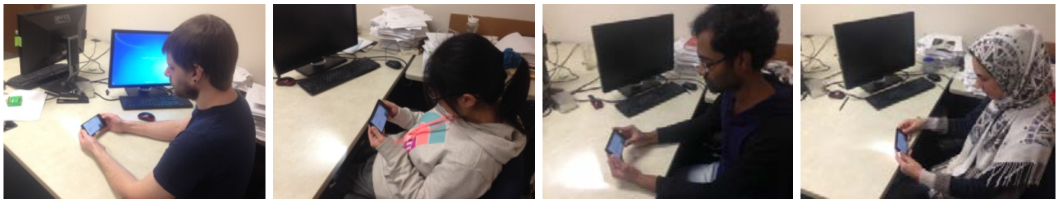


Figure 25. Some participants in my experiment using AttentiveLearner to learn video lectures.

Eighteen subjects (7 females) participated in my study (Figure 25). I decided to recruit 18 subjects because this number is comparable to similar studies on usability evaluation of learning technologies, e.g., Monserrat et al. [81] (18 subjects), Kovacs [69] (18 subjects), Monserrat et al. [82] (15 subjects), Kim et al. [65] (12 subjects). The average participant age was 24.9 (σ = 2.2) ranging from 22 to 30. All participants were undergraduate or graduate students at the University of Pittsburgh. All participants had little or no knowledge of Game Theory. Among the 18 participants, 8 took MOOC courses before the study. Only 2 subjects actually finished a MOOC course, suggesting a low completion rate in MOOC learning. Three subjects had experiences in using mobile MOOC learning apps.

My experiment was completed on a Nexus 5 smartphone with a 4.95 inch, 1920 x 1080 pixel display, 2.26 GHz quad-core Krait 400 processor, running Android 5.0.1. It has an 8 mega-pixel back camera with an LED flash.

4.3.3 Results

4.3.3.1 Subjective Feedback

Participants reported favorable experiences with AttentiveLearner (Figure 26), giving an average rating of 4.11 (σ = 0.68) on the overall experience of AttentiveLearner on a five-point Likert scale (1-very unsatisfied, 5-very satisfied).

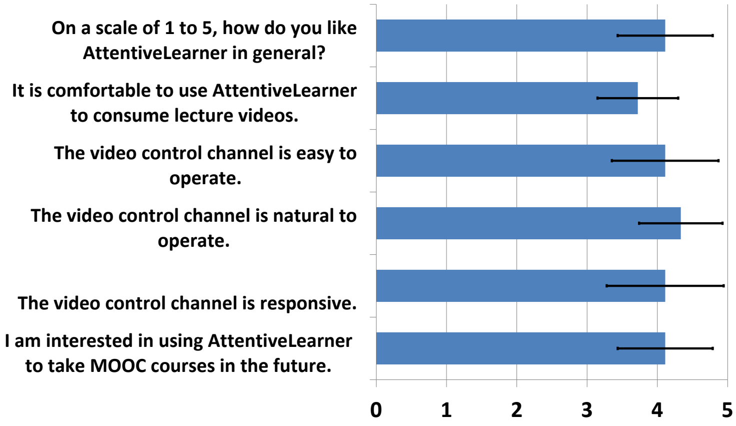


Figure 26. Subjective ratings of AttentiveLearner

Regarding the tangible video control interface, participants gave an average rating of 4.33 (σ = 0.59) on intuitiveness and an average rating of 4.11 (σ = 0.83) on responsiveness. All participants agreed that it was comfortable to cover-and-hold the lens while watching the video.

Fifteen subjects commented that they would continue to use AttentiveLearner to take MOOC courses in the future. When asked about what they like about AttentiveLearner, participants were most impressed by the flexibility of the video control channel:

“The lens-covering control is interesting. It is an easy and intuitive way to play/pause video.”

“I like the auto-pause feature. You just put the device aside and it automatically stops.”

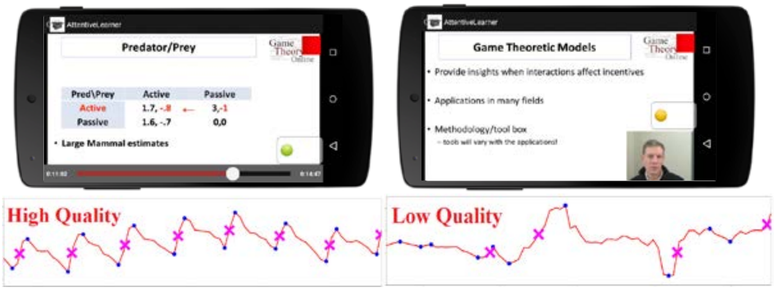
“A user can play/pause the video easily with one hand. No need to touch the screen which normally needs two hands.”

Some participants also believed the video control channel made them pay more attention to the video:

“It can help me focus on the lesson; you need to hold the mobile phone while listening to the instructors.”

“Because I’d like to keep covering the lens, I am paying attention to the video all the time.”

4.3.3.2 PPG Signals



Predator/Prey

Pred\Prey	Active	Passive
Active	1.7, -.8 ←	3,-1
Passive	1.6, -.7	0,0

- Large Mammal estimates

Game Theoretic Models

- Provide insights when interactions affect incentives
- Applications in many fields
- Methodology/tool box
  - tools will vary with the applications!

Figure 27. Top: Screenshots of AttentiveLearner in the experiment. Bottom: PPG signal at the time of the screenshot. Left: high-quality PPG, right: low-quality PPG.

Interruptions: Although I encouraged participants to pause the video when needed, I only observed a total of 17 user-initiated pauses (i.e., interruptions) from 7 subjects. The main reason for interruptions was finger or eye fatigue. However, two participants reported that they paused the video to take a closer look at the slides and digest the topic.

Interestingly, I found that 14 of the interruptions (82.3%) occurred after 8 minutes of video play; 12 of them (70.5%) occurred after 10 minutes of video play. This suggests that participants usually felt fatigue and needed a rest when they used AttentiveLearner to watch a video nonstop for more than 8 minutes.

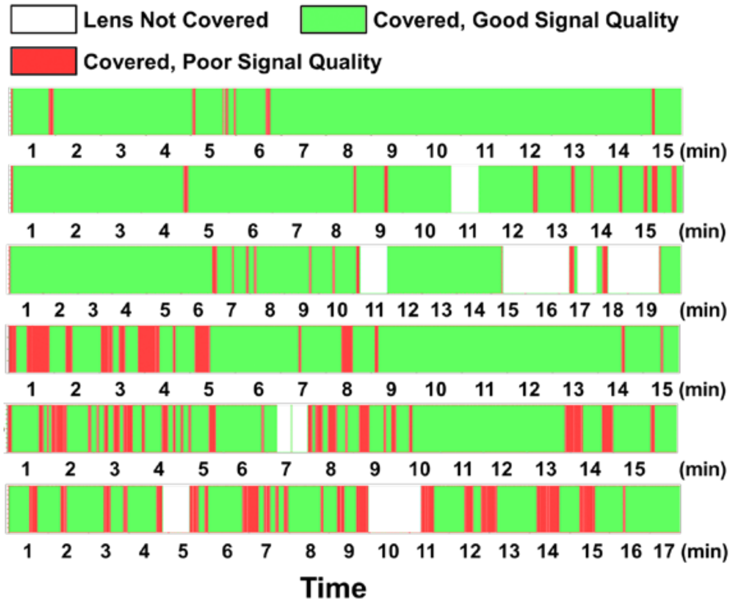


Figure 28. PPG signals of six participants while watching the first video clip.

Signal Quality: I analyzed the quality of PPG signals by investigating the RR-intervals (the cardiac interval between two heart beats) in a 5-second moving window. I used the heuristics that a window contains high quality PPG signal if at least 80% of RR-intervals in that window are within a given range (+/- 25% from the median). Figure 27 shows two sample sequences of PPG signals (left: high-quality; right: low-quality) and the corresponding screenshot of

AttentiveLearner. In 88.9% of the 72 video sessions (18 subjects x 4 videos), more than 80% of the signals were in high quality. This suggests that AttentiveLearner can collect high-quality PPG signals reliably from learners’ fingertips during video watching.

Figure 28 shows an illustration of the PPG signal quality collected in six learning sessions. The white areas were interruptions in the signal (video pauses). The green areas were high-quality PPG signals, and the red areas were low-quality signals. The percentage of high-quality signals for the six video sessions in Figure 28 were 97.74%, 95.70%, 96.76%, 84.71%, 81.22%, and 74.90% respectively from top to bottom. I observed that the low-quality signals were scattered across the whole video session and that each low-quality signal sequence usually had short durations (less than 30 seconds). Therefore, I can still extract high-quality PPG signals from major portions of the learning sessions even if the video session contained low-quality signals (sessions 4 – 6 in Figure 28).

4.3.4 Summary

This usability study showed the feasibility of using AttentiveLearner in extended MOOC learning sessions. Learners found the tangible video control interface in AttentiveLearner to be intuitive to learn, and accurate and responsive to use. The PPG signals collected via commodity camera phones were also reliable. Finger fatigue may happen after extended usage of AttentiveLearner. To avoid finger fatigue, the durations of lecture videos in AttentiveLearner should be no more than 8-10 minutes based on qualitative feedback from the closing questionnaire.

5.0 IMPLICIT COGNITIVE STATE INFERENCE: HEART RATE AS FINE-GRAINED FEEDBACK FOR MOOC LEARNING

In this section, I demonstrate the feasibility of using the PPG signals and heart rate implicitly captured by AttentiveLearner to predict learners' cognitive states. Specifically, I am interested in two cognitive states: boredom and confusion in MOOC learning. The contents of this chapter can be found in the published papers [130] and [134].

5.1 BACKGROUND

A user's changing heart rhythms affect not only the heart itself, but also the brain's ability to process information and manage emotion. Researchers have discovered that both the heart rate [120] and heart rate variability (HRV) [110] have strong correlation with user's physiological state, including cognitive workload and mental stress level, in contexts such as computer user interfaces [105], traffic control [105], longitudinal monitoring of emotion and food intake [19], and intelligent tutoring [55, 59]. Previous studies have already used heart rate and HRV to infer cognitive workload [75, 105, 127], mental stress [74, 114], attention [59, 102], or emotions [55]. In this section, I investigate the feasibility of using PPG signals and heart rate collected by AttentiveLearner to predict learners' cognitive states.

5.2 DATA COLLECTION AND FEATURE EXTRACTION

In Chapter 4.3, I presented an 18-participant user study to evaluate the usability of AttentiveLearner during actual MOOC learning sessions. In the same study, AttentiveLearner collected participants’ PPG signals throughout the whole learning session. In the study, participants were also instructed to rate the perceived interest levels and confusion levels of each topic in the videos on a 5-point Likert scale immediately after they watched each video. I used participants’ self-reported ratings on learning topics as the gold standard. A total of 522 user ratings (29 topics * 18 subjects) were collected. I excluded video sessions with the same rating for all topics, implying that the participant reported the same feeling throughout the session. The whole dataset contained 428 samples of interest/boredom predictions (23.83% of the topics were rated boring/uninteresting, rating $\leqslant$ 2) and 490 samples of confusion predictions (19.8% were rated confusing, rating $\geqslant$ 4).

For each video session, I used LivePulse to extract RR-intervals and heart rates from the corresponding PPG signals. I applied the following heuristics to eliminate outliers in RR-intervals:

- Discard RR-intervals corresponding to heart rates beyond 40 ~ 140 bpm;
- Discard RR-intervals corresponding to heart rates differ more than 20 bpm from the median over the video session;
- Discard RR-intervals corresponding to heart rates differ more than 10 bpm from the previous RR-interval.

I extracted 14 dimensions of features from raw PPG signals of each learning topic. The durations of a topic ranged between 33s to 2m46s (average: 1m41s). Among these features, 7 dimensions were global features extracted from the PPG signals of the entire topic section. The 7 global features were these: 1) Mean-HR; 2) SD-HR; 3) AVNN; 4) SDNN; 5) pNN50; 6) rMSSD; 7) MAD. These features were common short-term time domain heart rate and HRV features [106, 117] and Table 3 lists the descriptions for these features. The other 7 dimensions (e.g., Local Mean-HR) were local features extracted by averaging the same features in multiple fix-sized, non-overlapping local windows within the topic section (Figure 29). If the last local window overlapped with the beginning of the next topic, only signals that had fallen within the current topic were used. I also normalized the features in each video session for each participant.

Table 3. Descriptions of the heart rate and HRV features extracted in the study

Feature Name	**Explanation**
Mean-HR	Average of the heart rate
SD-HR	Standard deviation of the heart rate
AVNN	Average of the RR-intervals (or NN-intervals)
SDNN	Standard deviation of the RR-intervals
pNN50	Percentage of adjacent RR-intervals with a minimum difference of 50 ms in the corresponding time frame
rMSSD	Square root of the mean of the squares of difference between adjacent RR-intervals
MAD	Median absolute deviation of all RR-intervals

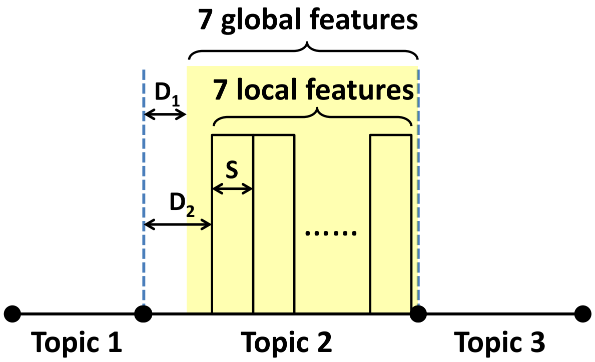


Figure 29. Feature Extraction from the PPG signal in each section of a video session.

5.3 RESULTS

I explored several supervised machine learning algorithms to predict learners’ interest/boredom and confusion states using the extracted features. The algorithms I tested were k-nearest neighbors (*kNN*), Naïve Bayes (*NB*), Decision Tree (*DT*), support vector machine with linear kernel (*LinearSVM*), and support vector machine with radial basis function kernel (*RBFSVM*). I used WEKA to train and optimize the classifiers. I explored these algorithms because they were commonly used machine learning algorithms which can be applied to almost any data problems in the absence of prior knowledge about the data and domain.

I used the leave-one-subject-out method to evaluate the performance of the models. Therefore, all results reported were user-independent. I calculated and reported Cohen’s Kappa because the distributions of class labels were skewed (23.8% and 19.8% topics were rated boring and confusing) so reporting accuracies alone would be insufficient. The optimal parameters of a classifier were chosen according to the best average Kappa over all subjects. To get the optimal performance, I tried 5 different delays D1 for extracting the global features (0s, 5s, 10s, 15s, 20s) × 5 different delays D2 for the first window (0s, 5s, 10s, 15s, 20s) × 12 window sizes S (10s, 15s, 20s … 60s) for extracting local features (Figure 29).

Table 4 lists the best performance (in Kappa) achieved by each classifier. The RBF-kernel SVM has best overall Kappa (0.297 and 0.269) for predicting both boring and confusing topics. The reason why RBF-SVM has the best performance might be that the relationship between the extracted features and learners’ cognitive state is non-linear. And compared to other models, the RBF-kernel can map the data to a much larger dimensional space; thus, RBF-kernel SVM is more likely to find the optimal solutions.

Table 4. The performance of the classifiers for boredom prediction and confusion prediction.

Prediction of the boring topics in a video session				
Model	**Accuracy**	**Precision**	**Recall**	**Kappa**
kNN	77.29%	**0.504**	0.325	0.258
NB	69.37%	0.376	0.373	0.162
DT	**78.56%**	0.523	0.180	0.191
LinearSVM	67.71%	0.397	**0.538**	0.237
RBFSVM	73.58%	0.462	0.499	**0.297**
Prediction of the confusing topics in a video session				
kNN	77.17%	0.396	0.316	0.211
NB	77.99%	0.358	0.164	0.116
DT	**81.96%**	0.523	0.208	0.224
LinearSVM	75.74%	0.402	**0.366**	0.223
RBFSVM	77.69%	**0.516**	0.353	**0.269**

The performance of my classifiers is comparable to existing systems that rely on dedicated physiological sensors to detect human affect (e.g., Hussain et al. [55] developed *user dependent* models with Kappa scores of 0.35 and 0.23 for detecting three-level valence and arousal; D'Mello et al. [32] extracted features from dialog, posture, and face to classify four different affect, and their *user dependent* models achieved the best Kappa score of 0.29). It is worth highlighting that the performance is achieved on today's mobile phones without any hardware modifications. The Kappa scores indicate that AttentiveLearner is capable of identifying the perceived boring and confusing parts of a video in a user-independent fashion.

One thing to notice is that the focus of this research is to explore the feasibility of detecting learners' cognitive states during mobile MOOC learning via a "sensorless" approach. Currently, the performance is still far from perfect. The relatively low Kappa/accuracy might be because of the following reasons: 1) Noises in the signal; 2) Heart rate and PPG signals might be affected by other confounding effects, such as physical movement or fatigue; 3) There exists a large variance among people. The diverse heart rate, PPG signal patterns, as well as the different perception of learning materials among participants may be additional factors that limit the performance of a user-independent model. I believe that the accuracy/Kappa could be improved using more powerful machine learning techniques or building user-dependent classifiers.

I found that the local window size had a significant impact on the classifier performance (Figure 30). A window size of 50 seconds had the best performance for predicting perceived boredom and a smaller window (30 seconds) led to the best performance for predicting confusing topics.

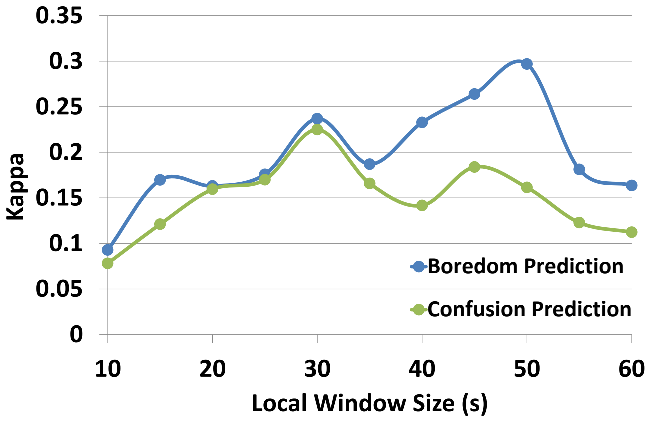


Figure 30. Classifiers' Kappa by the local window size.

5.4 EXTREME EVENTS AND AGGREGATED EVENTS

The previous study shows that it is feasible to detect boring and confusing topics in MOOC videos via PPG signals implicitly captured by AttentiveLearner. Although the Kappa scores indicated a strong correlation between HRV features and cognitive states, the prediction accuracy is still far from perfect. Therefore, I propose two solutions to get reliable and effective learning analytics from imperfect predictions: predicting *extreme personal learning events* and *aggregated learning events*. I define *extreme personal learning events* as a small fraction of events from a learner that are drastically different from other events based on a specific marker. I define *aggregated learning events* as the aggregated responses of all learners towards a learning topic.

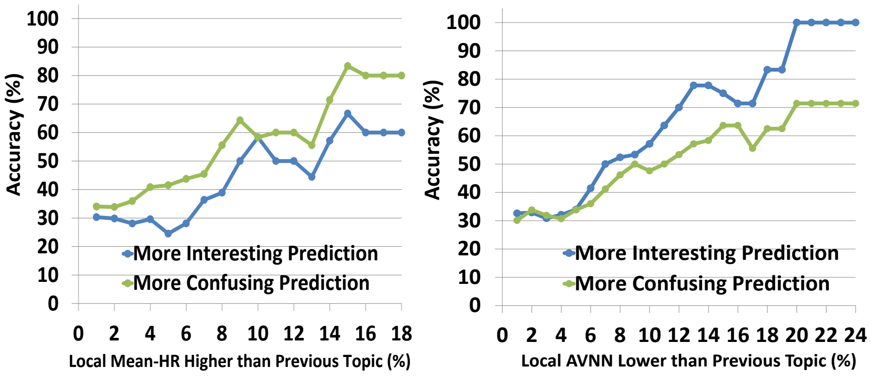


Figure 31. Predication accuracy when using local mean-HR (left) and local AVNN (right) as markers of "extreme personal learning events".

Figure 31 shows the use of the local mean-HR feature (left) and the local AVNN feature (right) as markers of "*extreme personal learning events*" and the corresponding prediction accuracies on relative interest/boredom and confusion. I can predict with 60% accuracy in relative confusion by restricting predictions to events where the local mean-HR was at least 11% higher than the previous topic. I can predict with 83.3% accuracy by restricting predictions to events where the local mean-HR was at least 15% higher than the previous topic (Figure 31 left). Similarly, the local AVNN can also be used as an effective marker for predicting relative interest/boredom and confusion (Figure 31 right). In my experiments, local mean-HR was a better marker for predicting relative confusion (Figure 31 left) and local AVNN was a better marker for predicting relative interest/boredom (Figure 31 right).

I also found that PPG signals from a group of learners could be aggregated for more accurate prediction of cognitive states. Such aggregated events are informative to instructors because they convey the overall feedback from students on a specific learning topic. For example, Figure 32 shows the aggregated histogram of “confusing” topic from both reported results (left) and predicted results (right) for the second lecture (“*Normal Form Definitions*”). It is clear that our aggregated prediction is consistent with participants’ ratings. In both histograms, the 7th topic has more confusion than any other topic, implying that this topic is challenging during learning. After investigating the corresponding lecture video, I found that the 7th topic contains an in-depth analysis of the “Team Games” concept. Such insights captured from learners’ physiological signals can help teachers refine instructional content for the future.

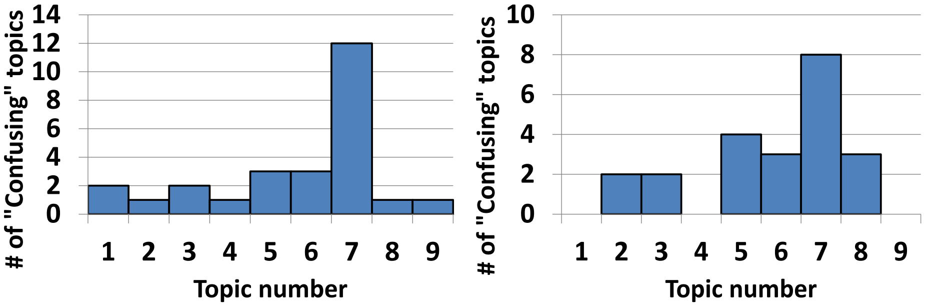
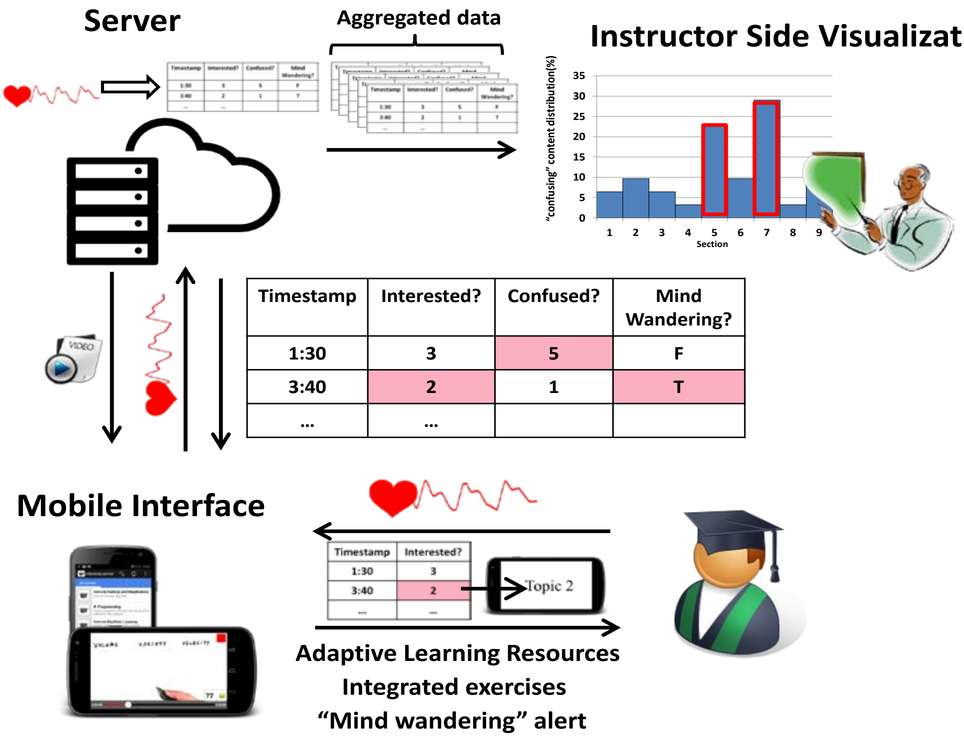


Figure 32. Histograms of “Confusing” topics of video clip 2.

Left: reported results. Right: predicted results.

5.5 INFRASTRUCTURE OF ATTENTIVELEARNER

Given the feasibility to infer learners' cognitive, affective states and attention from the implicitly captured PPG signals; I believe AttentiveLearner can bring many new opportunities to enrich large scale learning analytics and enable attentive and bi-directional learning on unmodified mobile phones. Figure 33 shows a hypothesized infrastructure of AttentiveLearner. Step 1: While a learner uses the AttentiveLearner mobile interface to consume lecture videos, the interface also implicitly captures her PPG signals and sends the information to the server. Step 2: On the server's side, pre-trained classifiers make predictions of the learner's cognitive states using the received PPG signal information. Then, the server sends the cognitive state information back to the mobile interface. Step 3: Knowing the learner's real-time cognitive states, the mobile interface can benefit learners by providing personalized learning materials and instructional paradigms. For example, when a learner is not interested in a topic, AttentiveLearner may switch to a different learning resource or use integrated exercises [48, 81] to engage the learner. AttentiveLearner can also use visual and tactile feedback to remind learners when they are "mind wandering". Such interactions have the potential to make the MOOC learning process more attentive. In Chapter 7, I will present an intervention technique, the Context and Cognitive State triggered Feed-forward, which effectively improve student engagement and efficacy in mobile MOOC learning.



Timestamp	Interested?	Confused?	Mind Wandering?
1:30	3	5	F
3:40	2	1	T
...	...		

Timestamp	Interested?
1:30	3
3:40	2
...	...

Figure 33. The infrastructure of AttentiveLearner.

As shown in Figure 33, AttentiveLearner can also benefit instructors by providing instructor side visualization of the aggregated information of learners’ physiological, cognitive, and affective states synchronized with the learning materials. An instructor can identify and reflect upon areas needing improvement within the curriculum. For example, which parts of a lecture are more confusing to students? Did my joke "wake up" the students? Or, were students bored by the end of the lecture? I believe this fine-grained, continual, implicitly feedback channel through learners' physiological signals can serve as a valuable complement to existing technologies such as log analysis [21, 51], questionnaires, and post-lecture reflections [48]. Such information can help instructors to identify both struggling students and lecture materials that need improvements, hence enabling bi-directional communications between learners and instructors.

6.0 IMPLICIT COGNITIVE STATE INFERENCE: UNDERSTANDING AND DETECTING DIVIDED ATTENTION IN MOBILE MOOC LEARNING

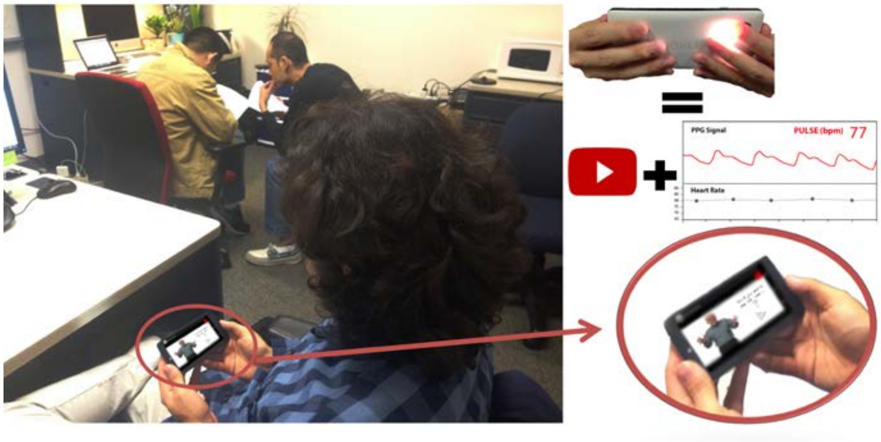


Figure 34. AttentiveLearner detects external distractions (colleagues discuss in the background) during mobile MOOC learning.

Learners tend to face more distractions due to the highly diversified learning environments and highly interruptive learning context when studying alone with one's mobile devices. Distractions could come from both external sources (e.g., background conversations, ambient noises) and multitasking (e.g., checking/updating social networking sites). When learners divide their attention between the learning materials and other tasks or external distractions, the interference hampers their intentional use of memory [57] and reduces the memory performance substantially [25]. Both outcomes hinder the knowledge encoding process and lead to decreased understanding of the learning materials. In this section, I investigate the impact of divided attention (DA) on both the learning process and the learning outcomes in the context of mobile MOOC learning. In this section, I demonstrated the feasibility of detecting whether a learner is dividing her attention as well as the type of DA via implicit physiological signal sensing on unmodified mobile phones (Figure 34). The contents of this chapter can be found in the published paper [132].

6.1 BACKGROUND

By definition, divided attention (DA) occurs when attention is divided among simultaneous stimuli [60]. DA is different from mind wandering (MW) [80] in that DA is either caused by *intentional* multi-tasking (internal distractions) or *passive* external distractions [60], while MW is an *involuntary* shift in attention from task-related thoughts to task-unrelated thoughts [76, 80] and is *stimulus independent* [63]. Existing DA research in learning technology is limited because MW can happen in any environment [63] while DA is more pervasive in informal learning.

Research on DA has been focusing on understanding people's capabilities to perform multiple tasks simultaneously [25, 57, 60]. Kahneman [60] systematically reviewed experiments on the parallel processing of simultaneous inputs and found that although parallel processing was possible, its effectiveness was often impaired due to the interferences among multiple activities. Craik et al. [25] conducted four experiments to explore the effects of DA on encoding and retrieval processes. Experimental results showed that DA during the encoding process was associated with large reductions in memory performance. The divided attention made the selection of information imperfect, resulting in delayed or slowed processes [112].

Although researchers have explored the use of various physiological signals, such as Electroencephalography (EEG) [52, 116], eye gaze [11, 12, 30], galvanic skin responses (GSR) [12, 13, 17, 52, 55, 59], and heart rate [52, 55, 59] to infer learners' attention state or "mind wandering" events in educational settings, little work has been done on detecting divided attention from attention. The work by Rodrigue et al. [104] is perhaps the most relevant research. In two three-participant experiments, Rodrigue and colleagues built user-dependent models (accuracies range from 79% to 99%) to detect DA from signals collected by a consumer-grade eye tracker and an EEG sensor. However, this research mainly focused on reading and the DA detection required dedicated devices for signal collection.

6.2 USER STUDY

I conducted another 18-participant study to investigate the impact of divided attention on both learning outcomes and learners' PPG signals in mobile MOOC contexts. I studied two typical types of distractions: 1) multitasking distractions (i.e., internal divided attention) where the subject's attention is divided between two sets of stimuli; and 2) unpredictable and intrusive auditory distractions (i.e., external divided attention).

6.2.1 Task

The basic task for participants was to watch four lecture videos (8 minutes each) with AttentiveLearner. I used two types of stimuli to create internal divided attention condition (e.g., multitasking) and external divided attention condition (e.g., distractive audio sound) while participants were watching MOOC videos. In the control condition (full attention), participants can focus on the video without any internal or external stimuli.

I adopted the color counting task [104] to introduce internal divided attention. While a participant was watching the lecture video using AttentiveLearner, a computer placed on the side spoke the names of six different colors in a random order at a speed of one-second per word (high divided attention) or five-seconds per word (low divided attention). The volume of the smartphone was set to the highest. The volume of the computer was set to be as low as possible while still allowing the subject to hear the colors clearly. Participants were told to focus on the video but also count the number of times a target color (e.g., the color "red") was spoken during the video. The participants reported the counted number of the target color after each video, which was compared to the ground truth to ensure that they indeed divided their attention between the two tasks.

To simulate an environment where a learner is distracted by external stimuli (external divided attention), such as unexpected and intrusive auditory distractions, the computer placed on the side of the learner played loud and energetic music while the learner was watching the lecture video. I chose to use music as an external stimulus because it is a common environmental sound during informal learning. The volume of the computer was set to high so that the participants were more likely to be distracted while still allowing them to hear the lecture video clearly (tested in a pilot study of three participants).

6.2.2 Participants and Apparatus

Eighteen subjects (6 females) participated in the study (Figure 35). The average age was 25.8 (σ = 2.73) ranging from 22 to 32. All participants were graduate students at the University of

Pittsburgh. None of the participants had previous knowledge of the learning materials used in the study. Four subjects had prior experience of using mobile apps for MOOC learning.

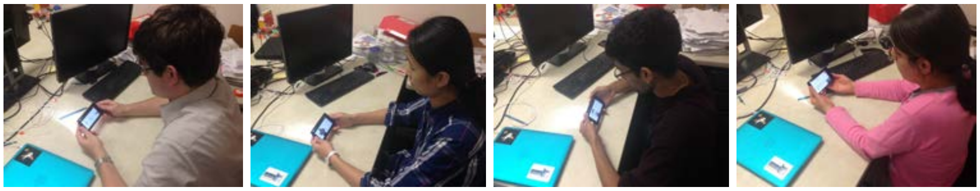


Figure 35. Some participants in the user study. The laptop on the left was used to play stimulus sound.

The experiment was completed on a Nexus 5 smartphone with a 4.95 inch, 1920 x 1080 pixel display. The device had an 8-megapixel back camera and a 2.26 GHz quad-core Krait 400 processor, running Android 5.0.1.

6.2.3 Procedure

The whole study took about one hour, and each participant was compensated with a $10 gift card. The study consisted of three parts:

Introduction: I ran a tutorial session and collected background information from participants.

MOOC Learning Session: Participants watched four video clips in four conditions (i.e., within-subjects design): control condition (i.e., Full Attention, **FA**), low internal divided attention condition (**LIDA**), high internal divided attention (**HIDA**), and external divided attention (**EDA**). The order of condition assigned to each video was counterbalanced by a Latin Square pattern. Before a participant watched each video, I also collected one-minute baseline PPG signals from her.

The four video clips were chosen by three reviewers to make sure all video clips had similar difficulties and durations. Two video clips were from the course "*Intro to Design of Everyday Things*" (https://www.udacity.com/course/intro-to-the-design-of-everyday-things--design101), and the other two were from the course "*Intro to iOS App Development with Swift*" (https://www.udacity.com/course/intro-to-ios-app-development-with-swift--ud585). Both courses were taken from the MOOC learning platform, Udacity. All clips were edited to exactly eight-minutes long.

Post-video Quiz and Self-Rating: After finishing each video clip, participants completed a five-question quiz on the topics covered in the video. The questions tested their ability to recall important information presented in the video. For example, "please explain the concept 'signifier'" and "what are the two ways to center the button horizontally and vertically in the container".

Participants also reported the perceived interestingness level and difficulty level for each video clip, as well as their perceived distraction level for each learning condition on a 7-point Likert scale.

6.3 EFFECT OF DIVIDED ATTENTION ON LEARNING

6.3.1 Subjective Feedback

The perceived distractions were 1.33 ($\sigma = 0.47$), 3.11 ($\sigma = 1.59$), 4.11 ($\sigma = 1.24$), and 4.97 ($\sigma = 1.18$) for the four conditions **FA**, **EDA**, **LIDA**, and **HIDA** respectively. Repeated measures of Analysis of variance showed a significant main effect (F (3, 15) = 9.28, p < 0.0001) of the perceived distractions among the four conditions. Pairwise mean comparison (t-tests) with Bonferroni correction showed that the control condition (**FA**) was rated significantly less distractive than all other conditions (t(17) = 4.74, p < 0.001; t(17) = 8.71, p < 0.001; t(17) = 12.27, p < 0.001). *External Divided Attention (***EDA***)* was significantly less distractive than the *high internal divided attention (***HIDA***)* (t(17) = 4.28, p < 0.001). The difference between the two levels of internal divided attentions, i.e., **HIDA** and **LIDA,** was also significant (t(17) = 3.38, p = 0.0036). The results suggested that both external divided attentions and internal divided attentions had a significant impact on learners' perceived distraction levels in mobile MOOC learning.

6.3.2 Effect of Divided Attention on Learning Performance

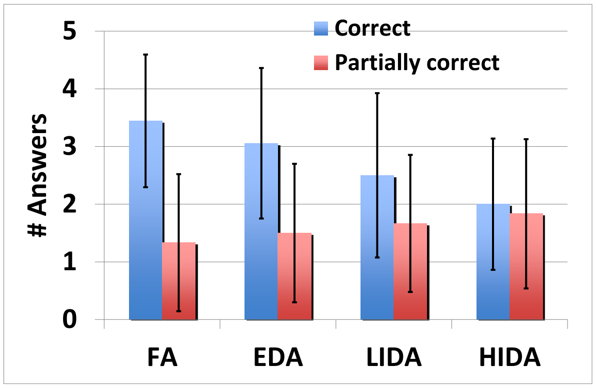


Figure 36. Participants' quiz performance by conditions.

Questions in the post-video quiz were graded by the following rubrics: a participant received 1 point for a complete and accurate answer; the participant received 0.5 points for a reply with a correct general idea but missing critical details; the participant did not get any points if the answer was incorrect or missing. Participants' average scores for the four conditions were 4.11

($\sigma = 0.66$), 3.83 ($\sigma = 0.90$), 3.33 ($\sigma = 0.92$), and 2.92 ($\sigma = 0.98$) respectively. Repeated measures of Analysis of variance showed a significant main effect (F (3, 15) = 1.12, p < 0.01) of the learning outcomes. Pair-wise mean comparison (t-tests) with Bonferroni correction showed that participants performed significantly better in *control condition* (**FA**) than both **LIDA** (t(17) = -3.0, p < 0.01) and **HIDA** (t(17) = -4.13, p < 0.001). Although the learners' average performance in **FA** was better than that in **EDA**, the difference was not significant (t(17) = -1.17, p = 0.256). This suggests that IDA is more detrimental to learning than EDA.

Figure 36 shows the average number of questions the participants answered entirely correct and partially correct. Compared to the **FA** condition, participants had less completely correct answers, but more partially correct answers in the divided attention conditions. This is especially true for the **HIDA** condition, where the number of entirely correct answers were only 58.1% of that in the **FA** condition (2 vs. 3.44), while the number of partially correct answers increased by 37.5% (1.83 vs. 1.33). I found that learners in the divided attention condition were more likely to miss important details. Their answers also showed partial and shallower understanding of the learning materials.

6.4 DIVIDED ATTENTION DETECTION

I explored the use of PPG signals *implicitly* captured during mobile MOOC learning to predict whether a learner has divided attention, as well as its type and intensity. My work is different from Rodrigue et al. [104] in two major ways: 1) My method can run on unmodified smartphones and does not rely on external eye trackers and EEG sensors; 2) My approach focuses on mobile MOOC learning rather than reading on desktop computers. Since the divided attention detection algorithm in [104] was user-dependent, for direct comparison purposes, I begin with user-dependent models, and then build user-independent models to estimate performances in the "*cold-start with no adaptation*" situation.

6.4.1 User-Dependent Classification

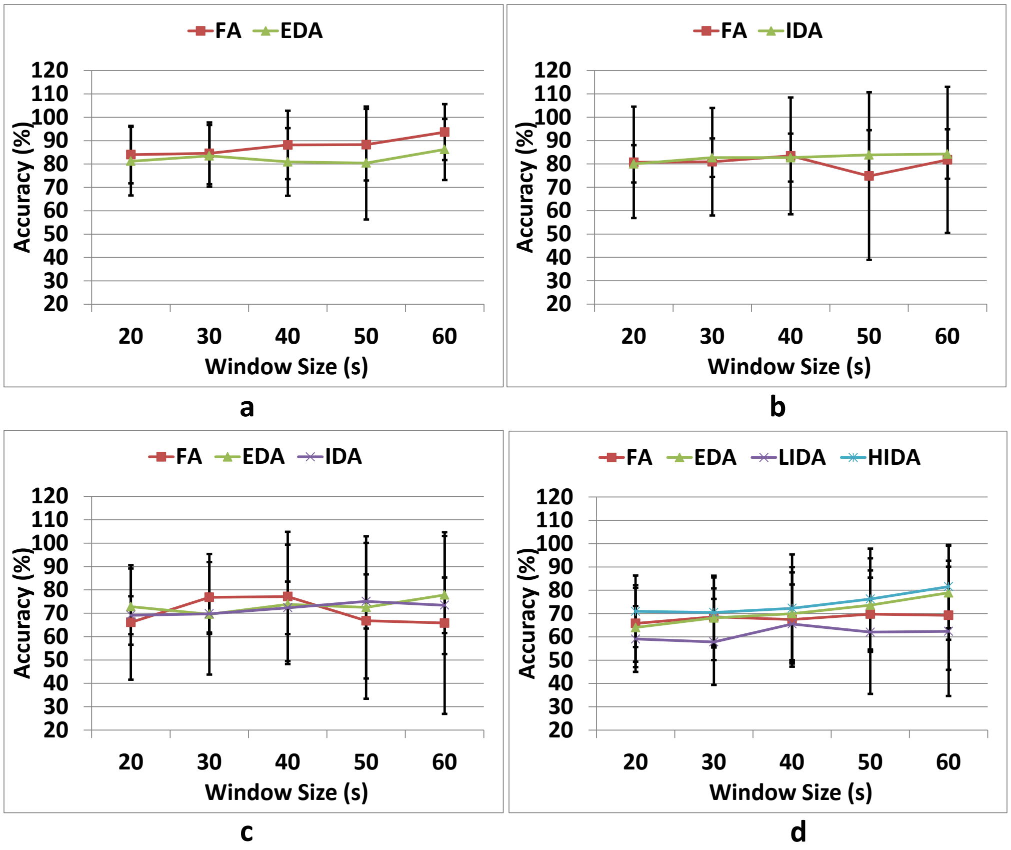


Figure 37. Accuracies for detecting divided attention. a: FA and EDA only; b: FA and IDA; c: FA, EDA, and IDA; d: all four conditions.

The PPG signals collected from each lecture video first went through a second-order Butterworth filter with a cutoff frequency of [0.75, 3.3] Hz. The signals were then segmented into small, non-overlapping consecutive windows to detect divided attention. I explored 5 different window sizes

(20s, 30s ... 60s). I used the LivePulse algorithm to extract RR-intervals and heart rates from each window. I applied the same heuristics to eliminate outliers in RR-intervals as the study presented in Chapter 5. Nine dimensions of heart rate and HRV features were extracted from each window: 1) Mean-HR; 2) SD-HR; 3) AVNN; 4) SDNN; 5) rMSSD; 6-8) pNN12, pNN20, pNN50; and 9) MAD. Definitions of these features were presented in the previous study (Chapter 5.2). All these features were normalized using the corresponding features from the one-minute baseline PPG signals collected before the corresponding video. I used WEKA to train and optimize different classification algorithms. Support vector machine (SVM) with a radial basis function (RBF) kernel yielded the best overall performance.

I built divided attention prediction models for each participant and evaluated them with 10-fold cross validation. Figure 37 shows the average prediction accuracies across the 18 subjects by different window sizes for various classification tasks: 1) detect **EDA** from **FA** (binary classification, Figure 37. a); 2) detect **IDA** from **FA** (binary classification, Figure 37.b); 3) detect **FA**, **EDA** and **IDA** (three-way classification, figure 37.c); 4) detect **FA**, **EDA**, **LIDA** and **HIDA** (four-way classification, figure 37.d).

Overall, the classification accuracies increase with the growth of window sizes. When using 60s windows, the classifiers achieved an accuracy of 88.54% when detecting **EDA** (86.22% accuracy) from the control condition **FA** (93.65% accuracy) (Figure 37.a); 84.49% when detecting **IDA** (84.24% accuracy) from the control condition **FA** (81.77% accuracy) (Figure 37.b); 74.13% when detecting the **FA** (65.80% accuracy), **EDA** (77.81% accuracy), and **IDA** (73.43% accuracy) (Figure 37.c); and 72.74% when detecting all four conditions (69.26%, 78.91%, 62.37% and 81.57% accuracy for **FA**, **EDA**, **LIDA** and **HIDA**) (Figure 37.d).

It is worth noting that I used accuracy as the performance metric (rather than Kappa which is used in [11, 94]) because the class label distributions in my experiments were balanced. In comparison, the distributions of "Mind Wandering" labels in [11] and the boring/confusing topics in my previous study were highly skewed. The corresponding Kappa scores across the 18 subjects were 0.77 for **FA**, **EDA** detection, and 0.591 for **FA**, **IDA** detection.

The classification accuracy of the binary classifiers is comparable to EEG and eye-gaze based methods (accuracies range from 79% to 99%) [104]. I also investigated multiclass classification to differentiate the *type* and *intensity* of divided attention, which were not presented in [104]. Please note that a strict performance comparison with [104] was not possible due to the task difference (mobile MOOC learning vs. speed reading on PCs) and the number of participants (18 vs. 6). However, it is still inspirational to show the feasibility to detect divided attention on unmodified mobile devices at a performance comparable to dedicated eye-trackers and EEG sensors.

6.4.2 User-Independent Classification

While personalized models usually have better performances, it is necessary to estimate the expected system performance to simulate the "*cold-start with zero adaptation*" scenario. Therefore, I also built user-independent models to detect divided attention.

In addition to the PPG signal processing I have done for the user dependent models, I found the following two techniques can improve the performance of user-independent models[3]: 1) Smoothing the RR-intervals and interpolating the RR-intervals at 20Hz; and 2) adding four

[3] These techniques do not improve user-dependent models according to my experiments.

additional HRV features, i.e., pNN5, (percentage of adjacent RR-intervals with a difference longer than 5ms), SDANN (standard deviation of AVNN in all k segments of a window), SDNNIDX (mean of SDNN in all k segments of a window), and rMSSD/SDNNIDX.

Table 5. The performance of the user independent models for divided attention detection.

Condition	Chance	Accuracy	Precision	Recall	Kappa
FA vs. EDA	50.0%	72.2%	0.75	0.67	0.44
FA vs. LIDA	50.0%	75.0%	0.71	0.83	0.50
FA vs. HIDA	50.0%	83.3%	0.80	0.89	0.67
FA + EDA vs. LIDA + HIDA	50.0%	83.3%	0.90	0.75	0.67
FA vs. LIDA vs. HIDA	33.3%	63.0%	0.67	0.63	0.44
FA vs. EDA vs. LIDA vs. HIDA	25.0%	50.0%	0.52	0.50	0.33

I used RBF-SVM for classification. I found that segmenting the signals into larger window size generates better prediction accuracy. Therefore, PPG data of an entire video session (8 minutes) was treated as an instance in the user independent model. This made the dataset contain 72 instances (18 instances per condition). The leave-one-subject-out method was used to evaluate my user-independent models. Table 5 lists the performance of different classifiers. Although the accuracies were lower than corresponding performances in user-dependent models, both accuracies and Kappa scores were far above chance. It is expected that the detection accuracy of **HIDA** is higher than **LIDA** and **EDA**, considering that **HIDA** can cause stronger changes in physiological arousal due to the higher level of divided attention.

I also ran a linear regression analysis to gain further insights on the relationship between HRV features and different attentional states. I found that for detecting **FA** and **EDA**, the important features were MAD (p = 0.0167) and pNN5 (p = 0.0258); while important features for

FA, IDA prediction were MAD (p = 0.0054), SDNN (p = 0.0158), and SDANN (p = 0.0326). For **FA**, **EDA**, **LIDA**, **HIDA** prediction, the only important feature was MAD (p = 0.0325). It can be seen that MAD was the most important features for the user-independent classifiers of divided attention.

6.5 DISCUSSIONS

The classification accuracies of the user-independent models were much lower than the user-dependent models in my experiments. In addition to the inherent challenges in building user-independent models, there are three additional reasons I need to take into account. First, although all participants reported that **FA** to be less distractive than **DA**, the difference could be small (i.e., less than 1 on a 7-point Likert scale) for some participants; second, several participants reported that they had a habit of listening to music while studying, so they do not consider the loud music in the **EDA** condition a major distraction; third, although 16 out of 18 participants reported that **IDA** is more distractive than **EDA**, two participants reported the opposite. Such differences in the perception of **IDA** and **EDA** among my participants may be additional factors that limit the performance of a user-independent model. Inspired by the recent advances in speech recognition, the accuracies of user-independent recognition can be further improved with more training data from highly diversified users and sub-group modeling/adaptation.

Since the user dependent model achieved much higher accuracy than the user independent model, the user independent model could be used initially when there is not enough personal data for the user dependent model. AttentiveLearner then switches to a personalized model when more training data of the user are gathered after a couple of sessions. In this way, I could deal with the “cold start” problem for the personalized system.

The automatically detected divided attention information can be used to improve mobile MOOC learning. First, since different types of learning activities have different requirement for attention, the system could switch to less attention-demanding learning activities, such as discussion forums, when consistent divided attention is detected. This allows the system to assign learning tasks properly so that learners’ can achieve optimal learning efficacy. Second, appropriate intervention technologies could be developed in the system to directly address divided attention. For example, the system can use visual and tactile feedback to remind learners to focus on the video alone when divided attention is detected. The system may also provide customized reviews on the sections when divided attention is detected.

7.0 ADAPTIVE INTERVENTIONS TO IMPROVE LEARNER ENGAGEMENT AND PERFORMANCE

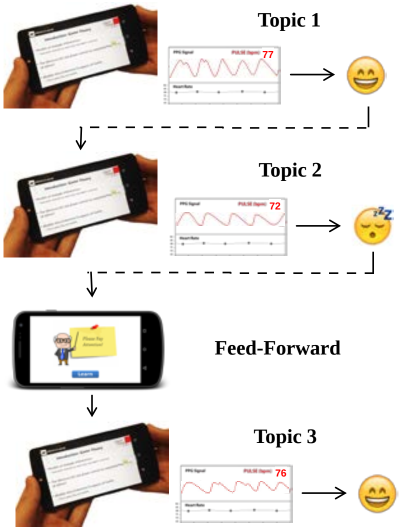


Figure 38. The proposed system detects disengagement by analyzing implicitly captured PPG signals. The system uses feed-forward to remind learners when they are disengaged.

I have demonstrated AttentiveLearner as a “sensorless” approach that implicitly captures learners’ physiological signals and infers their cognitive states. While the previous chapters focus on improving instructors' understanding of the MOOC learning process via *offline analytics*, this chapter demonstrates a novel intervention technology, *context and **c**ognitive state triggered adaptive **f**eed-**f**orward (C2F2)*, which provides *real-time* predictions of the learner’s cognitive states during learning and adaptively responds to the predicted cognitive states. C2F2 monitors a learner’s engagement while she is watching lecture videos and adaptively reminds the learner of important upcoming content when she is disengaged (Figure 38). C2F2 is built on top of AttentiveLearner and provides real-time predictions of the learner's engagement state for each learning topic using the implicitly captured PPG signals. Moreover, the system proactively initiates feed-forward interventions to re-engage the learner if she is currently disengaged and the upcoming topic is important. My prediction is that by monitoring and responding to disengagement at the right time, *C2F2* will yield superior learning performance compared with a standard non-interactive learning system.

In this section, I present the design and evaluation of C2F2 and perform a comparison between my PPG-based engagement detection method and state-of-the-art electroencephalography (EEG) based methods [115, 116]. The contents of this chapter are modified from a published paper [131].

7.1 BACKGROUND

Disengagement/boredom is one of the most frequent affective states and is persistent across various learning environments [7, 103]. Previous studies have shown that boredom was a negative affective state which interactive learning environments should focus on detecting and quickly responding to [7, 24, 42, 91, 103, 119]. Craig et al. [24] investigated the role of affective states in learning through a user study with 38 low domain knowledge undergraduate students. Results of the study showed a significant negative correlation between learning gains and boredom. Baker et al. [7] analyzed data on students' affective-cognitive states as they used three educational environments. They found that boredom was the most persistent state and was the only state that led students to game the system, which was known to be associated with poorer learning. Boredom could disengage learners from educational activities and seriously decrease learners' abilities to acquire knowledge [119]. Boredom was also found to adopt a persistent temporal quality [42, 103], where students were less likely to be re-engaged once they were disengaged.

Given the harmful effects of boredom on learning, it is important for intelligent educational systems to maintain learner engagement and regulate boredom state during learning. The proposed intervention technique, *C2F2*, is a disengagement repair technique specifically designed for mobile MOOC learning.

7.2 DESIGN OF C2F2

7.2.1 The C2F2 Intervention

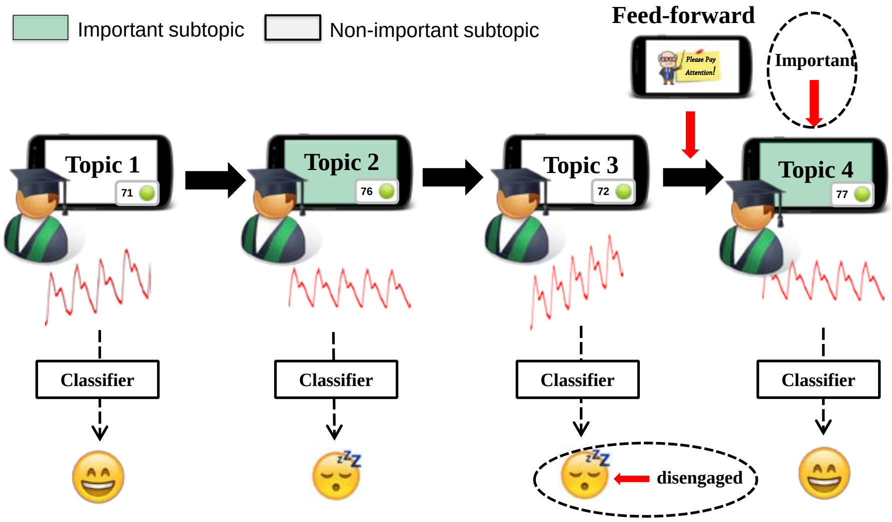


Figure 39. The working mechanism of C2F2.

A feed-forward reminder is presented after Topic 3 because the learner is disengaged watching the Topic 3 video and the next topic (Topic 4) is important.

The design of C2F2 intervention is based on two key assumptions of MOOC learning. First, if a learner becomes disengaged watching one video, she is likely to stay disengaged watching similar videos shortly. I made this assumption based on the temporal persistence nature of disengagement/boredom [7, 42]. Going to the next video alone is unlikely to increase the learners' engagement, as the basic learning activity (video watching) is unchanged, and the follow-up videos usually have the same teaching style on relevant topics. Therefore, I develop C2F2 to repair disengagement and maintain sustained engagement across multiple video sessions. My second assumption is that not all parts in a lecture video are of equal importance. Some segments present key concepts or methods, while others may present less relevant or duplicate content. The inclusion of topic importance helps us isolate and quantify key factors that influence the learning outcomes. Because of these two assumptions, I consider both the learner's cognitive states and the intrinsic importance of the upcoming learning topic to determine the timing of intervention.

C2F2 uses feed-forward reminders to draw the learner's attention to the video when he/she is disengaged. *C2F2* is triggered before a video if the following two conditions are met at the same time: 1) the system detects that the learner is in a disengagement/boredom state watching the last video; and 2) the next video is an important subtopic and will be assessed in tests or exams (Figure 39, the feed-forward after Topic 3). If only one condition is satisfied, or none condition is satisfied, the system directly presents the next subtopic. By considering both contents of the video, and learners' real-time cognitive states, I hope to effectively regulate their disengagement/boredom state without frustrating them with too many feed-forward.

7.2.2 Disengagement Detection

When a learner is watching lecture videos with *C2F2*, the system also implicitly captures her PPG signals at the same time. The raw PPG signals are processed by LivePulse to extract RR-intervals and instant heart rates. Outliers of the RR-intervals are removed using the same heuristics as previous studies. 24 dimensions of heart rate features are then extracted. Similar to the study presented in Chapter 5, half of these features are global features extracted from the PPG signals of the entire subtopic video. These global features are: 1) Mean-HR; 2) SD-HR; 3) AVNN; 4) SDNN; 5) pNN50; 6) rMSSD; 7) MAD; 8) pNN12; 9) pNN20; 10) SDANN (standard deviation of the averages of RR-intervals in all *k* bins); 11) SDNNIDX (mean of the standard deviations of RR-intervals in all *k* bins); 12) SDNNIDX/rMSSD. The other 12 dimensions are corresponding local features extracted by averaging the same features in multiple fix-sized, non-overlapping local windows within the subtopic video. For each participant, the features are normalized using the same features of a two-minute baseline PPG signal sequence collected before the learning session.

I used WEKA and LibSVM to train and optimize the classifier using data collected from a 10-participant pilot study reported in the evaluation section. The final prediction algorithm (RBF-SVM) can run in real time on mobile devices. The classifier predicts whether a learner was disengaged watching a subtopic video immediately after the learner watched that video. On a Nexus 5 smartphone, each prediction only takes on average 275 millseconds, which is hardly noticeable according to participants in the user study.

7.2.3 Feed-Forward Reminder Design

The feed-forward reminder prompts a learner of upcoming important topic so as to redraw her attention back to the videos. Figure 40 is the design of feed-forward reminders after two rounds of pilot studies (4 subjects each). The cartoon character acts as a learning companion, which attracts the learner's attention. I piloted with a number of messages displayed to the learner but finally chose to display a simple message "*Please Pay Attention!*". An audio response "*The next topic is very important. Let's pay more attention to it!*" is also played immediately after the feed-forward shows up. The learner needs to explicitly acknowledge the feed-forward reminder by pressing the "*Learn*" button.

Through my pilot studies, users commented that they preferred direct, concise messages, more than indirect, polite messages. Pilot study users also suggested that I should avoid using negative statements. According to one user, statements such as "*You should pay more attention*" or "*you are not paying enough attention*" discouraged him, especially when he thought he had already paid enough attention to the video. Therefore, in the audio message, I attribute the occurrence of feed-forward to "*the next topic is very important*" to avoid eliciting negative emotions from the user.

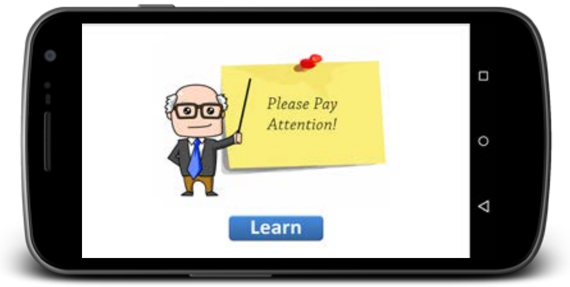


Figure 40. The feed-forward reminder presented to users.

My feed-forward design also adopted a high-interruptive presentation. The learner has to explicitly acknowledge it before watching the next video. According to [50], the high-interruptive presentation is more effective than low-interruptive indicators when the learner is in a negative learning state.

7.3 EVALUATION

I conducted a lab-based study to investigate the effectiveness of *C2F2* on learning. I have the following hypothesis: *In a given learning task, providing adaptive feed-forward before important topics when learners are disengaged will increase learning performance compared with a no feed-forward baseline*.

To gain a thorough understanding of C2F2, I implemented three alternative designs of feed-forward interventions. The first design provides *no feed-forward*, the second provides *context only feed-forward* which presents feed-forward before randomly selected important subtopics, and the third provides *cognitive only feed-forward* which presents feed-forward after learner disengagement is detected regardless of whether the next subtopic is important or not.

7.3.1 Experiment Design

I conducted a between-participant study in which I manipulated the presentation timing of feed-forward within a mobile MOOC learning system. The independent variable was the *type of feed-forward intervention* received by the participant: (1) *no feed-forward*, (2) *context only feed-*

forward, (3) *cognitive only feed-forward* and (4) *C2F2*. The dependent variables included participants' recall of the video content, their learning gains, and their perceptions of the system.

In the experiment, participants used our mobile MOOC client to study an introductory lecture about computer and network security, a topic that participants were unlikely to have prior knowledge, while being "representative" as a real-world STEM learning topic. The lecture is divided into six videos based on the subtopics: "*Cryptography Basis*", "*Computer Virus and Worms*", "*AIC Principles*", "*Cyber Crimes*", "*Access Control*", and "*Session Hijacking*". The length of each video has been adjusted to exactly 4 minutes and 30 seconds, leading to a total instructional time of 27 minutes.

Because our feed-forward technique also considers the importance of subtopics, I selected the second ("*Computer Virus and Worms*"), third ("*AIC Principles*"), fifth ("*Access Control*") and sixth ("*Session Hijacking*") videos as the important ones as these four videos convey the most essential and relevant topics. In comparison, the first and the fourth video clips either contain some trivia, non-technical content, or duplicate content from previous videos, thus they are the non-important videos in the study.

To assess learning performance, participants were asked to answer eight multiple-choice questions for each subtopic/video. Unlike previous studies [35, 42, 116] which have the evaluation session after the whole learning session, I chose to present the evaluation questions for a subtopic immediately after the learner watched the video of that subtopic. This design was fair to all subtopics and minimized the effect of different memory abilities among participants. The evaluation questions were asked for all subtopic videos (including the non-important ones) to ensure that participants had no idea of which subtopics were important and should be given more attention. However, I only considered participants' performance on the four important subtopics. One thing to note is that previous research also showed that in-video quizzes could potentially improve learner engagement [64]; however, the effect of quizzes on engagement is out of the scope of this study.

I first conducted a pilot study to train and optimize the disengagement prediction classifier. I recruited 10 participants (4 females) between 23 to 33 years old (μ=27.8, σ=2.8) for the pilot study. All participants were graduate students at the University of Pittsburgh. Participants watched the six subtopic videos introduced earlier using AttentiveLearner (no feed-forward was presented). Immediately after watching each video, participants were instructed to rate their perceived engagement levels while watching the video on a 5-point Likert scale. Participants' self-reported ratings on the subtopics were used as the ground truth when evaluating performance of the classifiers. Of the 60 ratings (6 videos x 10 participants), 51.67% indicated disengaging learning experience (rating <= 3). I used the leave-one-subject-out method to evaluate the performance of classifiers. Therefore, all results reported were user-independent. The RBF-kernel SVM had best overall Kappa (Kappa = 0.349, accuracy = 68.33%) predicting learner disengagement.

In real-world usage scenarios, the system can present a feed-forward reminder whenever it detects that the learner is disengaged, leading to various numbers of feed-forward reminders per learning session depending on the learner's engagement state. Therefore, in the study, I intentionally controlled the number of feed-forward to avoid confounders. All systems, except for the *no feed-forward* system, will present two feed-forward reminders to the learner for the six subtopic videos.

For the *context only feed-forward* system, feed-forward reminders are presented before two randomly selected important subtopics. For the *cognitive only feed-forward* and *C2F2* system, I designed an algorithm that decides two optimal positions to present feed-forward reminders. The algorithm gives higher priority to the videos participants are more likely to be disengaged with by setting different classification thresholds (determined by the probability estimation of LibSVM) of the disengagement classifier for the six subtopic videos (thresholds are 0.8, 0.6, 0.5, 0.6, 0.5, 0.5 respectively). The adaptive thresholds are determined based on participants' average engagement ratings reported for the six videos in the pilot study. For example, because none participants reported disengagement experience for the first video, it has a high classification threshold. The system will stop presenting any feed-forward if it has already presented two feed-forward reminders. If a learner is always predicted as being engaged, the feed-forward reminders will be presented before the last two (important) subtopic videos. In this way, the same number of feed-forward is guaranteed for all participants.

7.3.2 Procedure

The whole study took about one hour, and each participant was compensated with a $10 gift card after completing the MOOC course. The study consisted of four phases:

Introduction. Participants first signed an informed consent and completed a demographics questionnaire. Next, participants were instructed to use C2F2 to watch a forty-second warm-up video to get familiar with the tangible video control interface.

Initial Quiz. Participants were required to take an eighteen-question multiple-choice quiz (three questions for each subtopic) to assess their prior knowledge of the learning topic.

MOOC Learning and Evaluation. 48 participants were randomly assigned to one of the four experimental conditions (12 participants each condition). Depending on the experimental condition, participants used one of the four mobile MOOC systems with different feed-forward interventions.

After participants watched a subtopic video, they immediately evaluated this video with a *Subjective Impression Questionnaire* (B.4). Participants also took an 8-question, multiple choice quiz, which tested their understanding of the subtopic video they've just watched. After participants completed the questionnaire and quiz, they continued to learn the next subtopic.

During the MOOC learning and evaluation phase, my participants also wore a Neurosky MindWave headset which measured and stored their EEG data during learning.

Qualitative Feedback. Participants first completed the *Subjective Impression Questionnaire* (B.4) of the whole learning session. Next, each participant took a post-experiment (B.3) questionnaire to obtain their subjective evaluations of the mobile MOOC application.

7.3.3 Participants and Apparatus

Forty-eight subjects (28 males and 20 females) participated in my study (Figure 41). Each of the four conditions was gender balanced (seven males and five females). The number of participants were determined by running a prior power analysis (assumptions of the standardized group mean difference of the participants was based on [116]). Moreover, the number is on par with many previous studies on similar topics in educational setting, for example, Szafir et al. [116] (48 subjects, 4 learning conditions), Ogan et al. [86] (12 subjects, 1 learning condition). The average participant age was 23.4 (σ = 3.5) ranging from 18 to 32. All participants were undergraduate or graduate students recruited from the University of Pittsburgh by flyers posted around the campus. Prior familiarity with the lecture used in the study was low; the average pre-lecture quiz score is 12.31% (σ = 12.9%). No significant differences in pre-lecture quiz score were found across conditions.

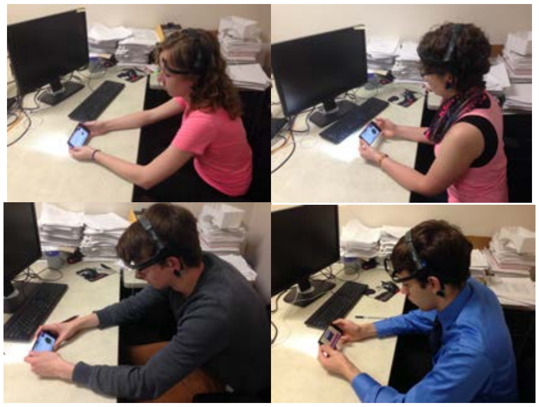


Figure 41. Sample participants in my experiments. Participants also wore an EEG headset during learning.

My experiment was completed on a Nexus 5 smartphone with a 4.95 inch, 1920 x 1080 pixel display, 2.26 GHz quad-core Krait 400 processor, running Android 5.0.1. It has an 8 mega-pixel back camera with an LED flash.

7.4 RESULTS

7.4.1 Signal Quality

The mobile MOOC systems collected PPG signals while participants were watching lecture videos and stored them on the mobile device. I have collected a total of 1305 minutes of PPG signals from the 48 participants (average 27 min 11s per participant). I analyzed the quality of collected PPG signals by investigating the RR-intervals in a 5-second moving window. I found that of the 288 (48 x 6) video sessions, 66.7% of them contained more than 90% high-quality PPG signals, 82.29% of them contained more than 80% high-quality PPG signals, and 89.23% of them contained more than 70% high-quality PPG signals. Only 5.9% of them contained less than 60% high-quality PPG signals. This suggested that AttentiveLearner generally collected reliable PPG signals from learners' fingertips during video watching.

7.4.2 Feed-Forward Accuracy

To verify that my system was working correctly, I first checked if the feed-forward intervention was indeed presented at the correct time. Participants' self-reported engagement levels were used as the ground truth. A feed-forward was presented at the correct place if the learner was disengaged while watching the last video and the next video was important. I also excluded participants whose ratings suggested consistent engagement throughout the whole learning session as in this case the position of feed-forward interventions was likely to make no difference. For the *context only condition*, 39.13% feed-forward was presented at the right position, for the *cognitive only condition*, 27.79% feed-forward was presented at the right position, and for *C2F2*, 62.5% feed-forward was presented at the right position. If I did not consider presenting feed-forward before important videos as a constraint, then in the *cognitive only condition*, 56.6% feed-forward was presented at the right position.

One thing to note is that in all feed-forward experimental conditions, the participant received two feed-forward reminders. Some of these feed-forward reminders are extra, that is, they are triggered not because of the learner's cognitive state, but triggered to balance the number of total feed-forward received. Disregarding these extra reminders, only 12.25% feed-forward was triggered at a wrong place in *C2F2* condition, 37.5% feed-forward was triggered at a wrong place in *cognitive only* condition, and 41.67% feed-forward was triggered at the wrong place in *context only* condition. If I did not consider presenting feed-forward before important videos as a constraint, then in the *cognitive only* condition, only 12.5% feed-forward was triggered at the wrong place. Therefore, in *C2F2* condition, my algorithm generally presented feed-forward in the correct position.

7.4.3 Learning Performance

My experiment was based on the concept that different feed-forward interventions would affect learning, thus I first utilized analysis of variance (ANOVA) to analyze the effect of feed-forward interventions on participants' learning performance. I looked at participants' performance on the post-video quizzes only for the important subtopics (4 x 8 = 32 questions in total).

Information Recall, measured by the percentage of correctly answered questions, were on average 63.57% (σ = 17.75%), 65.16% (σ = 17.47%), 68.71% (σ = 15.72%) and 76.39% (σ = 12.17%) in the *no feed-forward*, *context only feed-forward*, *cognitive only feed-forward*, and *C2F2* conditions respectively (Figure 42). A one-way between subject ANOVA found no significant effect of the type of feed-forward interventions on *Information Recall*: F (3, 44) = 1.4754, p = 0.2343. Post-hoc pairwise t-tests with Bonferroni correction suggested no significant difference between the *C2F2* condition and the *no feed-forward* condition (t (22) = 0.1281, p = 0.0602, d = 0.8161). However, the large effect size (Cohen's *d* > 0.8) indicated the possibility of a significant relationship between these two conditions.

I used proportional learning gains, computed as *(post-test – pre-test scores)/(1 – pre-test scores)*, to measure *Learning Gains*. Average *Learning Gains* were 60.03% (σ = 19.38%), 60.50% (σ = 19.54%), 64.17% (σ = 15.72%) and 72.18% (σ = 13.43%) in the *no feed-forward*, *context only feed-forward*, *cognitive only feed-forward*, and *C2F2* conditions respectively. No significant effect of the type of feed-forward interventions on *Learning Gains* was found: F (3, 44) = 1.2030, p = 0.3198. Post hoc pairwise t-tests with Bonferroni corrections revealed no significant differences between the *C2F2* condition and the *no feed-forward* condition (t(22) = 0.1214, p = 0.1008, d = 0.7026).

Although I did not observe significant learning differences between the conditions for all participants, the large effect size of the t-tests between the *C2F2* condition and the *no feed-forward* condition indicated that there probably existed a significant interaction in the data worth further investigation. Therefore, I divided participants in each condition into two groups. Based on participants' scores on the quizzes, I divided participants in each condition into the bottom half performers (six participants) who earned the lower score on the quizzes and the top half performs.

The bottom half performers had an average *Learning Gains* of 43.75% (σ = 11.16%), 44.45% (σ = 11.59%), 52.74% (σ = 5.43%) and 61.94% (σ = 4.32%) respectively. I observed significant main effect of type of feed-forward interventions for the bottom half performers on *Information Recall*, F (3, 20) = 6.11, p = 0.004, and on *Learning Gains*, F(3, 20) = 5.68, p = 0.0056. Post hoc pairwise t-tests with Bonferroni correction (α = 0.05/6 = 0.0083) revealed that for the bottom performers, they learned significant better in the *C2F2* condition than in the *no feed-forward* condition (t = 0.1829, p = 0.0018) and the *context only feed-forward* condition (t = 0.1749, p = 0.0025). Although the sample size (N = 6) is small, the small p value and the large effect size suggested a high practical significance. However, I did not observe significant main effect of type of feed-forward interventions for the top half performers on *Information Recall*, F(3, 20) = 0.7580, p = 0.5308, and on *Learning Gains*, F(3, 20) = 0.3714, p = 0.7745. Figure 42 shows the major results of the study.

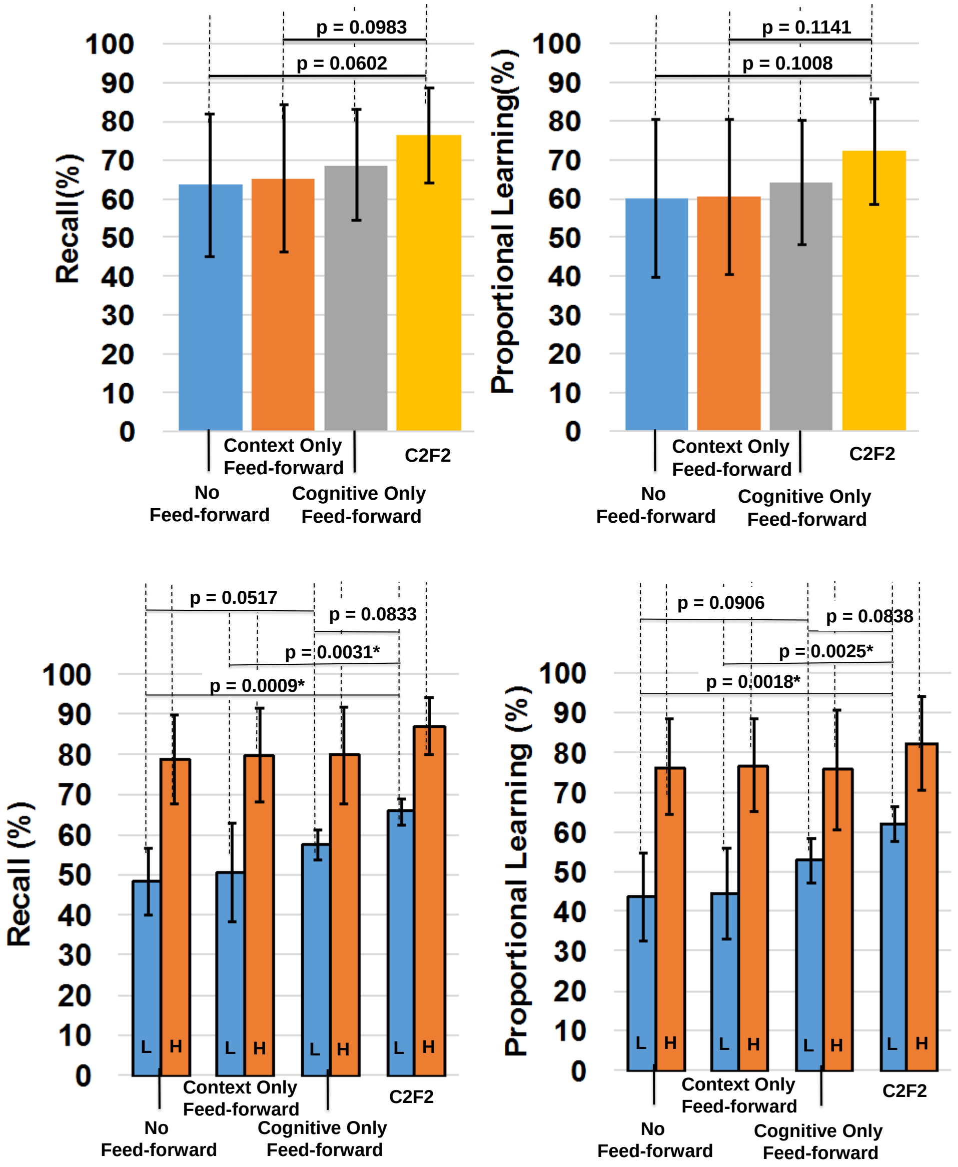


Figure 42. Recall rate and learning gains of different conditions.

Top: general results of all participants, bottom: detailed results for bottom performers (L) and top performers (H).

Marginal and significant p values are reported. (*) denotes significant differences.

The results suggested that *C2F2* were especially effective for bottom performers. A closer look at participants' learning performance showed that the top four performers in the no feed-forward condition did well enough and achieved comparable performance as the top performers in other conditions. Looking at these participants' self-reported impression on the lectures, three of them reported consistent engagement and attention throughout the whole session. On the other hand, four of the six participants with the lowest scores across all conditions were from the no feed-forward condition. Two of them reported consistent disengagement since the fourth video in the lecture. This finding suggested that *C2F2* was useful for learners who became disengaged from learning and lacked the self-regulation ability to refocus on the learning content. For these learners, *C2F2* prevented them from staying in a disengaged state and reoriented their attention back to the learning materials.

My results also suggested that presenting feed-forward based on topic importance alone did not improve learning. Presenting feed-forward based on learners' cognitive state was more effective. This is because cognitive-state triggered feed-forward directly addresses the learner's disengagement state, thus it is more effective at helping the learner maintain sustained engagement and attention throughout the whole learning session.

7.4.4 Subjective Feedback

Participants reported an average rating of 3.67 (σ = 0.89), 4 (σ = 0.60), 4.17 (σ = 0.72), 4.17 (σ = 0.58) on a 5-point Likert Scale for the mobile MOOC system in the *no feed-forward, context only feed-forward, cognitive only feed-forward* and *C2F2* conditions respectively. I found no significant main effects of the type of feed-forward interventions on participants' impressions of the learning session (e.g., attention and engagement level, the effort put into the lecture, perceived learning, etc.).

Participants were generally very positive towards feed-forward. They commented that the feed-forward intervention indeed reoriented their attention to the video when they were disengaged:

"I thought it is a good idea. I think it grabbed my attention when I was zoning out. So overall pretty good."

"The feed-forward alert really helps me re-engage when my mind starts wandering."

"It was helpful when I knew I needed to pay more attention. It was distracting when I felt that I was paying attention."

Some participants reported that the feed-forward was presented at the wrong place, especially in the *context only feed-forward* condition. The self-perceived accuracy of whether the feed-forward was presented at the right place could affect how a learner responds to feed-forward:

"I think the feed-forward alerts were presented at random places. It showed up when I paid a lot of attention and did not show up when it should. So I did not find it useful and just ignored it."

Some problems of the feed-forward were also identified from the experiment. One participant mentioned, "*I paid extra attention for the 4th video when I saw the alert (feed-forward), and then paid less attention in the following video.*" This suggested that asking learners to pay more attention to one video could potentially make them pay less attention to another video. The feed-forward intervention is also not necessarily helpful for everyone. One participant commented, "*When I am learning, extrinsic motivation often isn't helpful for me. If I do not find a topic interesting, it is hard to pay attention even if I'm told to pay attention*".

Participants also reported that the content of a video affected their overall engagement for that video, "*I could tell I preferred the one video about computer virus. It was more interesting for me and also easier to follow.*"

7.5 COMPARISON WITH EEG

Another goal of the study was to directly compare the performance my disengagement prediction method (camera-phone based PPG-sensing) with the EEG-based engagement monitoring method, which is the current state-of-the-art technique to infer users' engagement and attention state [115, 116] from physiological signals. Therefore, all participants in the study were asked to wear a Neurosky Mindwave EEG headset (Figure 42) during the learning session. This setup was similar to previous studies investigating the use of EEG-monitored attention during learning [115, 116]. Among the 48 participants in the experiment, the EEG signals from 11 participants were either incomplete (due to the device problem) or partially unusable (highly corrupted by noises). Therefore, I compared the performance of the PPG-based method and EEG-based method using data from the remaining 37 subjects.

I performed off-line analysis and used the EEG-based engagement-monitoring algorithm in [115, 116] to calculate and filter an attention index. A participant's engagement level for a given video was determined by calculating the mean of the attention index recorded during that video. I used participants' self-reported engagement ratings for each video as the ground truth. I evaluated performance of the method using three measures: accuracy of using the EEG attention index to identify the video (of the six videos in the lecture) with the lowest engagement for each subject (*acc1*); accuracy of detecting the bottom two videos with the lowest engagement for each subject (*acc2*), and accuracy of detecting the bottom three videos (*acc3*) with the lowest engagement for each subject.

For direct comparison, the same measures were also applied to evaluate the performance of my PPG-based engagement prediction method. I used a Ranking SVM algorithm (SVM^{rank}) to predict the ranks of learners' engagement levels for the six videos they watched. Based on the ranking, I was also able to predict the video(s) with the lowest engagement. The same set of PPG features as well as signal processing methods used in 7.2.2 were also used. The leave-one-subject-out evaluation was utilized to evaluate the performance of the ranking model.

The EEG-based engagement prediction method achieved the best accuracy of 55.56% for *acc1*, 62.5% for *acc2,* and 75.93% for *acc3* when the regularization constant was set to 0.02. This means that using the average EEG attention index, I could correctly identify the video which a learner showed the least engagement with 55.56% accuracy. On the other hand, my PPG-based method achieved 69.44% accuracy for *acc1*, 68.05% accuracy for *acc2*, and 76.85% accuracy for *acc3*. The PPG-based method outperformed the EEG-based method, especially detecting the video with the lowest engagement (*acc1*). The reason why my PPG-based classifiers achieved better accuracy might be the specifically trained machine learning models. The EEG-based engagement monitoring method directly calculates the attention index and does not have any learning involved.

One benefit of the EEG-based engagement monitoring method is that the method can capture finer-grained attention changes as the attention index is updated every one second. On the contrary, I used PPG signal sequences of a few minutes to predict learners' general attention and engagement over a period. Another limitation of my PPG-based engagement prediction method is that it is less robust against different learning environments than the EEG-based method. One important characteristic of heart rate and heart rate variability is that they are easily affected by the surrounding environments or the learner’s physical state. In real usage scenarios, learners are likely to study in more interrupted, highly diverse environments (e.g., standing, sitting, public transit). Therefore, one important issue is the confounding effects of the environments on PPG signals. On the contrary, the Mindwave headset collects EEG measurements from the FP1 region of the cortex, which is directly correlated with concentration. In the future, I shall investigate how to infer reliable engagement information from noisy, highly interrupted PPG signals.

One problem I observed during my experiment was that wearing the EEG headset for an extended time could cause physical discomfort. More than ten participants complained about the pain caused by the ear clip and headband of the EEG headset. In my study, participants were instructed to wear the device before the learning session and take off the device after the whole learning session. The sensor tip on forehead could get detached from the participant’s skin due to incorrect adjustment of the device or user movement. This was the main reason why I was not able to collect complete good quality EEG data from the 11 participants excluded from this analysis.

7.6 DISCUSSIONS

One limitation of the proposed *C2F2* technique is that the reminder is only presented before an entire subtopic video. Participants reported that they would also like to receive *within video reminders* immediately after they mind wandered. In this way, they could quickly redraw their attention to the video. However, the accuracy of predicting whether a participant was MW at a moment using the implicitly captured PPG signal is moderate at best (highest precision 40% and highest recall 65% in [94]). Such accuracy is insufficient to support fine-grained reminders within a video.

Another problem is that the feed-forward intervention will be presented when the system detects that the learner is not engaged/paying attention while watching the last video. Although feed-forward could potentially regulate learner's disengagement state, the learner was still disengaged *before* the *C2F2* reminder. To address this problem, *C2F2* technique could be used together with other techniques, such as adaptive review, to improve learning. After the system detects that the learner is disengaged for the last video, the system could present a short review video or slide, or use exercises to help the learner review content of the last video. My colleagues at the University of Pittsburgh have developed the AttentiveReview [95] intervention technique which recommends review materials adaptively from the PPG signals captured by AttentiveLearner.

The current design requires lessons to be divided into small subtopic videos. *C2F2* could make learners pay attention for a while, but learners could still become disengaged halfway through a video if the video was long and boring. Smallwood et al. [111] found an increased mind wandering with time on task. Therefore, it is important to identify the maximum duration of a learning topic/video which allows learners to maintain sustained engagement. Based on subjective feedback from the study, most participants commented that they could stay focused for 3 to 5 minutes after seeing the feed-forward reminder. For longer videos, brief in-video alert in the middle of the video could be used.

8.0 DYNAMICS OF AFFECTIVE STATES DURING MOOC LEARNING

The previous chapters mainly focus on the static presence of affective-cognitive states, such as *Boredom* and *Confusion*, at specific parts of the lecture video. However, learners' cognitive states are dynamic and periodically change [33]. An understanding of the temporal dynamics of certain classes of affective states is necessary for a satisfactory model that integrates affective states and MOOC learning. For example, physiologists have shown that *Flow/Engagement* is the optimal experience during learning [27] when learners are completely focused and engaged. To keep learners in the optimal *Flow* state, I should understand how learners enter other affective states and how to revert them back.

In this chapter, I present a fine-grained analysis of the rapid dynamics of affective-cognitive states that naturally occur during a MOOC learning session.

8.1 BACKGROUND AND MOTIVATION

A series of studies have explored the affective-cognitive states that occur during learning with technologies [7, 24, 29, 36, 40]. Graesser and his colleagues collected online measures of affect in various ways, such as observations by trained judges, self-report ratings, emote aloud protocols, and biological detection through physiological signals [24, 36, 40]. These studies have revealed that *Boredom*, *Engagement/Flow*, *Confusion* and *Frustration* dominate learning experiences, with *Delight* and *Surprise* occasionally occurring but considerably less frequent. In a selective meta-analysis of 24 studies which monitored student affect during interactions with learning technologies, D’Mello [29] found that *Engagement/Flow*, *Boredom*, *Confusion*, *Curiosity*, *Happiness*, and *Frustration* occurred more frequently during learning than other affect. Baker et al. [7] studied the incidence, persistence, and impact of students’ affective-cognitive states during their interactions with three different computer-based learning environments. They found that *Confusion* and *Engagement* were the most common states within all three learning environments. *Boredom* was the most persistent state and was associated with poorer learning and problematic behaviors.

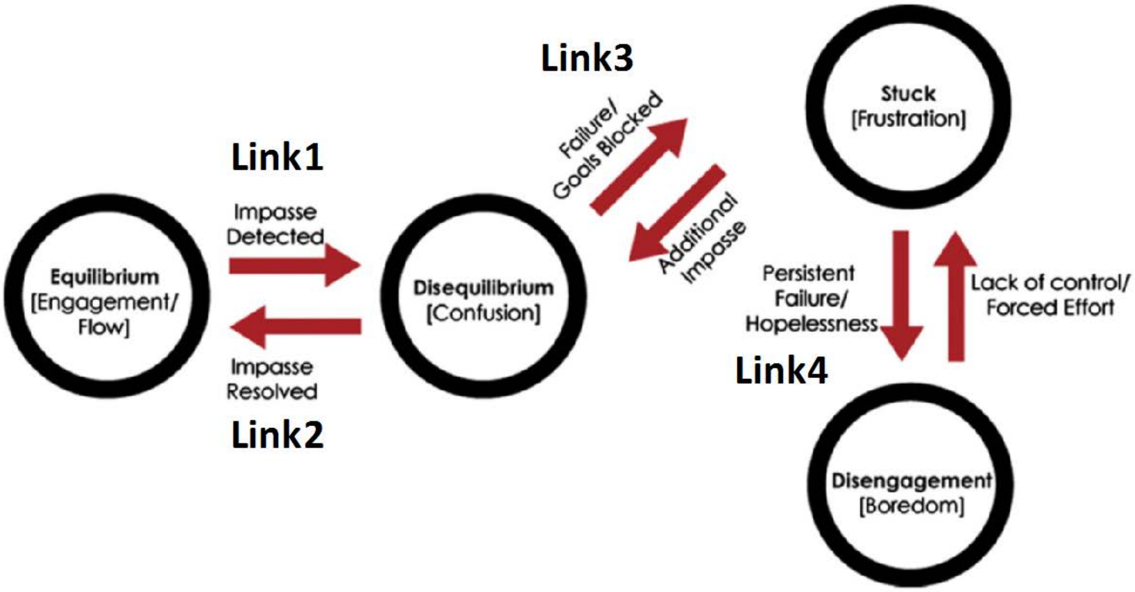


Figure 43. Model of affect dynamics in complex learning [33]

One important aspect of the learner affect that has been explored by researchers is the temporal dynamics of affective states. D’Mello et al. [28, 31, 33, 41] proposed and tested a model of affect dynamics (Figure 43) during complex learning. The model emphasizes the importance of cognitive disequilibrium in complex learning and posits the bi-directional transitions between *Engagement/Flow* (cognitive equilibrium) and *Confusion* (cognitive disequilibrium). The model assumes that learners in a base state of *Engagement/Flow* will experience cognitive disequilibrium and *Confusion* when they are confronted with contradictions, anomalies, or obstacles to goals (Link 1). Learners revert to *Engagement/Flow* if they resolve the impasse and restore equilibrium (Link 2). However, when the impasse cannot be resolved, and important goals are blocked, learners will experience *Frustration* (Link 3), which, if unresolved, will eventually lead to *Boredom*, a point at which the learner is disengaged from the learning process (Link 4). This model specifies how affective states evolve, morph, interact, and influence learning and engagement in complex learning contexts.

Although a series of studies have explored affective states in complex learning environments, little work has been done in the MOOC context. Different from complex learning environments which allow students to gain knowledge through interactive problem-solving activities, MOOC students learn primarily by watching lecture videos. Due to the differences between the learning activities (actively solving problems vs. passively watching videos), many findings in the complex learning domain might not apply to MOOC learning. Additionally, previous studies on student affect in understanding learners' affective experience in relation to the course as a whole on a weekly/module level [38, 39]. For example, Dillon et al. [38] conducted a study to measure a range of self-reported student affect throughout a MOOC course. They used "course-level" affect surveys in which students reported their emotions in relation to the course as a whole at the start of even-numbered modules. There is little work on the moment-to-moment affective state transitions within each learning session.

To expand the limited research of affective states in MOOC contexts and gain a holistic understanding of the dynamics of MOOC learning, I investigated the moment-to-moment affective states during a typical MOOC learning session. This piece of work alone has two unique contributions:

- Through a 22-subject user study, I identify common affective states that naturally occur during a MOOC learning session and the dynamic temporal transitions between these states. This work extends the model of affect dynamics in complex learning environments proposed by D’Mello et al. [28, 31, 33, 41] to MOOC contexts.
- I demonstrate the feasibility of predicting a learner’s moment-to-moment affective states by analyzing the PPG signals implicitly captured by mobile cameras.

8.2 MOOCS VS. COMPLEX LEARNING

While previous studies mostly explored the incidence and transitions of students’ affective states in complex learning environments, this dissertation focuses on the same topic during MOOC learning sessions. I observed four major differences between MOOC learning (watching lecture videos) and complex learning (solving problems): 1) Levels of control over the pace and process of the learning activity; 2) Clarity of learning goals as well as achievement feedback; 3) Levels of interactions offered by the system during the learning process; 4) Require conceptual skills. Table 6 gives a detailed comparison between these two types of learning.

Considering these significant differences, the affect dynamics model for complex learning proposed by D’Mello et al. [33] might no longer apply to the MOOC contexts. It is necessary to propose a new model of affect dynamics specifically for MOOC learning.

Table 6. Major differences between complex learning and MOOC learning.

Content Differences	Complex Learning	MOOCs
1. Levels of control over the learning process	Flexible control over the speed and process of learning	Little control due to the constant information flow of the video
2. Achievement of learning goals	Goals are well-defined (solving the problem following each step) and can be clearly measured (pass or fail).	Goals are often not explicitly defined; Achievement of goals depends on individual learner's standards and expectations
3. Levels of Interactivity	Constant feedback in trial-and-error	Little direct feedback within the video
4. Required conceptual skills	Deeper level of comprehension skills is required (e.g., generate references, diagnose and solve problems, transfer acquired knowledge, generate coherent explanation, etc.)	Shallower level of skills is required (extracting and memorizing information)

8.3 HYPOTHESES

I first came up with a list of hypotheses of learners' affective experiences during MOOC learning. Please note that my focus is the moment-to-moment affect that occurs during short (30 minutes to 1 hour) MOOC learning sessions, which could be very different from the affective experience in relation to the whole course on a weekly/module level [39]. Also, I focus on the primary learning activity within MOOCs, i.e., watching lecture videos. I have three main hypotheses:

1. **The video content plays an important role in the incidence and temporal transitions of learners' affective states in MOOC learning.**

The affect dynamics model for complex learning [33] is motivated by the *cognitive disequilibrium* theory, which emphasizes the critical role of uncertainty that occurs when an individual is confronted with impasses and obstacles. However, the shallower MOOC learning activity mainly requires learners to memorize key phrases and facts from the video and is unlikely to explicitly present impasses and obstacles to challenge learners (*difference 4*). Therefore, the impasse-driven dynamics may not apply for MOOC learning.

One explanation for the origins of emotions in the academic domain is the *control-value* theory of achievement emotions [89]. This theory posits that subjective control over the activities and outcomes, and the subjective values of these activities and outcomes, are central to the arousal of achievement emotions. For example, a student experiences learning-related enjoyment when she has a sense of being able to master the material (high control) and intrinsically values the material (high interest). Inspired by this theory, I assume that perception of the learner regarding the learning materials plays an important role in the incidence and transitions of affective states in MOOC contexts.

2. **Learners are less likely to experience frustration with MOOC learning than with complex learning.**

Frustration is a primary affective state that frequently occurs in complex learning, but I suspect that MOOC learners are less likely to experience *Frustration*. In complex learning, learners experience *Frustration* when they constantly make mistakes and are unable to resolve the impasse [33]. Since MOOC learners are unlikely to be challenged and experience failure in the process of watching a lecture video, I assume *Frustration* will not frequently occur.

3. Transitions to boredom are more easily triggered in MOOC learning.

In the affect dynamics model for complex learning [33], to enter *Boredom/Disengagement*, learners will experience the following sequence of transitions: *Engagement/Flow* → *Confusion* → *Frustration* → *Boredom*. I assume learners are more likely to become bored with the MOOC learning activity for two reasons: there is little interaction and explicit feedback in the learning process (*difference 3*), and there is a lack of clear learning goals to motivate the learner (*difference 2*). According to the control-value theory [89], I think that it is possible for a learner to directly enter *Boredom* from *Engagement/Flow* if she finds the video no longer appealing (low perceived value).

8.4 USER STUDY

To test the above hypotheses, I conducted a user study in which I observed the affective states of college students while they took a mini-MOOC course (30-minute lecture videos). I had two major goals for the study. First, I would like to investigate the dynamics of affective states in a MOOC learning session. Second, I was interested in exploring the feasibility of using implicitly captured PPG signals to predict learners' moment-to-moment affective states.

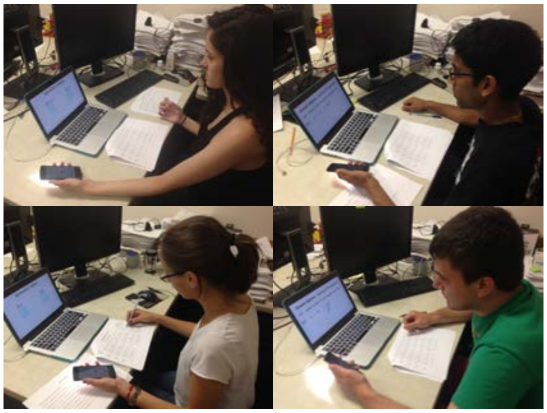


Figure 44. Some participants in my study.

They were watching the MOOC videos with a laptop computer. Participants also held a smartphone running the LivePulse application to collect the PPG signals from their fingertip.

8.4.1 Participants

Due to funding constraints, I managed to recruit 22 University of Pittsburgh students (12 males, 10 females) for the study (Figure 44). While more participants are desirable, the scale of this study was comparable to some previous studies investigating affective states in educational systems, e.g., Diana et al. [37] (19 subjects), Kizilcec et al. [67] (22 subjects), D'Mello et al. [28, 31, 41] (28 subjects). The average age was 22.68 ($\sigma = 4.04$) between19 and 34. The participants were recruited via flyers posted around the campus, and they had various backgrounds: half of them had an engineering background (majored in CS, electrical engineering, etc.), 9 of them majored in liberal arts (history, neuro-science, etc.) and the other 2 went to the medical school. All participants reported little or no prior knowledge of the MOOC course used in the study.

8.4.2 Materials and Equipment

The mini-MOOC course used in the study was a section of the Coursera course "*Cryptography*" taught by Professor Dan Boneh from Stanford University (https://www.coursera.org/learn/crypto/home/week/1). There were three lecture videos in this mini-MOOC course: *Introduction to Ciphers*, *One-Time-Pad*, and *Stream Ciphers*. Duration of the three videos were 10m18s, 11m45s and 8m30s respectively. I chose the course because this was a topic that participants were unlikely to have prior knowledge while being "representative" as a real-world STEM learning topic. Moreover, the course was chosen because the difficulty of its content varied at different parts. The first half of the lecture presented basic concepts and simple ciphers, while the second half of the lecture talked about more challenging content, including mathematical deduction and proofs. I believe such variety in the lecture content could help me elicit diverse emotional responses from my participants. Also, given that the participants had various background, I suspected that some participants (e.g., the students with engineering background) might be more interested in the course than the other students, this allowed me to investigate the effect of interest on learners' affect in MOOC contexts (Section 8.6).

The three lecture videos each contained one or two multiple choice questions (five questions in total) somewhere in the middle to test the learners' understanding of the topics. Learners were given 15 seconds to think about the question before the instructor revealed the answer. The questions were similar to the in-video quizzes [70], which were commonly found in MOOCs on Coursera.

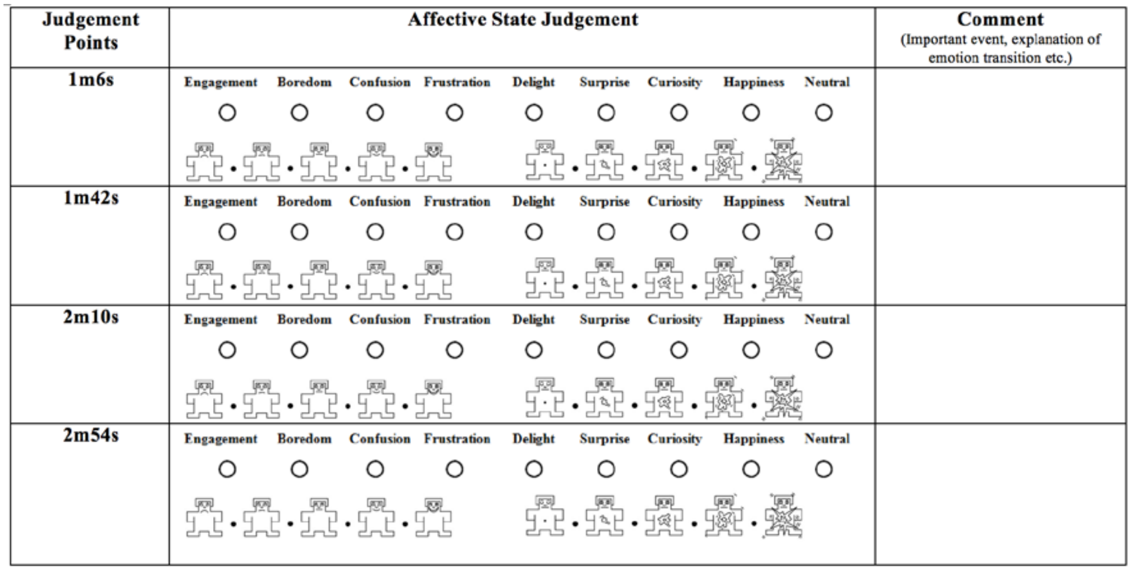
Participants watched the lecture videos on a MacBook Pro with a 13-inch screen. I also recorded participants' PPG signals using a smartphone running the LivePulse application while participants watched the lecture videos. The Nexus 5 smartphone had a 4.95 inch, 1920 x 1080 pixel display, 2.26 GHz quad-core Krait 400 processor, running Android 5.0.1. It had an 8 mega-pixel back camera with an LED flash.

8.4.3 Procedure

Participants completed the mini-MOOC course by watching the three lecture videos one by one. To collect learners' moment-to-moment affective states during learning, I asked participants to provide judgment of their affective states at fixed affect judgment points in the video at which the video paused automatically. Participants reported the affective states they experienced at that instant and clicked the play button to resume watching the video. Affect judgment points were at the end of each concept, usually composed of a definition and a few sentences of explanation, or before the instructor presented the answer to a question. The intervals between two consecutive judgment points ranged from 21s to 80s (average 42.39s), depending on how much time the instructor spent on a concept. There were 16, 16, 15 affect judgment points in the three videos respectively, resulting in 47 affect judgment points in total across the whole lecture.

There are two major differences between my method to collect affect reports and the method used by D'Mello et al. [33]. First, in [33], students provided judgment of their affective states *after* the learning session by viewing the face and screen videos recorded during the learning session. In a pilot study with four subjects, all participants commented that they had a hard time discriminating their affective states based on the recorded face videos alone. I assume that compared with complex learning, learners were presumably more affectively neutral and showed less facial expression changes in the shallower MOOC learning session, thus the retrospective method might not be applicable. Therefore, instead of using the *retrospective* method, I decided to collect affect self-report *during* the learning session, similar to [23, 58].

However, collecting affect report *during* the learning session had the problem of disrupting the learning process with probes [49]. To minimize the interruptions, I did not obtain affect judgments at short, fixed intervals (e.g., 20 seconds) as [33], which could easily interrupt a learner in the middle of a statement. Based on participant feedback in a multiple subject pilot study, learners normally did not experience significant affective state transitions when the instructor talked about the same concept. Therefore, I chose to obtain affect judgments at the end of each concept/slide to avoid interruptions.



Judgement Points	Affective State Judgement	Comment (Important event, explanation of emotion transition etc.)
1m6s	Engagement ○ Boredom ○ Confusion ○ Frustration ○ Delight ○ Surprise ○ Curiosity ○ Happiness ○ Neutral ○	
1m42s	Engagement ○ Boredom ○ Confusion ○ Frustration ○ Delight ○ Surprise ○ Curiosity ○ Happiness ○ Neutral ○	
2m10s	Engagement ○ Boredom ○ Confusion ○ Frustration ○ Delight ○ Surprise ○ Curiosity ○ Happiness ○ Neutral ○	
2m54s	Engagement ○ Boredom ○ Confusion ○ Frustration ○ Delight ○ Surprise ○ Curiosity ○ Happiness ○ Neutral ○	

Figure 45. The interface participants used to report their affective states at each affect judgement point

Figure 45 shows the interface participants used to reported their affective states at each affect judgment point. For each affect judgment point, participants were provided with a checklist of nine states to mark along with definitions of each state. While [33] measured seven states (*Engagement/Flow, Boredom, Confusion, Frustration, Surprise, Delight, Neutral*), I added another two states, *Curiosity* and *Happiness*, which were also reported to occur during learning with technology [29]. The definitions were presented on a piece of paper that participants retained throughout the study. Definition of the first seven states can be found in [33]. *Curiosity* was defined as "*being curious, in regards to the desire to gain knowledge or information.*" *Happiness* was "*being pleased and glad about the condition or situation.*" Participants were asked to choose the affect that best described their state at the exact instant from the checklist. Participants were also asked to rate the level of valence (displeasure to pleasure) and arousal (deactivation to activation) they experienced using the Self-Assessment Manikin's (SAM) [15].

The whole study took about one and a half hours, and each participant was compensated with a $15 gift card after completing the MOOC course. The study contained the following three phases.

Overview. I collected participants' background information and gave a brief introduction to the study. Participants also completed an initial quiz (7 multiple choice questions) to assess their prior knowledge of the class.

MOOC Learning. Participants took the mini-MOOC course. They were required to provide judgments on their affective states at the affect judgment points during learning. Participants' PPG signals were captured throughout the entire learning session.

After participants watched a lecture video, they immediately evaluated this video with the *Subjective Impression Questionnaire* (B.4). Participants also completed a post-lecture test, containing 7 to 9 questions, testing their understanding of the video they had just watched. After the participant completed the questionnaire and quiz, they continued to the next video.

Qualitative Feedback. Each participant completed a closing questionnaire about when and why they experienced each of the nine affective states during learning.

8.5 RESULTS

8.5.1 Affect Distribution

I collected a total of 1034 self-reported affect judgments for the 22 participants' MOOC sessions. Table 7 shows the number of instances of each affect.

Table 7. The number and percentage of each affective state collected in the study.

Affect	Whole Session	Video1	Video2	Video3
Engagement	368 (35.7%)	117 (33.2%)	112 (31.8%)	139 (42.4%)
Boredom	136 (13.2%)	47 (13.4%)	52 (14.8%)	37 (11.2%)
Confusion	150 (14.5%)	57 (16.2%)	55 (15.6%)	38 (11.5%)
Frustration	24 (2.3%)	5 (1.4%)	13 (3.7%)	6 (1.8%)
Delight	17 (1.6%)	10 (2.8%)	4 (1.1%)	3 (0.9%)
Surprise	21 (2.0%)	5 (1.4%)	3 (0.9%)	13 (3.9%)
Curiosity	122 (11.8%)	48 (13.6%)	40 (11.4%)	34 (10.3%)
Happiness	35 (3.4%)	9 (2.6%)	15 (4.3%)	11 (3.3%)
Neutral	161 (15.6%)	54 (15.3%)	58 (16.5%)	49 (14.8%)

I analyzed the proportional occurrences of the affective states experienced by the 22 participants. A repeated measures ANOVA indicated that there was a statistically significant difference in the frequency of occurrences of these states, F(8, 168) = 21.8, $p < 0.0001$. *Engagement/Flow* was the most frequent affect during MOOC learning, followed by *Boredom, Confusion*, *Curiosity*, and *Neutral*. There were only a few occurrences of *Frustration, Delight*, *Surprise*, and *Happiness*.

As expected, there were fewer occurrences of *Frustration* in MOOC contexts than in complex learning (*Hypothesis 2*). According to [33], persistent failure and goal block lead to *Frustration*. In the MOOC learning session, the learners were not given specific learning goals which they had to achieve, thus they were less likely to be frustrated when their learning goals were blocked. Also, watching MOOC videos is generally less mentally demanding than solving complex problems. The learners were less likely to get stuck on the lecture videos than on complex problems. For these reasons, *Frustration* is less likely to happen in MOOCs than in complex learning contexts.

8.5.2 Subjective Feedback of the Origins of Affective States

Before I quantitatively analyze the dynamic transitions of affective states during MOOC learning, I first reviewed the qualitative feedback from participants in the closing questionnaire regarding when and why they experienced each affective state in the study. The qualitative feedback provides insights of what causes and modulates learners' affective states during MOOC learning sessions. I categorized the causal attributions of each affective state into two groups: *internal attributions*, which are related to the learner, and *external attributions*, which are about the learning material. A full summarization of the causal attributions of different affective states is presented in 0.

8.5.3 Dynamics of Affective States

To identify the frequently occurring transitions between different affective states, I used the transition likelihood metric **L** [28, 33] to compute the likelihood of transition between any two affective states. The metric can be represented as L(M_t → M_{t+1}), where M_t is the state at time t (the current states), and M_{t+1} is the next state at t+1. L(M_t → M_{t+1}) is defined by equation 1.

$$L(M_t \rightarrow M_{t+1}) = \frac{\Pr(M_{t+1}|M_t) - \Pr(M_{t+1})}{1 - \Pr(M_{t+1})} \qquad (1)$$

This definition considers the base rate of the subsequent emotions and penalizes associations that are not greater than an expected amount of association ($\Pr(M_{t+1}|M_t) - \Pr(M_{t+1})$). The metric specified in Equation 1 was used to compute the relative

likelihood that individuals in an affective state at time t_i, will remain in the same state or change to another affective state at time t_{i+1}. When analyzing transitions between different states, I used the same data recoding method in [33] to eliminate repetitions between states. For example, X→Y→Y→Z was converted to X→Y→Z. This process reduced the length of the time series to 635 states with a mean of 28.9 states per participant (SD = 9.3). This data recoding procedure was not used when analyzing persistence in the same state (i.e., the students being in the same state for two successive observations).

Our investigation focuses on the frequently occurring states (*Engagement, Boredom, Confusion, Curiosity, Neutral*) and *Frustration,* a primary negative affect in complex learning [7, 33]. To determine whether there is any relationship between immediate and next state, I performed one-sample, t-tests to test whether the likelihood was significantly greater than or equal to zero. Figure 46 presents the descriptive statistics for the likelihood that each of the 6 investigated affective states immediately follows another. I now present analysis for each state.

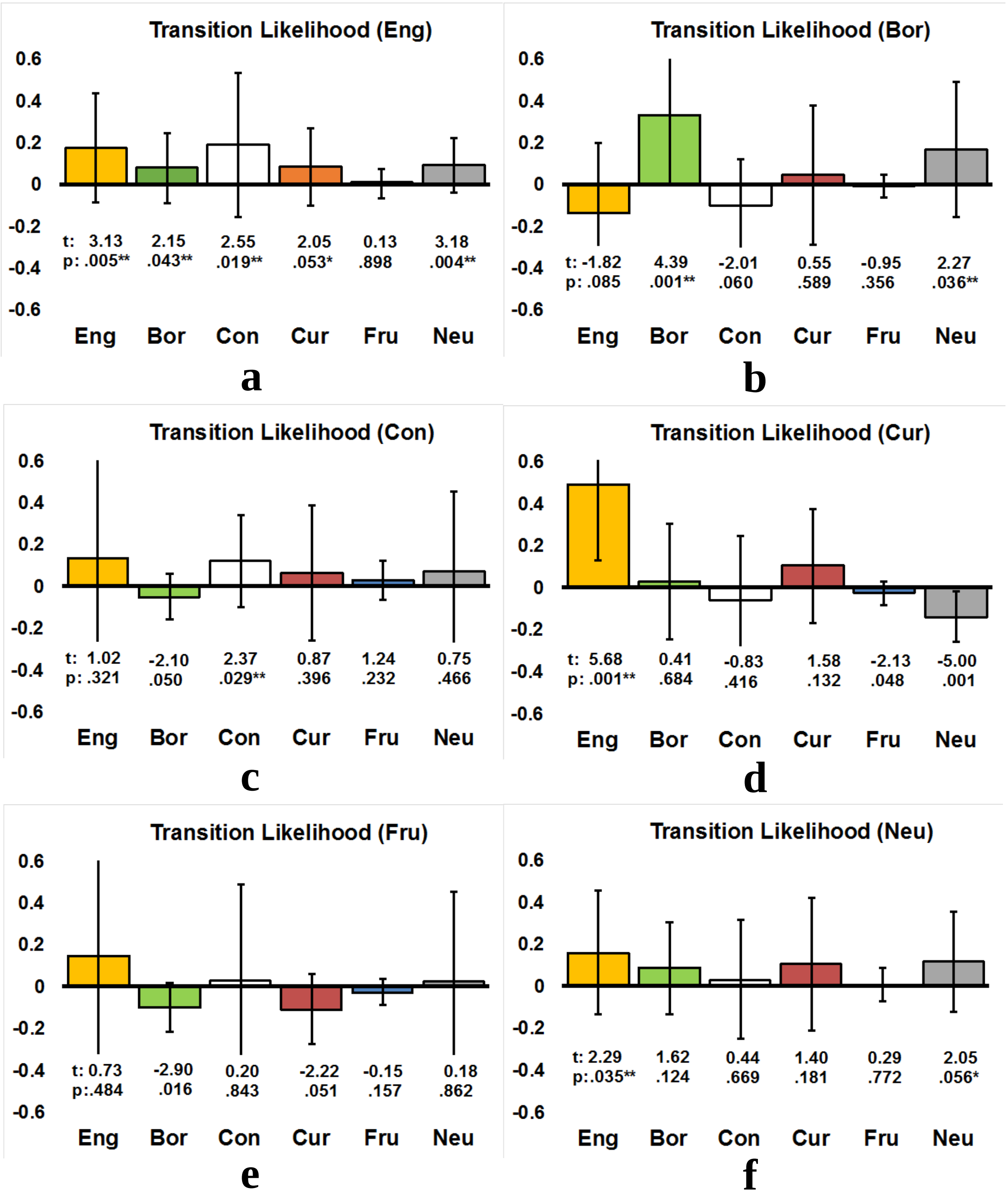


Figure 46. Likelihoods that each of states immediately follows (a) Engagement, (b) Boredom, (c) Confusion, (d) Curious, (e) Frustration, (f) Neutral.

8.5.3.1 Engagement

Engagement is the most frequently occurring state during MOOC learning. Based on participants' subjective feedback, they felt engaged when they were interested in the video content and were able to comprehend the video. This was evident by their comments: "*When I was first starting out, I was interested in learning a new topic.*" [S13], *"When content, although very unfamiliar, seemed simple enough for me to comprehend and follow or accessible."* [S6]. This finding also confirms the control-value theory [89] which predicts that learners feel engaged when they have high perceived control (ability to understand the content) and positive values (interest in the content) over the learning materials.

Figure 46a presents the average likelihood that each of the 6 investigated affective states immediately follow *Engagement*. It also presents results of the one-sample t-tests, which indicate that transitions from *Engagement* to *Engagement* (t(21) = 3.13, p = 0.005), *Engagement* to *Confusion* (t(21) = 2.55, p = 0.019), and *Engagement* to *Boredom* (t(21) = 2.15, p = 0.043) are significantly more likely than chance. The transition of *Engagement* to *Curiosity* is marginally significant (t(21) = 2.05, p = 0.053).

Unlike D'Mello's model [33], I observed that the *Engagement* → *Boredom* transition significantly occurred. Instead of experiencing the *Engagement* → *Confusion* → *Frustration* → *Boredom* process, learners could directly enter *Boredom* from *Engagement*. Based on subjective feedback related to the 29 direct Engagement → Boredom transitions, 51.7% occurred because the material did not appeal to the learner (low perceived value). The content was not interesting, redundant or trivial: *"topic is boring"* [S7], *"took too long on the example"* [S9], "*already knew the content*" [S19], *"same stuff repetitive"* [S22]. 41.4% occurred because the learner could not cope with the lecture (low control): "*lost*" [S6], "*too hard to follow*" [S9], or *"can't understand*

and do not know what to do" [S3]. The other 6.9% occurred because the learner's mind wandered or she felt tired. This finding that *Boredom* in MOOC contexts could be caused by either low perceived value or low control is consistent with Pekrun's control-value theory [29, 89, 91] that *"boredom occurs when value is low, when skills exceed challenges (too high control) and when challenges exceed skills (too low control)"* [29].

The *Engagement* → *Confusion* was also a significant transition. Learners experienced this transition when they did not understand a statement, symbol or notation presented in the video (*"Must've missed something. What's 'key space'?"* [S17], *"unsure of why E is randomized*" [S12]), or when they encounter a question (*"The question has me a little confused."* [S4]).

There is also a marginally significant *Engagement* → *Curiosity* transition. Participants indicated that they felt curious when the instructor asked a question, and they were interested to know the answer (*"want to know the answer or why something is the way it is"* [S20], *"when I was unsure of an answer or unsure how a problem was going to be addressed"* [S19]) or when they encountered a new idea and were curious to know more (*"sometimes I was intrigued and wanted to know more about an idea or concept generally"* [S3], "*Why is E "randomized"? Tell me more.*" [S4]).

8.5.3.2 Boredom

Three learners did not report any experience of *Boredom*. Thus the degree of freedom of the t-tests was 18. Learners in the state of *Boredom* were most likely to remain bored (Figure 46b). The transition from *Boredom* to *Boredom* was significant compared to chance level ($t(18) = 4.39$, $p < 0.001$). The small p-value showed that *Boredom* was persistent, which was also found in complex learning systems [7]. Given the strong persistence of *Boredom*, when instructors create MOOC videos, it is important to adopt engaging video production and interaction techniques so that learners are less likely to become bored.

The *Boredom* → *Engagement* and *Boredom* → *Confusion* transitions were rare in the MOOC learning activity as I observed a negative transition likelihood and a marginally significant difference. However, the *Boredom* → *Neutral* transition occurred above chance (t(18) = 2.27, p = 0.036). Given that learners had no apparent emotions when they were neutral, the *Neutral* state could act as a transitional state for the learner to move out of *Boredom* state and enter another state.

8.5.3.3 Confusion

I found a significant transition from *Confusion* to *Confusion* (t(18) = 2.37, p = 0.029). *Confusion* was rarely followed by *Boredom* because the transition likelihood was negative at a significant level (t(18) = -2.10, p = 0.050). The *Confusion* → *Engagement*, *Confusion* → *Curious*, and *Confusion* → *Frustration* transitions all occurred at chance levels. Three learners did not report experiencing the confused states. Thus the degree of freedom was 18.

I did not find any significant transitions from *Confusion* to another state, so I further investigated whether the amount of *Confusion* could impact the transitions of affective states. The 19 participants reporting *Confusion* were divided into two groups based on the amount of *Confusion* they were experiencing during the learning session. 9 subjects reported more than 15% confusion states (Avg: 27.7%, SD: 9.6%) and they were in the *strong Confusion* group. The other 10 subjects were in the *mild Confusion* group (Avg: 7.0%, SD: 3.5%). I observed significant transitioning from *Confusion* to *Frustration* in the *strong Confusion* group (t(8) = 2.41, p = 0.042). However, the *Confusion* → *Frustration* transition were rare (t(9) = -2.13, p = 0.059) in the *mild Confusion* group. This difference suggests when a learner consistently feels confused, she likely becomes frustrated. Participants' comments also supported this transition from consistent *Confusion* to *Frustration*: "*I was constantly confused which affected my focus. I became frustrated*" [S9], "*The lecturer made me feel frustrated. I often would not tell if I felt frustrated or simply confused*" [S3], "*I felt frustrated when I was feeling confused for 3 to 5 minutes and felt no sense of achievement*" [S14].

I did not observe a strong transition from *Confusion* to *Engagement* as in D'Mello's affect dynamics model [33], which posits that *Confusion* is a key signature of the cognitive disequilibrium state in complex learning. Learners must engage in effortful problem-solving activities to resolve the impasse and restore *Equilibrium/Engagement*. On the other hand, participants in my study reported that *Confusion* → *Engagement* happened when *"(the instructor) made it clearer"* [S9], "*narrator repeated his words once. It was more thorough*" [S6], *"more interesting and better understanding"* [S5], *"New concept has me interested now"* [S6]. Therefore, the occurrence of *Confusion* → *Engagement* depends more on the content and flow of the video (e.g., the video provides follow-up explanations to clarify a confusing idea, or it moves from one topic to a more interesting topic, etc.) than the learner actively solving the problem independently. Because of this large reliance on the content and flow of the video, the learner might remain confused if their questions or doubts are not answered in the video. The learner can become frustrated as the video plays and the content gets more and more difficult before she is ready to advance.

8.5.3.4 Curiosity

Figure 46d shows that learners in the *Curiosity* state are highly likely to transition to the *Engagement* state (t(17) = 5.68, p < 0.001). This transition was also reflected by participants' feedback comments, “*Once you feel curious, you become more engaged”* [S20]. All other transitions occurred randomly.

8.5.3.5 Frustration

Because participants rarely experienced frustration in my study (only 11 participants reported experiencing frustration for a few times), I did not observe any significant transitions after *Frustration*. Different from D’Mello’s affect dynamics model [33], I observed that the *Frustration* → *Boredom* transition was rare and occurred significantly below chance (t(10) = -2.90, p = 0.016). This further proves that learners could easily enter *Boredom* instead of going through the *Frustration* → *Boredom* process. Learners were also unlikely to enter the *Curiosity* state from the *Frustration* state (t(10) = -2.22, p = 0.051).

The transitions from *Frustration* to *Engagement*, *Confusion*, and *Frustration* all occurred at random. I found that Frustration → Engagement transitions might occur when frustration is caused by temporary, yet irritating video production problems (e.g., “*annoyed by choppy voice or writing*” [S21], “*this person’s handwriting can be a little hard to read*” [S4]). The learner returned to the engaged state when the problem was diminished. However, to confirm these speculations and identify significant transitions related to *Frustration*, more affective state data needs to be collected.

8.5.3.6 Neutral

I found a significant transition from *Neutral* to *Engagement* (t(18) = 2.29, p = 0.035), and a marginally significant transition from *Neutral* to *Neutral* (t(18) = 2.05, p = 0.056). Given the significantly occurred Boredom → Neutral transition, there exists this Boredom → (Neutral) → Engagement transition. The *Boredom* → *Engagement* transition occurred when the learner had an increased interest level ("*interested to know info*" [S5], "*new subject*" [S22]) or exhibited a better understanding of the content ("*Okay, what he is saying makes sense now.*" [S3]).

8.5.3.7 Discussions

Based on participants' feedback of when they experienced each affective state, I thought the *Confusion* → *Boredom* transition might be significant: *"The second and third videos got boring after I became confused. I didn't have the background knowledge"* [S9], *"When things dragged on and I didn't fully understand (I got bored)"* [S5], "*When the content seems way over my head and the lecture was too difficult to follow (I got bored)*" [S6]. Based on these comments, I assume that a learner might first feel confused and lost when she missed a fundamental concept in the lecture. As the lecture went on, the learner could eventually get bored if she couldn't follow the lecture and give up on the lecture.

Unlike my expectation, the *Confusion* → *Boredom* transition was rare and below chance (t(18) = -2.10, p = 0.050). One explanation for this is that *Boredom* and *Confusion* might co-exist, but participants were only required to report the dominant affective states at each judgment point in the study. Some participants did report the co-existence of *confusion* when they were bored: "*Notation still confusing*" [S9], "*I have no idea what he was talking about.*" [S3]. Previous studies have also reported the co-existence of *Boredom* and *Confusion* [37, 39]. Another explanation is that the *Confusion to Boredom* transition might happen at a fast rate. Pekrun et al. [91] found that students with a self-concept of low aptitude and low interest in the learning material may not believe that the effort will help them master the material, thus they may quickly become bored and disengaged. Since the intervals between two consecutive affect judgment points was as short as twenty seconds to one minute, the *confusion* to *boredom* transition could have happened at a fast rate of less than one minute.

8.6 FACTORS INFLUENCING LEARNER AFFECT

The quantitative analysis of affect dynamics displays the general moment-to-moment affective state transitions that individuals undergo during MOOC learning. I also performed quantitative analysis to identify the external and internal factors that influence learners' affective experience during MOOC learning.

In the study, other than the affective state judgments, I also asked participants to provide other information which I thought might influence their affective experience during MOOC learning. These include knowledge base (F1, measured by whether they have engineering background or not), interest in the course (F2), perceived usefulness of the course (F3), perceived learning capability (F4, measured using the Perceived Competence Scale), perceived interestingness of the learning material (F5) and perceived difficulty of the learning material (F6). F1 was collected in the entrance survey, F2 to F4 were collected before the participant watched each video, and F5 and F6 were collected after the participant watched each video.

To analyze if learners' affective experiences were correlated with these factors, I calculated the proportional occurrence of each affective state by each participant and ran linear regression analyses. Since *Boredom* could be caused by either a lack of perceived values (disinterested in the content) or a lack of control (high mental demand), I manually labeled each boredom either as *Boredom1* (caused by low value) or *Boredom2* (caused by low control) according to the learner's comments during the lecture, their subjective feedback after the lecture, as well context of the judgment point (e.g., *Boredom* after *Confusion* likely belongs to *Boredom2*). Table 8 displays the results of the linear regression analysis.

Table 8. Linear regression analysis of the proportional occurrence of each affective state and the six factors.

	F1	**F2**	**F3**	**F4**	**F5**	**F6**
Engagement	-0.204 (0.839)	-1.106 (0.273)	-0.016 (0.987)	0.434 (0.665)	**3.560** **(0.001*)**	-0.289 (0.773)
Boredom	0.908 (0.368)	-0.801 (0.426)	-0.300 (0.765)	0.063 (0.950)	**-6.113** **(<0.001*)**	-0.011 (0.991)
Boredom1	**2.792** **(0.007*)**	0.132 (0.896)	-0.600 (0.551)	0.502 (0.618)	**-4.868** **(<0.001*)**	**-1.718** **(0.091)**
Boredom2	-0.840 (0.404)	-1.104 (0.274)	0.287 (0.775)	-0.251 (0.803)	**-3.159** **(0.003*)**	1.4336 (0.157)
Confusion	**-2.150** **(0.036*)**	0.581 (0.564)	0.178 (0.859)	**-2.156** **(0.035*)**	1.396 (0.168)	**2.723** **(0.009*)**
Frustration	-0.573 (0.569)	**1.962** **(0.055)**	-1.00 (0.319)	0.408 (0.685)	-0.675 (0.502)	**1.883** **(0.065)**
Curiosity	**-2.038** **(0.046*)**	0.533 0.596	**2.054** **(0.045*)**	-0.841 0.404	1.304 0.197	**-2.820** **(0.007*)**

I have some interesting findings. For example, both *Confusion* and *Frustration* had a strong positive correlation with the perceived difficulty of the learning material. However, *Confusion* was negatively correlated with the knowledge base and the perceived learning capability. This suggested that when the learner did not have necessary background knowledge or did not possess the learning skills to understand the lecture, she likely became confused. On the other hand, Frustration had a strong negative correlation with interest, suggesting that when the learner saw the value of the lecture, but could not understand it due to its difficulty, she tended to feel frustrated because she was unable to master the knowledge which was valuable for her.

The two types of *Boredom* also had different relationships with the six factors. While both types of *Boredom* had a strong negative correlation with the perceived interestingness of the learning material, *Boredom1* was negatively correlated with the learning material difficulty, while *Boredom2* had a positive relationship. This confirmed my previous finding that both low mentally demanding (low difficulty) and high mentally demanding (high difficulty) could induce *Boredom*. Also, *Boredom1* was positively correlated with the knowledge base which suggested that learners were more likely to get bored if they were already familiar with the topic.

8.7 IMPLICATIONS

8.7.1 Providing materials with the right difficulty level

It is important to provide materials for a learner with the appropriate difficulty level. As I have discussed in the *Engagement* → *Boredom* transition, 51.7% of the transition occurred when the learning material had low perceived value. 41.4% of the transition occurred when the learner had low control over the learning activity. When mental demands are too low, there is an insufficient challenge and a lack of intrinsic value, thus producing *Boredom*. Conversely, when demands exceed capabilities and cannot be met, it may be difficult to detect meaning in the activity, thus reducing its value. Therefore, *Boredom* can be experienced when learning materials are too easy (low mental demands) or too hard (high mental demand).

As I have discussed earlier, the learner's knowledge base and learning capabilities could greatly impact her mental load as well as her overall affective experience while learning. Therefore, it helps the learning process to assess the student's learning capabilities and pre-course knowledge base before a lecture to provide materials accordingly. For example, the MOOC course could provide some optional introductory videos for learners lacking the necessary background and skills to take the course.

8.7.2 Distinguishing the types of boredom detected

Because *Boredom* is found to be associated with poor learning, researchers have explored the use of physiological signals, such as EEG signals and eye gaze data, to detect *Boredom/Disengagement* in educational systems [42, 116]. However, none of these work detects the types of *Boredom,* which can be triggered when one's mental demand is either high or low.

Affect detection in a MOOC environment should distinguish the types of *Boredom* detected because different interventions should be deployed to handle these two types of *Boredom*. To diminish the *Boredom* caused by easy activities, the system should introduce more challenging content. For example, the system could skip the more basic explanatory video sections and jump to the important part. The system could also use interventions such as in-video quizzes to challenge learners and improve their engagement. On the other hand, for *Boredom* caused by high mental load, the system should increase the learner's understanding of the lecture. It could slow down the playback speed, rewind the video to go over the important parts, or provide adaptive reviews after the lecture.

8.7.3 Addressing learners' confusion state

D'Mello's affect dynamics model [33] posits the central role of *Confusion* and cognitive disequilibrium in deep learning. A major assertion that emerges from the model is that the *Engagement/Flow* ←→ *Confusion* oscillations are beneficial to learning, and the system can introduce *Confusion* to place learners in a state of cognitive disequilibrium in which they will have to stop, think, reason, and be active problem solvers. Unlike complex learning, MOOC learners are unlikely to engage in this productive deep thinking process because: 1) As opposed to deep learning activities, the MOOC activity mainly requires learners to memorize key phrases and facts in videos; 2) Due to the constant information flow of MOOC videos, learners are unlikely to spend much time thinking about the concepts/questions presented in the video. Therefore, is *Confusion* still a desirable state in MOOCs, and should the system even purposely induce *Confusion*?

My analysis of the transition from *Confusion* (cognitive disequilibrium) to *Engagement* (cognitive equilibrium) suggests that this transition relies more on the information provided by the video than on the effortful reasoning and problem-solving by the learner. Because this transition relies heavily on the content and flow of the video, which might vary greatly among different videos for different learners, I did not observe a significant *Confusion* → *Engagement* transition in the MOOC learning session; rather I observed that a confused learner is most likely to stay confused. As *Confusion* accumulates, it is likely that the learner will eventually become frustrated when she can no longer follow the lecture. Apart from the persistence of *Confusion* and the *Confusion* → *Frustration* transition, I also observed a possible fast transition from *Confusion* to *Boredom*. As previous work suggests [49], some students "*may not have enough skills and knowledge of math to experience much confusion other than an initial bewilderment* *and quick escape."* If a student believes that the learning material exceeds her aptitude and her effort will not help to master the material, she could quickly become bored and give up.

From the above discussion, it can be seen that *Confusion* in MOOC contexts is unlikely to elicit deep inquiry and learning gains; rather, it could easily lead to *Disengagement* and *Frustration* if the challenge is too difficult for the learner. Therefore, rather than intentionally causing *Confusion*, we should make the videos clear and easy to understand for general learners. Based on participants' subjective feedback on what lead to confusion, I summarized a list of rules to help instructors produce better MOOC videos that eliminate unnecessary *Confusion* of learners:

- Avoiding making assertions or claims without any explanation or support.
- Adding necessary transitions between two topics (for example, explain why introducing the new topic and its relevance to the previous one).
- When presenting a concept involving much information (e.g., proving a mathematics equation), slow down the instructional pace. Avoiding putting information on a single slide all at once, clearly explaining the concept step by step.
- Using diagrams and demonstrations might help the learner understand a concept.
- Using standard notations and symbols.
- Avoiding sloppy and unrecognizable handwriting.

8.7.4 Curiosity leading to better engagement

I found that the transition from *Curiosity* to *Engagement* occurred quite often. Thus, one effective way to improve student engagement is to make them feel curious about the upcoming content. I had discussed earlier how learners experienced *Curiosity* (0), the most common answer was, "*when the speaker posted a question and I was interested to hear what the answer was*" [S4]. Therefore, in-video quizzes, which are adopted by many MOOC learning platforms (e.g., Coursera, Udacity), are indeed effective for improving learner engagement. Some participants also indicated that statements such as "*we will address this issue later*", "*we will now talk about this problem in detail*" pique their interests. Therefore, instructors could use these prompts now and then to draw learners' attention.

8.8 PPG SIGNALS AND MOMENT-TO-MOMENT AFFECTIVE STATES

I also explored the feasibility of using the PPG signals collected in the study to predict learners' moment-to-moment affective states. This is different from our previous cognitive state detection tasks (presented in Chapter 5, 6, 7), which dealt with predicting learners' cognitive states over a period (e.g., within each learning topic). The moment-to-moment affective state detection is more difficult as learners' affect typically fluctuate dynamically and could change rapidly.

8.8.1 Signal Quality

I first analyzed the quality of implicitly captured PPG signals by investigating the RR-intervals in 5-second moving windows (measurement presented in Chapter 4.3.3.2). I only collected partial PPG data from *S1* because he accidentally closed the LivePulse application during the study. Of the remaining 21 participants, I found that most participants had good quality PPG signals except for *S4*. The PPG data from *S4* was of extremely low quality (12.9%), suggesting that *S4* did not cover the camera lens fully during the study. For the remaining 20 participants, an average of 89.10% of the PPG signals collected had high quality. 85.7% of the 60 video sessions (20 subjects X 3 videos) were in high quality. Figure 47 shows illustrations of the quality of the PPG signals, captured using the LivePulse application from 6 participants (signal quality: 95.8%, 96.3%, 95.3%, 82.7%, 84.64%, 74.9% respectively).

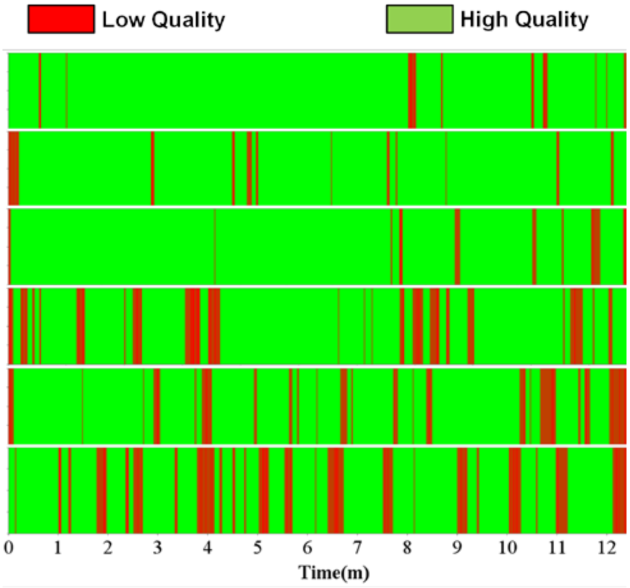


Figure 47. Quality illustration of the PPG signals from six participants while they were watching the third video clip.

Figure 48 shows a comparison of the HRV spectrograms (normalized amplitude) when a learner is in a general *Engagement/Neutral* state vs. deep *Confusion* state during learning in a one-minute time segment. The HRV spectrograms were computed by calculating the power spectral density from the RR intervals. Each topic used a one-minute sliding window with one-second increments to configure power spectral density. I found that in general there was less high-frequency power (0.15 ~ 0.4 Hz) in the HRV spectra when the learner was in a deep *Confusion* state, suggesting that the learner was under a higher cognitive workload.

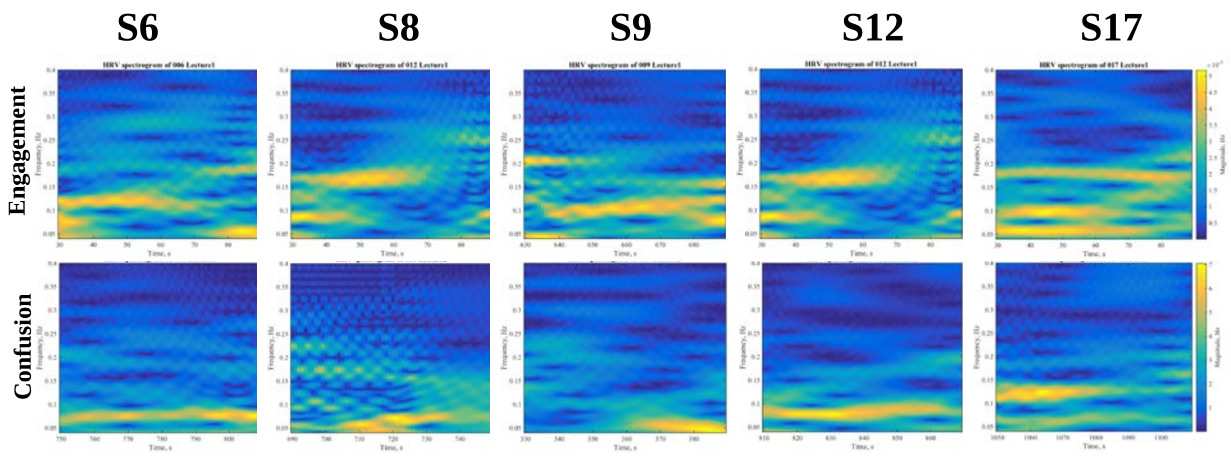


Figure 48. Heart rate variability spectrogram (normalized amplitude) of five participants (S6, S8, S9, S12, S17) when they were in a neutral/engagement state (top row) and when they were in a deep confusion state (bottom row).

8.8.2 Moment-to-Moment Affect Detection

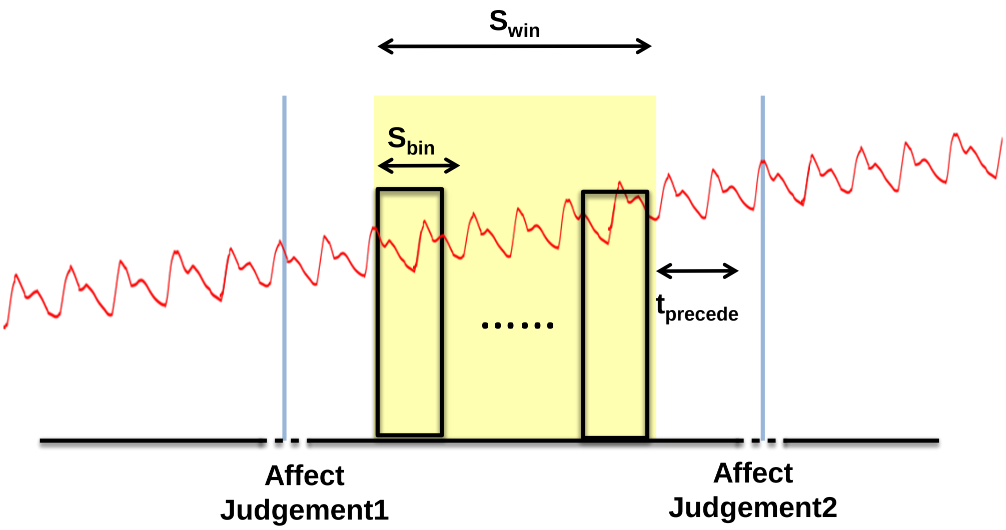


Figure 49. Extracting features for each affect judgment point.

I did not consider S1 and S4 in the following analysis due to the missing and highly corrupted PPG signals. For the collected PPG signals of the remaining 20 subjects, I first used LivePulse to extract the RR-intervals. The RR signals were smoothed to reduce noises. *Heart rate variability (HRV) features* were extracted from the PPG signal segment right before each affect judgment point (Figure 49) and used to detect learners' affective states at that judgment point. 11 dimensions of HRV features were extracted from a context window in the PPG segment: 1) AVNN; 2) SDNN; 3) rMSSD; 4-7) pNN5, pNN10, pNN20, pNN50; 8) MAD; 9) SDANN; 10) SDNNIDX; and 11) rMSSD/SDNNIDX. For each participant, all features were rescaled to [0,1] to eliminate individual and dimensional variance. I explored three parameters when extracting the HRV features: the size of the context window (20s, 30s, 40s), the preceding time offset (0s, 3s, 5s, 10s), and the size of the bin (3s, 5s, 10s).

To have a long enough PPG sequence to make a prediction, I removed those affect judgment points which were close to the previous affect judgment point (interval < 30s), leading to a total of 33 affect judgment points for each participant. Also, if the window used for extracting HRV features of the current affect judgment point reached and exceeded the previous affect judgment point, then only signals after the previous affective judgment point were used to ensure that there were no interferences between the PPG sequences of two consecutive affect judgment points. Finally, I removed those affect judgment points which had low-quality (<50%) PPG sequences. After these operations, the data set contained 643 entries for training and testing the affective state classifiers.

Using self-reported affect judgments as the gold standard, I performed the following detection tasks: 1) Task 1: detecting whether the learner is in *Engagement*, *Boredom*, or *Confusion* state (yes or no, binary classification); 2) Task 2: detecting whether the learner is in a negative state; and 3) Task 3: detecting the occurrence of critical events.

I performed detection Task 1 because these three states are the most important states that occur during MOOC learning. When building classifiers for a certain affective state (e.g., *Boredom*), I excluded participants who did not report experiencing that affective state from the dataset. Therefore, the final dataset contained 643 entries for the *Engagement* prediction task (20 participants, 32.97% *Engagement* state); 577 entries for the Boredom prediction task (18 participants, 17.33% *Boredom* states); and 544 entries for the Confusion prediction task (17 participants, 18.75% *Confusion* states.).

For Task 2, valence ratings of each judgment were used to identify negative states. For each participant, based on the valence ratings, I marked each affect judgment point either as positive (rating >= 3) or negative (rating <= 2). I excluded participants who did not report experiencing any positive or negative states. The final dataset consisted of 511 entries (16 participants, 15.46% negative states).

For Task 3, critical events were identified using the arousal ratings of each affect judgment. For each participant, based on the arousal rating, I marked each affect judgment point either as critical (rating >= 4) or not (rating <= 3). I excluded participants who did not report any critical events. The final dataset had 478 entries (15 participants, 14.22% critical events).

I used the Support Vector Machine (SVM) with a radial basis function (RBF) to build the classifiers. I built both *user-independent* models and *user-dependent* models. The *leave-one-subject-out* evaluation was used to evaluate the user-independent models. User-dependent models were built for each participant and evaluated with *10-fold cross-validations*. Table 9 lists Kappa's best performance for each classification task.

The Kappa score indicated a clear relationship between learners' affective states and their PPG signals. I achieved the best performance predicting the critical or high arousal events when participants had stronger emotions. This is expected, as one might assume that stronger emotions will also lead to stronger changes in physiological responses. Moreover, *Engagement* prediction was more accurate than *Boredom* and *Confusion* prediction. As I have discussed, *Boredom* and *Confusion* might co-exist, which could affect the prediction performance for *Boredom* and *Confusion*. Also, the performance of user dependent models are much better than the user independent models. The user dependent models are more accurate because there might exist significant differences among participants in the PPG signal as well as the perception of affective states in learning.

Table 9. The performance of different moment-to-moment affective state prediction tasks.

Detection	**User-independent**		**User dependent**	
	Acc.	Kappa	Acc.	Kappa
Engagement	70.75%	0.1512	61.96%	0.2770
Boredom	83.57%	0.0766	83.71%	0.1387
Confusion	80.09%	0.0701	83.72%	0.2054
Negative Events (low valence)	85.46%	0.1071	84.97%	0.1815
Critical Events (high arousal)	84.80%	0.2332	84.60%	0.2854

Compared with the cognitive state prediction tasks presented in Chapter 5 and 6, the classifiers in the current tasks had worse performance, suggesting that it is indeed more difficult to predict moment-to-moment affective states as opposed to predicting the general cognitive states over a period. The worse performance might be due to two reasons: 1) In the current prediction tasks, the PPG signal sequence of each sample was much shorter than the ones in the previous tasks (~20s vs. > 1min), indicating less information and more noises in the data; 2)

Participants only reported one dominant affect at each affect judgement point. The coexistence of affect at each point was not considered and manifested in the label of the samples. However, it is worth highlighting that our performance was achieved on current mobile phones without any hardware modifications. I also only used the PPG signals and did not use any other contextual features. To the best of my knowledge, this is the first work to investigate the prediction of moment-to-moment affective states in the MOOC contexts.

8.9 DISCUSSIONS AND SUMMARY

One limitation of this research is that the dynamics model I presented for MOOC learning is essentially a first-order Markov model showing the links between primary affective states. However, a more detailed model is possible if hidden states are considered. Although the feedback from participants suggests that both the video content and individual differences could affect how learners experience and regulate their affective states, the current model does not consider these possible hidden states. To improve the current model, a new study can be conducted to gather more information about the video content and the learner at each affect judgment point.

Another limitation pertains to the generalizability of the findings. The proposed model was tested on 22 students from various backgrounds who were on the same course. This raised the question of whether the model would be supported by a larger population of students using a different course. Considering that there are other alternative methods to track learner emotions (e.g., observations by external judges [14, 36, 103]). In the future, the models should be tested on a larger population using different courses and alternate affective state tracking methods.

To sum up, in this chapter, I present a 22-participant study to investigate the dynamic temporal transitions of affective states during a MOOC learning session. I discuss differences of MOOC learning and complex learning from the perspective of affect dynamics and present pedagogical implications of the results. I also showed that PPG signals implicitly captured by the built-in cameras on unmodified mobile phones can be used to detect moment-to-moment affective states, especially the high arousal and critical events. This research promotes a better understanding of the dynamic MOOC learning process.

9.0 CONCLUSIONS

9.1 CONTRIBUTIONS

This dissertation presents AttentiveLearner, a mobile learning system which captures and uses learners' physiological signals *implicitly* during mobile MOOC learning *without* leveraging any dedicated sensors. I explored the use of physiological signals collected from learners via AttentiveLearner to understand, model, and improve learning in mobile MOOC contexts. To support this dissertation, I had a list of hypotheses (Section 1.4) and these hypotheses are supported by the following specific contributions:

Chapter 3 presented the design and evaluation of the tangible video control channel of AttentiveLearner. A general on-lens finger gesture based interaction technique, *LensGesture*, was first proposed. I then optimized the Static LensGesture specifically to meet the unique requirement of AttentiveLearner. Through off-line benchmarking and an 18-subject user study, I verified that using the on-lens finger-covering gesture to control video play was both accurate and responsive (Hypothesis A.1). Moreover, I systematically investigated various usability concerns regarding the new video control channel, and showed that the tangible video control interface in AttentiveLearner was user-friendly (Hypothesis A.2).

Chapter 4 first demonstrated *LivePulse*, a real-time heart rate measurement algorithm based on commodity-camera-based PPG sensing. A *12-participant* user study verified the feasibility of using LivePulse to collect PPG signals and measure heart rate (Hypothesis B.1). I conducted another *18-participant* user study to investigate the usability of AttentiveLearner during actual MOOC learning sessions. Results of the study suggested that AttentiveLearner could collect high quality PPG signals reliably from the learner during actual MOOC sessions (Hypothesis B.2).

In **Chapter 5** and **Chapter 6**, I showed that the PPG signals, implicitly recorded by the built-in camera of mobile phones, can be used to infer learners' interests and perceived confusion levels towards the learning topics (**Chapter 5**), as well as their divided attentional state (**Chapter 6**) during learning. Two 18-participant user studies were conducted to collect data and build the cognitive state prediction models. The models achieved comparable performance as existing research that used dedicated physiological sensors to detect human cognitive states. Results of these two chapters proved Hypothesis C.1.

In **Chapter 7**, I demonstrated an intervention technique, Context and Cognitive State triggered Feed-Forward (*C2F2*), which proactively reminded a learner of important upcoming content when learner disengagement was detected. I performed a *48-participant* user study to evaluate effectiveness of C2F2 and found that C2F2 could improve both information recall and learning gains for bottom performers when compared with a non-interactive system (Hypothesis D.2). The effectiveness of C2F2 demonstrated the feasibility and efficacy of building end-to-end, affect-aware mobile MOOC systems on top of AttentiveLearner to benefit MOOC learners (Hypothesis D.1).

Finally, in **Chapter 8**, I investigated the dynamics of moment-to-moment affective states in MOOC contexts through a 22-participant user study. I identified several different affective state transition patterns between MOOC learning and complex learning, which verified

Hypothesis E.1. Using the PPG signals collected in the study, I explored the feasibility of moment-to-moment affective state predictions. Performance of the classifiers indicated a clear relationship between learners' affective states and their PPG signals, especially for the high arousal events. This finding verified Hypothesis E.2. Moreover, important pedagogical implications were discussed in this chapter.

Recall that this dissertation has the following thesis statement:

By proposing a "sensorless" approach to collect photoplethysmography (PPG) signals implicitly from users on unmodified mobile devices, this dissertation explores novel technologies to monitor learners' cognitive and affective states, and provide cognitive state triggered adaptive interventions, which can effectively improve learning in mobile MOOC contexts.

Chapter 3 and 4 demonstrated the feasibility of AttentiveLearner, the "sensorless" approach, to implicitly capture learners' physiological signals on unmodified mobile phones without any extra sensor. In Chapter 5, 6 and 8, I showed that learners' cognitive and affective states can be inferred by analyzing the PPG signals collected by AttentiveLearner. These three chapters verified that AttentiveLearner can promote a deeper understanding of learners' cognitive and affective states in MOOC contexts. In Chapter 7, I proved that affect-aware interventions (e.g., C2F2) can be built on top of AttentiveLearner and improve learning performance. Chapter 5, 6, 8 showed that from the instructor's perspective, AttentiveLearner can provide a rich, fine-grained feedback channel for them to understand the learners. Chapter 7 showed that from the learner's perspective, AttentiveLearner can directly benefit learners by providing affect-aware adaptive interventions. In all, I showed that AttentiveLearner can help us improve learning in the mobile MOOC contexts.

To achieve the goal of this dissertation, I also contributed with a series of user studies. A summary of these user studies is presented in Appendix C.

9.2 FUTURE WORK

My current research has built the foundation of AttentiveLearner as a rich and effective feedback channel to understand, model, and improve learning in mobile MOOC contexts. I believe that there are many research directions that can be pursued to improve and utilize AttentiveLearner. First, my colleagues plan to conduct large-scale, longitudinal studies on AttentiveLearner in learners' everyday environments soon. Second, security and privacy issues arise when learners' physiological signals are transmitted, stored, and visualized on the server-side. It is important to explore security algorithms and policies to provide rich feedback without disclosing unnecessary privacy from learners. Third, although my current cognitive state prediction models are user-independent, the PPG signals were collected when learners were learning the same material. This could impact the scalability of my method since the learning materials vary across different courses. A *course-independent* model is necessary for the robust detection of affective states and wide adoption of AttentiveLearner. Fourth, my current research has only scratched the surface of using physiological signals to understand and improve MOOC learning. There are other physiological signal modalities we can explore. For example, the front camera of smartphones can be used as a facial expression sensor and provide new complementary signals to understand and improve MOOC learning. I expect that different physiological signal channels might be effective under different circumstances. It is interesting to explore "*for who, under what circumstances, what signal channel is effective*". Finally, currently, the C2F2 intervention is based on PPG signals alone. It will be interesting to explore intervention techniques from multiple modalities (e.g., PPG, EEG, facial expressions, and eye gaze), especially investigating the interactions among multiple signal sources so as to identify the optimal timing to trigger interventions.

AttentiveLearner uses the mobile device to monitor physiological signals, but there are many new opportunities of measuring and using physiological signals beyond the mobile platform. For example, recent work has shown that physiological measurement of heart rate, breath rate, and HRV can be accurately captured remotely (3 meters away from the camera), via photoplethysmography using a low-cost digital camera [77, 79]. Digital cameras could also capture the user's facial features and behaviors to infer her affective and cognitive state [44, 78, 101]. Furthermore, it is reported that wireless signals could be used to monitor the vital signs (breathing and heart rate) of multiple people without body contact [1]. Moreover, with the rapid development of wearable technology, many wearable devices, such as smartwatches, fitness wristbands, and headphones, can monitor users' heart rate and other physiological signals. Therefore, there are many opportunities for us to collect physiological signals *implicitly* without purchasing dedicated physiological sensors. These multiple channels of signals from different sources could complement each other and provide better classification accuracies in detecting various affective-cognitive states and learning events.

I have already integrated AttentiveLearner into the OpenEdx, an open-source MOOC platform that powered edx.org, one of the major MOOC providers nowadays. The AttentiveLearner mobile client should be available for everyone to download at http://www.attentivelearner.com soon. One question I cannot resist asking myself is, "*What if millions of people use AttentiveLearner?*" With millions of users, I would obtain a great deal of relevant and useful information to understand MOOC learners and improve the performance of the cognitive state predictions. For example, I would gain a better understanding of which HRV features were *user-independent* and which features were *course-independent*. Moreover, I would be able to investigate the impact of different environments (e.g., classrooms vs. public transit, sitting vs. walking) on learners' PPG signals. This could help me exclude confounding effects of the environments on PPG signals and get more reliable cognitive state information. Finally, I could explore how different techniques of learning analytics (e.g., log analysis, post-lecture feedback, and AttentiveLearner) can be integrated to provide informative feedback to both instructors and students.

APPENDIX A

THE CAUSAL ATTRIBUTIONS OF AFFECTIVE STATES IN THE VIDEO-BASED MOOC LEARNING ACTIVITY

Table 10. The causal attributions of affective states in MOOC learning

Affect	Internal Attributions (learner)	External Attributions (video)
Engagement	1. Interested in the video content. 2. Able to understand and comprehend the video content well.	New or interesting concept/topic is introduced and is presented clearly.
	S6: "When content, although very unfamiliar, seemed simple enough for me to comprehend and follow or accessible." (2) *S13: "When I was first starting out, I was interested in learning a new topic." (1)*	*S19: "Material was interesting and presented clearly."* *S20: "When the lecture is clear and in a good pace to catch up with."*
Boredom	1. Disinterested in the video content. 2. Unable to keep up with the lecture. 3. Physical or mental tiredness.	1. The content is easy/obvious, redundant, trivial or unrelated to the main topic. 2. Too much information is presented at the same time and/or presented at a fast pace so that the learner cannot process or remember all information.

	S1: "Anything math related. I don't find this type of field interesting so it's difficult to pay attention." (1) *S6: "When the content seemed way over my head and the lecture was too difficult to follow (most of section 3)." (2)* *S17: "When I am tired and the material isn't all that interesting." (1, 3)*	*S11: "When the lecture seems to be straying from its goals." (1)* *S20: "When the explanation was too wordy, the definition was unclear, or the instructor introduced theories without talking about why we need them". (1, 2)* *S22: "Number/Equation crunching." (2)* *S9: "Too much mathematic talking." (2)*
Confusion	1. Do not understand a statement, symbol or notation. The learner may lack prerequisite knowledge for the course in the first place. 2. Encounter a question, either raised explicitly by the instructor (i.e., in-video quiz) or came up with by the learner.	1. Instructions are fast, unclear, and/or missing important information (e.g., making an assertion without any support, no explanation for symbols, etc.). 2. Sudden switch to a new concept without necessary transitions.
	S23: "I felt confused when I didn't understand the overall point." (1) *S6: "When I got lost during the lectures, especially because the content was extremely unfamiliar." (1)* *S11: "When I might have missed a point or not connected the ideas, or had a question." (1, 2)*	*S18: "Sometimes, lecturer would explain a concept, then immediately move on without letting the concept sink in." (1, 2)* *S16: "Felt confused when there were a lot of symbol being used at the same time; couldn't understand what they all meant." (1)* *S17: "When they explain too quickly. I would usually have gone back in the video and rewatched" (1)*
Frustration	1. Unable to achieve the learning goal. This happens when the learner feels extremely confused and cannot follow up the video. 2. Unsatisfied about the actual details of something in the lecture (e.g., the explanation of a theory etc.).	1. Instructions are too fast to allow the learner to process and understand the lecture. 2. Poor video production: crappy hand-writing, lags in the video-over, etc.
	S7: "My expectations were not met as I didn't understand (the lecture), and I became frustrated." (1)	*S21: "Annoyed by choppy voice or writing." (2)* *S3: "Some concepts were brushed*

	S14: "When I was feeling confused for 3 to 5 minutes and felt no sense of achievement." (1) *S9: "The second video while trying to learn the symbols and notations used. I was constantly confused which affected my focus. I became frustrated." (1)* *S17: "When things aren't explained to my satisfaction." (2)*	*past and I felt frustrated when I couldn't think through the material at my own pace." (1)* *S5: "When they went through things too fast that I could not follow." (1)*
Curiosity	Interested to know answers to a question or know more about an idea or concept.	The instructor asks a question and seeks answer
	S21: "Want to know the answer or why something is the way it is." *S19: "When I was unsure of an answer or unsure how a problem was going to be addressed. When the instructor said 'We'll address this later'."* *S3: "Sometimes I was intrigued and wanted to know more about an idea or concept generally."*	*S4: "The speaker posted a question and I was interested to hear what the answer was."* *S18: "Every now and then, they'd ask questions that I was curious to know the answer to."*
Surprise	Something unexpected happened or is presented in the video.	
	S10: "When something new I learned was something I couldn't predict/hypothesize." *S18: "At one point, the lecturer asked a question that I thought I got right, but it turned out to be wrong."*	
Happiness	The learner achieved something and feel accomplished (e.g., correctly answer a question, fully understand a concept, etc.)	
	S21: "Got answer correct, or understand (the content), feeling encouraged." *S14: "Once I felt I've got the idea or my confusion was gone."*	
Delight	The learner answered a difficult question correctly and completed a section	
	S2: "Answered questions correctly (not the trivial ones)." *S11: "Usually when some point has been completed and I feel like the lecture helped me understand."*	

APPENDIX B

QUESTIONNAIRES USED IN THE USER STUDIES

B.1 LENSGESTURE EVALUATION QUESTIONNAIRE

1. I find the LensGesture input method useful.

1 – Strongly Disagree 2 – Disagree 3 – Neutral 4 – Agree 5 – Strongly Agree

2. I find LensGesture useful as a Main Input channel.

1 – Strongly Disagree 2 – Disagree 3 – Neutral 4 – Agree 5 – Strongly Agree

3. I find LensGesture useful as an Assistant Input channel.

1 – Strongly Disagree 2 – Disagree 3 – Neutral 4 – Agree 5 – Strongly Agree

4. I find LensGesture accurate

1 – Strongly Disagree 2 – Disagree 3 – Neutral 4 – Agree 5 – Strongly Agree

5. I find LensGesture responsive

1 – Strongly Disagree 2 – Disagree 3 – Neutral 4 – Agree 5 – Strongly Agree

6. I find LensGesture easy to perform

1 – Strongly Disagree2 – Disagree 3 – Neutral 4 – Agree 5 – Strongly Agree

7. I will use LensGesture input method in the future on my mobile phone

1 – Strongly Disagree 2 – Disagree 3 – Neutral 4 – Agree 5 – Strongly Agree

8. What do you like or dislike about LensGesture?

9. Do you have any suggestions for us to improve the LensGesture?

B.2 ATTENTIVELEARNR EVALUATION QUESTIONNAIRE

1. On a scale of 1-5, how do you like the AttentiveLearner mobile application in general?

1 – Strongly Dislike 2 – Dislike 3 – Neutral 4 – Like 5 – Strongly Like

2. I find it comfortable to use the AttentiveLearner mobile application to consume lecture videos.

1 – Strongly Disagree 2 – Disagree 3 – Neutral 4 – Agree 5 – Strongly Agree

3. I find the AttentiveLearner mobile application useful for consuming lecture videos.

1 – Strongly Disagree 2 – Disagree 3 – Neutral 4 – Agree 5 – Strongly Agree

4. I find the video control channel (cover the lens to play video, uncover the lens to pause video) easy to operate.

1 – Strongly Disagree 2 – Disagree 3 – Neutral 4 – Agree 5 – Strongly Agree

5. I find the video control channel intuitive.

1 – Strongly Disagree 2 – Disagree 3 – Neutral 4 – Agree 5 – Strongly Agree

6. I find the video control channel responsive.

1 – Strongly Disagree 2 – Disagree 3 – Neutral 4 – Agree 5 – Strongly Agree

7. I am interested in using the AttentiveLearner mobile application in the future to learner more courses.

1 – Strongly Disagree2 – Disagree 3 – Neutral 4 – Agree 5 – Strongly Agree

8. What do you like or dislike about the video control channel (cover the lens to play video, uncover the lens to pause video)?

9. Do you feel any fatigue/uncomfortable during the video watching session? If so, when do you feel tired/uncomfortable? what's the optimal duration of a video lecture should be?

10. What do you like or dislike about the AttentiveLearner mobile application?

11. Do you have any suggestions for us to make the application better?

B.3 C2F2 EVALUATION QUESTIONNAIRE

1. On a scale of 1-5, how do you like the mobile MOOC application you used in general?

1 – Strongly Dislike 2 – Dislike 3 – Neutral 4 – Like 5 – Strongly Like

2. I am happy to download the application and use it to take lessons on my mobile phone:

1 – Strongly Disagree 2 – Disagree 3 – Neutral 4 – Agree 5 – Strongly Agree

If you have NOT seen a feed-forward alert (the "Pay Attention" warning before a topic), please skip the following questions

3. On a scale of 1-5, do you find the feed-forward alert helpful?

1 – Strongly Disagree 2 – Disagree 3 – Neutral 4 – Agree 5 – Strongly Agree

4. On a scale of 1-5, do you find the feed-forward alert distractive?

1 – Strongly Disagree 2 – Disagree 3 – Neutral 4 – Agree 5 – Strongly Agree

5. On a scale of 1-5, do you find the feed-forward alert triggered at the right time?

1 – Strongly Disagree 2 – Disagree 3 – Neutral 4 – Agree 5 – Strongly Agree

6. What do you like or dislike about the feed-forward alert.

7. Do you have any suggestions to improve the feed-forward alert?

B.4 SUBJECTIVE IMPRESSION QUESTIONNAIRE

I paid attention in this learning session

|-----------------|-------------------|-------------------|-----------------|

Strongly Disagree Disagree Neutral Agree Strongly Agree

I was engaged during this learning session

|-----------------|-------------------|-------------------|-----------------|

Strongly Disagree Disagree Neutral Agree Strongly Agree

I felt bored during this learning session

|-----------------|-------------------|-------------------|-----------------|

Strongly Disagree Disagree Neutral Agree Strongly Agree

I found the lecture hard to follow in this learning session

|-----------------|------------------|------------------|-----------------|

Strongly Disagree Disagree Neutral Agree Strongly Agree

I put effort into this learning session

|-----------------|------------------|------------------|-----------------|

Strongly Disagree Disagree Neutral Agree Strongly Agree

I think I learned the lecture well

|-----------------|------------------|------------------|-----------------|

Strongly Disagree Disagree Neutral Agree Strongly Agree

APPENDIX C

A SUMMARY OF USER STUDIES

Table 11. A summary of all user studies presented in this dissertation

Chapter	Study Description	No. of Subjects	Type	Major Result
3	Static LensGestures sample image collection	9	Data Collection	791 sample images collected; gesture detection accuracy: 97.9% (full-covering gesture), 93.2% (partial-covering gesture)
	Dynamic LensGestures sample image collection	12	Data Collection	957 sets of sample gestures collected; gesture detection accuracy: 91.3%
	LensGesture usability evaluation	16	System Evaluation	1. LensGesture response time: 789 ~ 1815ms 2. Target acquisition followed Fitt's Law 3. Speed of LensGesture-enabled keyboard vs. standard keyboard: 13.4 vs. 11.7 wps
	Static LensGesture sample image collection for AttentiveLearner	10	Data Collection	483 sample images collected; accuracy: 99.59%

	Tangible video control channel response time evaluation	18	System Evaluation	Tangible video control vs. traditional control: 625.9 vs. 426.6 ms.
4	LivePulse usability evaluation	12	System Evaluation	1. Raw PPG signals from LivePulse and the pulse oximeter were highly consistent 2. MER of heart rate estimation: 3.9%
	AttentiveLearner usability evaluation in actual MOOC learning; PPG signal collection for boredom/confusion detection	18	System Evaluation/ Data Collection	1. Average user experience ratings: 4.11 2. Collected PPG signals were reliable: 88.9% sessions, more than 80% signals were of high quality 3. Boring topic prediction performance: 0.297 Kappa; confusing topic prediction: 0.269 Kappa
6	PPG signal collection for divided attention detection	18	Data Collection	1. Learners' average performance in FA, EDA, LIDA, HIDA: 4.11, 3.83, 3.33 and 2.92 2. Accuracy of attentional state prediction (user-dependent): 72.74% ~ 88.54%
7	Evaluation of C2F2 against other feed-forward intervention strategies (no feed-forward, context-only feed-forward, cognitive-only feed-forward)	48	System Evaluation	*C2F2* were effective for bottom performers. Average *Learning Gains* was 43.75%, 44.45%, 52.74% and 61.94% in the four conditions respectively. *C2F2* vs. *no feed-forward* condition (t = 0.1829, p = 0.0018); C2F2 vs. *context only feed-forward* (t = 0.1749, p = 0.0025)
8	Investigation of the affect dynamics in	22	Data Collection	1. A set of common affective states and transitions of affective

	mobile MOOC contexts; PPG signal collection for moment-to-moment affective states detection			states in MOOC contexts were identified 2. Performance of moment-to-moment affect detection: 0.07 ~ 0.23 Kappa (user-independent); 0.14 ~ 0.29 Kappa (user-dependent) 3. Implications for MOOC design

BIBLIOGRAPHY

1. Adib, F., Mao, H., Kabelac, Z., Katabi, D., & Miller, R. C. (2015, April). Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 837-846). ACM.
2. Afzal, S., & Robinson, P. (2011). Designing for Automatic Affect Inference in Learning Environments. *Educational Technology & Society*, *14*(4), 21-34.
3. Akselrod, S., Gordon, D., Ubel, F. A., Shannon, D. C., Berger, A. C., & Cohen, R. J. (1981). Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control. *science*, *213*(4504), 220-222.
4. Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014, April). Engaging with massive online courses. In *Proceedings of the 23rd international conference on World wide web* (pp. 687-698). ACM.
5. Anitha, A. Motion Estimation Algorithm For Video Compression. *CVR JOURNAL OF SCIENCE & TECHNOLOGY*, 55.
6. Arroyo, I., Woolf, B. P., Burelson, W., Muldner, K., Rai, D., & Tai, M. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal of Artificial Intelligence in Education*, *24*(4), 387-426.
7. Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners’ cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, *68*(4), 223-241.
8. Balakrishnan, G., Durand, F., & Guttag, J. (2013). Detecting pulse from head motions in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3430-3437).
9. Banitsas, K., Pelegris, P., Orbach, T., Cavouras, D., Sidiropoulos, K., & Kostopoulos, S. (2009, November). A simple algorithm to monitor hr for real time treatment applications. In *2009 9th International Conference on Information Technology and Applications in Biomedicine* (pp. 1-5). IEEE.
10. Baudisch, P., & Chu, G. (2009, April). Back-of-device interaction allows creating very small touch devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1923-1932). ACM.
11. Bixler, R., & D’Mello, S. (2014, July). Toward fully automated person-independent detection of mind wandering. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 37-48). Springer International Publishing.

12. Bixler, R., Blanchard, N., Garrison, L., & D'Mello, S. (2015, November). Automatic Detection of Mind Wandering During Reading Using Gaze and Physiology. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (pp. 299-306). ACM.

13. Blanchard, N., Bixler, R., Joyce, T., & D’Mello, S. (2014, June). Automated physiological-based detection of mind wandering during learning. In *International Conference on Intelligent Tutoring Systems* (pp. 55-60). Springer International Publishing.

14. Bosch, N., Chen, H., D'Mello, S., Baker, R., & Shute, V. (2015, November). Accuracy vs. availability heuristic in multimodal affect detection in the wild. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (pp. 267-274). ACM.

15. Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, *25*(1), 49-59.

16. Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research & Practice in Assessment*, *8*.

17. Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on affective computing*, *1*(1), 18-37.

18. Cardiograph for iOS, https://itunes.apple.com/us/app/cardiograph/id441079429?ls=1&mt=8

19. Carroll, E. A., Czerwinski, M., Roseway, A., Kapoor, A., Johns, P., Rowan, K., & Schraefel, M. C. (2013, September). Food and mood: Just-in-time support for emotional eating. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on* (pp. 252-257). IEEE.

20. Coetzee, D., Fox, A., Hearst, M. A., & Hartmann, B. (2014, February). Should your MOOC forum use a reputation system?. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 1176-1187). ACM.

21. Coetzee, D., Fox, A., Hearst, M. A., & Hartmann, B. (2014, March). Chatrooms in MOOCs: all talk and no action. In *Proceedings of the first ACM conference on Learning@ scale conference* (pp. 127-136). ACM.

22. Coleman, C. A., Seaton, D. T., & Chuang, I. (2015, March). Probabilistic use cases: Discovering behavioral patterns for predicting certification. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale* (pp. 141-148). ACM.

23. Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, *19*(3), 267-303.

24. Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. *Journal of educational media*, *29*(3), 241-250.

25. Craik, F. I., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*(2), 159.

26. Cross, A., Bayyapunedi, M., Ravindran, D., Cutrell, E., & Thies, W. (2014, February). VidWiki: enabling the crowd to improve the legibility of online educational videos. In

Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (pp. 1167-1175). ACM.

27. Csikszentmihalyi, M., & Csikszentmihalyi, I. S. (Eds.). (1992). *Optimal experience: Psychological studies of flow in consciousness*. Cambridge university press.
28. D’Mello, S. (2012). Monitoring affective trajectories during complex learning. In *Encyclopedia of the Sciences of Learning* (pp. 2325-2328). Springer US.
29. D’Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology*, *105*(4), 1082.
30. D’Mello, S. (2016). Giving Eyesight to the Blind: Towards Attention-Aware AIED. *International Journal of Artificial Intelligence in Education, 26*(2), 645-659.
31. D’Mello, S., & Graesser, A. (2010). Modeling cognitive-affective dynamics with Hidden Markov Models. *Proceedings of the 32nd annual cognitive science society*, 2721-2726.
32. D’Mello, S., & Graesser, A. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction*, *20*(2), 147-187.
33. D’Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145-157.
34. D’Mello, S., Blanchard, N., Baker, R., Ocumpaugh, J., & Brawner, K. (2014). Affect-Sensitive Instructional Strategies. *Design Recommendations for Intelligent Tutoring Systems: Volume 2-Instructional Management*, *2*, 35.
35. D’Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., ... & Graesser, A. (2010, June). A time for emoting: When affect-sensitivity is and isn’t effective at promoting deep learning. In *International Conference on Intelligent Tutoring Systems* (pp. 245-254). Springer Berlin Heidelberg.
36. D’Mello, S., Picard, R., & Graesser, A. (2007). Towards an affect-sensitive autotutor. *IEEE Intelligent Systems*, *22*(4), 53-61.
37. Diana, N. E., & Febrian, A. (2016). Understanding Student Emotion in Video-Based Learning: Case Study-Programming Course. *Journal of Teaching and Education, 5*(01), 641-648.
38. Dillon, J., Ambrose, G. A., Wanigasekara, N., Chetlur, M., Dey, P., Sengupta, B., & D'Mello, S. K. (2016, April). Student affect during learning with a MOOC. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 528-529). ACM.
39. Dillon, J., Bosch, N., Chetlur, M., Wanigasekara, N., Ambrose, G. A., Sengupta, B., & D’Mello, S. K. Student Emotion, Co-occurrence, and Dropout in a MOOC Context.
40. D'Mello, S., & Graesser, A. (2007). Mind and body: Dialogue and posture for affect detection in learning environments. *Frontiers in Artificial Intelligence and Applications, 158,* 161.
41. D'Mello, S., & Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. *Cognition & Emotion*, *25*(7), 1299-1308.
42. D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of human-computer studies*, *70*(5), 377-398.

43. Drummond, J., & Litman, D. (2010, June). In the zone: Towards detecting student zoning out using supervised machine learning. In *International Conference on Intelligent Tutoring Systems* (pp. 306-308). Springer Berlin Heidelberg.

44. El Kaliouby, R., & Robinson, P. (2005). Real-time inference of complex mental states from facial expressions and head gestures. In *Real-time vision for human-computer interaction* (pp. 181-200). Springer US.

45. Fan, X., Luo, W., Menekse, M., Litman, D., & Wang, J. (2015, April). CourseMIRROR: Enhancing large classroom instructor-student interactions via mobile interfaces and natural language processing. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1473-1478). ACM.

46. Fowler, G. A. (2013). An early report card on massive open online courses. *The Wall Street Journal, 8*.

47. Gašević, D., Mirriahi, N., & Dawson, S. (2014, March). Analytics of the effects of video use and instruction to support reflective learning. In *Proceedings of the fourth international conference on learning analytics and Knowledge* (pp. 123-132). ACM.

48. Glassman, E. L., Kim, J., Monroy-Hernández, A., & Morris, M. R. (2015, April). Mudslide: A spatially anchored census of student confusion for online lecture videos. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1555-1564). ACM.

49. Graesser, A. C., D'Mello, S. K., & Strain, A. C. (2014). Emotions in advanced learning technologies. *International handbook of emotions in education*, 473-493.

50. Grawemeyer, B., Holmes, W., Gutiérrez-Santos, S., Hansen, A., Loibl, K., & Mavrikis, M. (2015, March). Light-bulb moment?: towards adaptive presentation of feedback based on students' affective state. In *Proceedings of the 20th International Conference on Intelligent User Interfaces* (pp. 400-404). ACM.

51. Guo, P. J., Kim, J., & Rubin, R. (2014, March). How video production affects student engagement: An empirical study of mooc videos. In *Proceedings of the first ACM conference on Learning@ scale conference* (pp. 41-50). ACM.

52. Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. (2010, September). Psycho-physiological measures for assessing cognitive load. In *Proceedings of the 12th ACM international conference on Ubiquitous computing* (pp. 301-310). ACM.

53. Han, T., Xiao, X., Shi, L., Canny, J., & Wang, J. (2015, April). Balancing accuracy and fun: Designing camera based mobile games for implicit heart rate monitoring. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 847-856). ACM.

54. Hinckley, K., & Song, H. (2011, May). Sensor synaesthesia: touch in motion, and motion in touch. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 801-810). ACM.

55. Hussain, M. S., AlZoubi, O., Calvo, R. A., & D'Mello, S. K. (2011, June). Affect detection from multichannel physiology during learning sessions with AutoTutor. In *International Conference on Artificial Intelligence in Education* (pp. 131-138). Springer Berlin Heidelberg.

56. Instant Heart Rate for iOS, https://itunes.apple.com/app/instant-heart-rate-measure/id395042892?mt=8

57. Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of memory and language*, *30*(5), 513-541.

58. Jaques, N., Conati, C., Harley, J. M., & Azevedo, R. (2014, June). Predicting affect from gaze data during interaction with an intelligent tutoring system. In *International Conference on Intelligent Tutoring Systems* (pp. 29-38). Springer International Publishing.

59. Jraidi, I., Chaouachi, M., & Frasson, C. (2013, December). A dynamic multimodal approach for assessing learners' interaction experience. In *Proceedings of the 15th ACM on International conference on multimodal interaction* (pp. 271-278). ACM.

60. Kahneman, D. (1973). *Attention and effort* (p. 246). Englewood Cliffs, NJ: Prentice-Hall.

61. Kapoor, A., & Picard, R. W. (2005, November). Multimodal affect recognition in learning environments. In *Proceedings of the 13th annual ACM international conference on Multimedia* (pp. 677-682). ACM.

62. Karsenti, T. (2013). The MOOC: what the research says. *International Journal of Technologies in Higher Education*, *10*(2), 23-37.

63. Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science*, *330*(6006), 932-932.

64. Kim, J., Glassman, E. L., Monroy-Hernández, A., & Morris, M. R. (2015, April). RIMES: Embedding interactive multimedia exercises in lecture videos. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1535-1544). ACM.

65. Kim, J., Guo, P. J., Cai, C. J., Li, S. W. D., Gajos, K. Z., & Miller, R. C. (2014, October). Data-driven interaction techniques for improving navigation of educational videos. In *Proceedings of the 27th annual ACM symposium on User interface software and technology* (pp. 563-572). ACM.

66. Kim, J., Guo, P. J., Seaton, D. T., Mitros, P., Gajos, K. Z., & Miller, R. C. (2014, March). Understanding in-video dropouts and interaction peaks in online lecture videos. In *Proceedings of the first ACM conference on Learning@ scale conference* (pp. 31-40). ACM.

67. Kizilcec, R. F., Papadopoulos, K., & Sritanyaratana, L. (2014, April). Showing face in video instruction: effects on information retention, visual attention, and affect. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 2095-2102). ACM.

68. Kop, R., & Fournier, H. (2011). New dimensions to self-directed learning in an open networked learning environment. *International Journal of Self-Directed Learning*, *7*(2), 1-18.

69. Kovacs, G. (2015, April). QuizCram: A Question-Driven Video Studying Interface. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 133-138). ACM.

70. Kovacs, G. (2016, April). Effects of In-Video Quizzes on MOOC Lecture Viewing. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale* (pp. 31-40). ACM.

71. Kumar, V. (2014, October). Enhancing video lectures with digital footnotes. In *2014 IEEE Frontiers in Education Conference (FIE) Proceedings* (pp. 1-3). IEEE.

72. Lee, Y. C., Lin, W. C., Cherng, F. Y., Wang, H. C., Sung, C. Y., & King, J. T. (2015, April). Using time-anchored peer comments to enhance social interaction in online educational

videos. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 689-698). ACM.

73. Leony, D., Muñoz-Merino, P. J., Ruipérez-Valiente, J. A., Pardo, A., & Kloos, C. D. (2015). Detection and evaluation of emotions in massive open online courses. *Journal of Universal Computer Science*, *21*(5), 638-655.
74. Lyu, Y., Luo, X., Zhou, J., Yu, C., Miao, C., Wang, T., Shi, Y. & Kameyama, K. I. (2015, April). Measuring photoplethysmogram-based stress-induced vascular response index to assess cognitive load and stress. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 857-866). ACM.
75. Mark, G., Wang, Y., & Niiya, M. (2014, April). Stress and multitasking in everyday college life: an empirical study of online activity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 41-50). ACM.
76. Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: the default network and stimulus-independent thought. *Science*, *315*(5810), 393-395.
77. McDuff, D. J., Hernandez, J., Gontarek, S., & Picard, R. W. (2016, May). COGCAM: Contact-free Measurement of Cognitive Stress During Computer Tasks with a Digital Camera. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 4000-4004). ACM.
78. McDuff, D., El Kaliouby, R., Demirdjian, D., & Picard, R. (2013, April). Predicting online media effectiveness based on smile responses gathered over the internet. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on* (pp. 1-7). IEEE.
79. McDuff, D., Gontarek, S., & Picard, R. W. (2014). Improvements in remote cardiopulmonary measurement using a five band digital camera. *IEEE Transactions on Biomedical Engineering*, *61*(10), 2593-2601.
80. Mills, C., D'Mello, S., Bosch, N., & Olney, A. M. (2015, June). Mind Wandering During Learning with an Intelligent Tutoring System. In *International Conference on Artificial Intelligence in Education* (pp. 267-276). Springer International Publishing.
81. Monserrat, T. J. K. P., Li, Y., Zhao, S., & Cao, X. (2014, April). L. IVE: an integrated interactive video-based learning environment. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 3399-3402). ACM.
82. Monserrat, T. J. K. P., Zhao, S., McGee, K., & Pandey, A. V. (2013, April). NoteVideo: facilitating navigation of blackboard-style lecture videos. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1139-1148). ACM.
83. Motti, V. G., Fagá Jr, R., Catellan, R. G., Pimentel, M. D. G. C., & Teixeira, C. A. (2009, June). Collaborative synchronous video annotation via the watch-and-comment paradigm. In *Proceedings of the seventh european conference on European interactive television conference* (pp. 67-76). ACM.
84. Nacke, L. E., Kalyn, M., Lough, C., & Mandryk, R. L. (2011, May). Biofeedback game design: using direct and indirect physiological control to enhance game interaction. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 103-112). ACM.

85. Nenonen, V., Lindblad, A., Häkkinen, V., Laitinen, T., Jouhtio, M., & Hämäläinen, P. (2007, April). Using heart rate to control an interactive game. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 853-856). ACM.
86. Ogan, A., Finkelstein, S., Mayfield, E., D'Adamo, C., Matsuda, N., & Cassell, J. (2012, May). Oh dear stacy!: social interaction, elaboration, and learning with teachable agents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 39-48). ACM.
87. Parr, C. (2013). Not staying the course. *Times Higher Education.*
88. Pavel, A., Reed, C., Hartmann, B., & Agrawala, M. (2014, October). Video digests: a browsable, skimmable format for informational lecture videos. In *UIST* (pp. 573-582).
89. Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational psychology review*, *18*(4), 315-341.
90. Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In *Handbook of research on student engagement* (pp. 259-282). Springer US.
91. Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., & Perry, R. P. (2010). Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology*, *102*(3), 531.
92. Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Positive emotions in education.
93. Pelegris, P., Banitsas, K., Orbach, T., & Marias, K. (2010, August). A novel method to detect heart beat rate using a mobile phone. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology* (pp. 5488-5491). IEEE.
94. Pham, P., & Wang, J. (2015, June). AttentiveLearner: improving mobile MOOC learning via implicit heart rate tracking. In *International Conference on Artificial Intelligence in Education* (pp. 367-376). Springer International Publishing.
95. Pham, P., & Wang, J. (2016, November). Adaptive review for mobile MOOC learning via implicit physiological signal sensing. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction* (pp. 37-44). ACM.
96. Picard, R. W., & Picard, R. (1997). *Affective computing* (Vol. 252). Cambridge: MIT press.
97. Poh, M. Z., McDuff, D. J., & Picard, R. W. (2011). Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering*, *58*(1), 7-11.
98. Putze, F., Hild, J., Kärgel, R., Herff, C., Redmann, A., Beyerer, J., & Schultz, T. (2013, March). Locating user attention using eye tracking and EEG for spatio-temporal event selection. In *Proceedings of the 2013 international conference on Intelligent user interfaces* (pp. 129-136). ACM.
99. Rico, J., & Brewster, S. (2010, April). Usable gestures for mobile interfaces: evaluating social acceptability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 887-896). ACM.
100. Risko, E. F., Buchanan, D., Medimorec, S., & Kingstone, A. (2013). Everyday attention: Mind wandering and computer use during lectures. *Computers & Education*, *68*, 275-283.

101. Robinson, P. (2014, November). Modelling emotions in an on-line educational game. In *Control, Decision and Information Technologies (CoDIT), 2014 International Conference on* (pp. 628-633). IEEE.

102. Roda, C., & Thomas, J. (2006). Attention aware systems: Theories, applications, and research agenda. *Computers in Human Behavior*, *22*(4), 557-587.

103. Rodrigo, M. M. T., & d Baker, R. S. (2011). Comparing the incidence and persistence of learners' affect during interactions with different educational software packages. In *New perspectives on affect and learning technologies* (pp. 183-200). Springer New York.

104. Rodrigue, M., Son, J., Giesbrecht, B., Turk, M., & Höllerer, T. (2015, March). Spatio-Temporal Detection of Divided Attention in Reading Applications Using EEG and Eye Tracking. In *Proceedings of the 20th International Conference on Intelligent User Interfaces* (pp. 121-125). ACM.

105. Rowe, D. W., Sibert, J., & Irwin, D. (1998, January). Heart rate variability: Indicator of user state as an aid to human-computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 480-487). ACM Press/Addison-Wesley Publishing Co..

106. Schaaff, K., & Adam, M. T. (2013, September). Measuring emotional arousal for online applications: evaluation of ultra-short term heart rate variability measures. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on* (pp. 362-368). IEEE.

107. Sears, A., Lin, M., Jacko, J., & Xiao, Y. (2003, June). When computers fade: Pervasive computing and situationally induced impairments and disabilities. In *HCI International* (Vol. 2, No. 03, pp. 1298-1302).

108. Seaton, D. T., Bergner, Y., Chuang, I., Mitros, P., & Pritchard, D. E. (2014). Who does what in a massive open online course?. *Communications of the ACM*, *57*(4), 58-65.

109. Shah, D. (2015, December). MOOCs in 2015: Breaking Down the Numbers. *Edsurge.com*.

110. Singh, D., Saini, B. S., & Kumar, V. (2008). Heart rate variability-A bibliographical survey. *IETE Journal of Research*, *54*(3), 209-216.

111. Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological bulletin, 132*(6), 946.

112. Smith, E. E., & Kosslyn, S. M. (2013). *Cognitive Psychology: Pearson New International Edition: Mind and Brain*. Pearson Higher Ed.

113. Sugimoto, M., & Hiroki, K. (2006, September). HybridTouch: an intuitive manipulation technique for PDAs using their front and rear surfaces. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services* (pp. 137-140). ACM.

114. Sun, D., Paredes, P., & Canny, J. (2014, April). MouStress: detecting stress from mouse motion. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 61-70). ACM.

115. Szafir, D., & Mutlu, B. (2012, May). Pay attention!: designing adaptive agents that monitor and improve user engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 11-20). ACM.

116. Szafir, D., & Mutlu, B. (2013, April). ARTFul: adaptive review technology for flipped learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1001-1010). ACM.

117. Task Force of the European Society of Cardiology. (1996). Heart rate variability standards of measurement, physiological interpretation, and clinical use. *Eur Heart J, 17,* 354-381.

118. Van der Sluis, F., Ginn, J., & Van der Zee, T. (2016, April). Explaining Student Behavior at Scale: The Influence of Video Complexity on Student Dwelling Time. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale* (pp. 51-60). ACM.

119. Vogel-Walcutt, J. J., Abich, J., & Schatz, S. (2012). Boredom in learning. In *Encyclopedia of the Sciences of Learning* (pp. 477-479). Springer US.

120. Vrijkotte, T. G., Van Doornen, L. J., & De Geus, E. J. (2000). Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension, 35*(4), 880-886.

121. Wang, J., Zhai, S., & Canny, J. (2006, October). Camera phone based motion sensing: interaction techniques, applications and performance study. In *Proceedings of the 19th annual ACM symposium on User interface software and technology* (pp. 101-110). ACM.

122. Weir, S., Kim, J., Gajos, K. Z., & Miller, R. C. (2015, February). Learnersourcing subgoal labels for how-to videos. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 405-416). ACM.

123. Whitehill, J., Williams, J. J., Lopez, G., Coleman, C. A., & Reich, J. (2015). Beyond prediction: First steps toward automatic intervention in MOOC student stopout. *Available at SSRN 2611750*.

124. Wigdor, D., Forlines, C., Baudisch, P., Barnwell, J., & Shen, C. (2007, October). Lucid touch: a see-through mobile device. In *Proceedings of the 20th annual ACM symposium on User interface software and technology* (pp. 269-278). ACM.

125. Wobbrock, J. O., Chau, D. H., & Myers, B. A. (2007, April). An alternative to push, press, and tap-tap-tap: gesturing on an isometric joystick for mobile phone text entry. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 667-676). ACM.

126. Wobbrock, J. O., Myers, B. A., & Aung, H. H. (2008). The performance of hand postures in front-and back-of-device interaction for mobile computing. *International Journal of Human-Computer Studies*, *66*(12), 857-875.

127. Wobrock, D., Frey, J., Graeff, D., De La Rivière, J. B., Castet, J., & Lotte, F. (2015, September). Continuous mental effort evaluation during 3d object manipulation tasks based on brain and physiological signals. In *Human-Computer Interaction* (pp. 472-487). Springer International Publishing.

128. Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: recognising and responding to student affect. *International Journal of Learning Technology*, *4*(3-4), 129-164.

129. Wu, T. Y., & Chao, H. C. (2008). Mobile e-learning for next generation communication environment. *International Journal of Distance Education Technologies, 6*(4), 1.

130. Xiao, X., & Wang, J. (2015, November). Towards Attentive, Bi-directional MOOC Learning on Mobile Devices. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (pp. 163-170). ACM.

131. Xiao, X., & Wang, J. (2016, November). Context and Cognitive State Triggered Interventions for Mobile MOOC Learning. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction* (pp. 378-385). ACM.

132. Xiao, X., & Wang, J. (2017, May). Detecting Divided Attention in Mobile MOOC Learning. In *Proceedings of the 35th Annual ACM Conference on Human Factors in Computing Systems*. ACM. (to appear)

133. Xiao, X., Han, T., & Wang, J. (2013, December). LensGesture: augmenting mobile interactions with back-of-device finger gestures. In *Proceedings of the 15th ACM on International conference on multimodal interaction* (pp. 287-294). ACM.

134. Xiao, X., Pham, P., & Wang, J. (2015, November). AttentiveLearner: Adaptive Mobile MOOC Learning via Implicit Cognitive States Inference. *Demo of the 2015 ACM on International Conference on Multimodal Interaction* (pp. 373-374). ACM.

135. Xu, Q., Li, L., & Wang, G. (2013, April). Designing engagement-aware agents for multiparty conversations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2233-2242). ACM.

136. Yang, D., Sinha, T., Adamson, D., & Rosé, C. P. (2013, December). Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the 2013 NIPS Data-driven education workshop* (Vol. 11, p. 14).

137. Yang, D., Wen, M., Howley, I., Kraut, R., & Rose, C. (2015, March). Exploring the effect of confusion in discussion forums of massive open online courses. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale* (pp. 121-130). ACM.

138. Yatani, K., Partridge, K., Bern, M., & Newman, M. W. (2008, April). Escape: a target selection technique using visually-cued gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 285-294). ACM.

139. Yin, Y., Ouyang, T. Y., Partridge, K., & Zhai, S. (2013, April). Making touchscreen keyboards adaptive to keys, hand postures, and individuals: a hierarchical spatial backoff model approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2775-2784). ACM.

140. Yuksel, B. F., Oleson, K. B., Harrison, L., Peck, E. M., Afergan, D., Chang, R., & Jacob, R. J. (2016, May). Learn piano with BACh: An adaptive learning interface that adjusts task difficulty based on brain state. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5372-5384). ACM.