Context and Cognitive State Triggered Interventions for Mobile MOOC Learning

Xiang Xiao[1,2] and Jingtao Wang[1,2]
[1]Department of Computer Science
[2]Learning Research and Development Center (LRDC)
University of Pittsburgh, PA, USA
{xiangxiao, jingtaow}@cs.pitt.edu

ABSTRACT
We present Context and Cognitive State triggered Feed-Forward (*C2F2*), an intelligent tutoring system and algorithm, to improve both student engagement and learning efficacy in mobile Massive Open Online Courses (MOOCs). *C2F2* infers and responds to learners' boredom and disengagement events in real time via a combination of camera-based photoplethysmography (PPG) sensing and learning topic importance monitoring. It proactively reminds a learner of upcoming important content (feed-forward interventions) when disengagement is detected. *C2F2* runs on unmodified smartphones and is compatible with courses offered by major MOOC providers. In a 48-participant user study, we found that *C2F2* on average improved learning gains by 20.2% when compared with a baseline system without the feed-forward intervention. *C2F2* was especially effective for the bottom performers and improved their learning gains by 41.6%. This study demonstrates the feasibility and potential of using the PPG signals implicitly recorded by the built-in camera of smartphones to facilitate mobile MOOC learning.

CCS Concepts
• **Human-centered computing→Ubiquitous and mobile computing→Ubiquitous and mobile computing systems and tools**

Keywords
MOOC; Heart Rate; Intelligent Tutoring Systems; Physiological Signals; Affective Computing, Mobile Interfaces

1. INTRODUCTION
Massive Open Online Courses (MOOCs) have emerged as a promising solution for delivering high-quality educational content on a large scale at a low cost. With more than 35 million registered students by December 2015 [25], MOOCs are becoming a common form of online course delivery. MOOC providers, such as Coursera, edX, and Udacity, also offer mobile apps to support "*learning on the go*". When used with mobile devices, MOOCs are "*ubiquitous, respond to urgent learning need, and flexibility of location and time to learn"* [32]. Lecture videos split into 3-15 minute segments are the primary delivery method for educational content in these MOOC mobile apps. Such small video clips can improve student engagement and are easy to consume on mobile devices.

Despite the rapid growth of MOOCs, educators and researchers have identified significant challenges surrounding MOOCs that must be addressed. Unlike traditional classrooms in which teachers can use facial expressions and behaviors to infer student disengagement, and increase or recapture student attention through interactions such as eye contact, questions, or in-class activities, the video lectures used in MOOCs are mostly static, non-interactive, and cannot monitor how well the student pays attention to the lecture. Therefore, students often lack sustained motivation to watch all lecture videos, leading to a high in-video drop-out rate (55.2% in [16]) and a low course completion rate (10% in [9], 7% in [21]). This problem could be even worse for mobile MOOC learning as students are more prone to "mind wandering" due to the highly diverse learning environments and highly interruptive learning contexts when studying alone with their mobile devices [24].

Although researchers have proposed various interaction techniques [14][15][20] to improve the interactivity of MOOC videos, none of these techniques can directly measure or repair learner disengagement during learning. Furthermore, most of these techniques cannot be directly deployed on mobile platforms due to the small size of mobile devices.

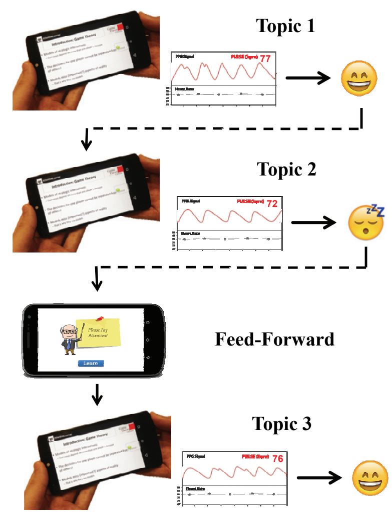

Figure 1. The proposed system detects learner disengagement by analyzing implicitly captured PPG signals. The system uses feed-forward to remind learners when they are disengaged.

In this paper, we propose a novel intervention technology, ***context and cognitive state triggered feed-forward (C2F2),*** to remind learners of their disengagement states in mobile MOOC learning.

ICMI'16, November 12–16, 2016, Tokyo, Japan
© 2016 ACM. 978-1-4503-4556-9/16/11...$15.00
http://dx.doi.org/10.1145/2993148.2993177

C2F2 monitors a learner's engagement while she is watching lecture videos and adaptively reminds the learner of upcoming important content when she is disengaged (Figure 1). *C2F2* is built upon AttentiveLearner [33], which captures learners' physiological signals in MOOC learning through implicit photoplethysmography (PPG) sensing on unmodified mobile phones. While AttentiveLearner focuses on improving instructors' understanding of the MOOC learning process via *offline analytics*, *C2F2* provides *real-time* predictions of the learner's engagement state for each learning topic using the implicitly captured PPG signals. Moreover, *C2F2* proactively initiates feed-forward *interventions* to re-engage the learner if she is currently disengaged and the upcoming topic is important.

This paper offers two major contributions:

- We present the design and evaluation of a context and cognitive state triggered intervention technology within a mobile learning system. To our knowledge, we are the first to systematically investigate the effectiveness of using cognitive state triggered proactive reminders (feed-forward) as an intervention to recognize and alleviate disengagement in mobile MOOC learning.
- We show the feasibility and effectiveness of using PPG signals implicitly recorded by mobile cameras to improve mobile MOOC learning. Our PPG-based intervention technique running on unmodified smartphones achieved better performance than state-of-the-art electroencephalography (EEG) based methods [27][28].

2. RELATED WORK

2.1 Disengagement in Learning

Disengagement/boredom is one of the most frequent affective states and is persistent across various learning environments [2]. Previous research has shown that boredom is a negative affective state that interactive learning environments should focus on detecting and quickly responding to [2][7][6][29]. Craig et al. [6] investigated the role of affective states in learning through a user study with 38 undergraduate students. Results of the study showed a significant negative correlation between boredom and learning gains. Baker et al. [2] analyzed data on students' cognitive-affective 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 (i.e., guess or abuse hint features so as to perform well in the learning environment without learning the material), which was known to be associated with poorer learning. Boredom could disengage learners from educational activities and seriously decrease their abilities to acquire knowledge [29]. Boredom was also found to adopt a persistent temporal quality [7], 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 during learning. The proposed intervention technique, *C2F2*, is a disengagement repair technique specifically designed for mobile MOOC learning.

2.2 Technologies for Improving MOOCs

Various techniques have been proposed to address challenges faced by MOOCs. Based on their design goals, these techniques can be grouped into three categories: 1) enhance learner engagement by improving the quality and interactivity of MOOC videos [1][18][15][22][14][20][30]; 2) promote student-student or student-instructor communications [5][10][4][19]; and 3) post-hoc analysis of clickstream data and major video interaction events [17][35][12][16].

Researchers have used video annotations (e.g. digital footnotes [18], sub-goal labels [30], etc.) to augment video lecture interactions [30][18] and enhance the video viewing experience [1]. Furthermore, various navigation controls have been proposed to help users browse and skim videos [15][22]. There also exist other techniques which add interactive elements in the video, such as interactive exercises [14] and embedded comment threads/assessments [20]. One problem of these interaction techniques is a lack of personalization for individual learners. Some techniques also require extra video production efforts.

The techniques to promote communications within MOOCs include asynchronous communication techniques, such as discussion forums [5] and post-lecture reflections [10]; synchronous communication techniques, such as chat-room systems [4]; and hybrid techniques which combine elements of both synchronous and asynchronous communication, such as time-anchored commenting [19]. These techniques either promote student-student interactions [4][19], or support student-instructor feedback [10]. However, most of these interactive techniques rely heavily on learners' active participation while prior research indicated low participation rates in activities or class discussions in MOOC contexts [3].

For post-hoc video/log analyses, researchers analyzed both activities within learning sessions (e.g. click-level interactions [12][16] and visual attention [17] within MOOC videos) and activities in the follow-up discussion forums [35]. Such analyses can reveal insightful information, such as the correlation between video productions and student engagement [16][17], and factors that contribute to course dropout [35][16]. Although post-hoc analyses can provide insightful information about MOOC learning, they reveal learners' behaviors in learning rather than their actual learning process. There is still little *direct* measurement of learners' actual learning process in MOOCs.

2.3 Detecting and Responding to Affective and Cognitive States in Education

Researchers have built various learning systems which detect and respond to learners' affective and cognitive states [27][28][7][8][31]. These systems collect physiological signals, such as heart rates [33][23], facial features [31], eye gaze [7], and EEG signals [27][28], and use machine learning algorithms to predict students' affective and cognitive states (e.g. boredom, confusion, and mind wandering) in learning. These affect-sensitive systems then dynamically respond to the sensed affective and cognitive states using pedagogical strategies, such as direct feedback [7][8], pedagogical agent[31], and adaptive activities[27] .

Gaze Tutor [7] 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 the gaze-sensitive statements were associated with a significant improvement for students with high aptitude, while not as effective for students with average aptitude.

To predict learners' affective states, the Affective AutoTutor [8] monitors multi-channel physiological signals. 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.

Szafir et al. [22] developed an adaptive-review technology which monitored learners' attention using their EEG signals and adaptively provided reviews on topics with low-attention levels.

One common problem with most of these systems is the requirement of dedicated sensors, such as cameras or EEG headsets to collect physiological signals. In comparison, our system is built on top of AttentiveLearner [33], which uses the built-in camera of mobile devices to implicitly collect and analyze real-time PPG signals, thus making it easy to be adopted beyond lab settings.

3. SYSTEM DESIGN

We designed and implemented a novel intelligent learning system optimized for mobile MOOC learning. Similar to existing mobile MOOC clients such as Coursera, edX, and Udacity, lectures in the system are organized as short video clips, each clip presenting a coherent, semi-independent subtopic. The new system adopts the same *tangible video control channel* and *implicit PPG sensing module* of AttentiveLearner [33]. It monitors a learner's engagement of each video clip in real-time by implicitly sensing and analyzing PPG signals of the learner. When learner disengagement is detected, it attempts to regain learner attention through *context and cognitive state triggered feed-forward* (*C2F2*) intervention.

3.1 AttentiveLearner

In AttentiveLearner, the built-in back camera is converted to a video control channel. A learner uses his/her fingertip to cover and hold the back camera lens to play a lecture video and uncover the lens to pause the video. The detection of lens finger-covering gesture is based on the *Static LensGesture* in [34]. Moreover, while the learner is watching lecture videos by covering the back camera lens, AttentiveLearner also implicitly collects his/her PPG signals by analyzing the learner's fingertip transparency changes captured by the back camera (commodity camera based PPG sensing [13]). AttentiveLearner uses the LivePulse algorithm [13] to extract learners' PPG signals. This algorithm is reported to be accurate, with a 3.9% mean error rate of estimating instant heart rate when users are in a resting condition [13].

The AttentiveLearner mobile interface was shown to be comfortable and natural to use through two controlled user studies [33]. Various usability concerns, such as battery life, accuracy and speed of the lens-covering gesture detection, and quality of the collected PPG signals, have been addressed in [33].

Figure 2. Primary Interface of *C2F2* (left: camera lens is uncovered, video paused; right: lens is covered, video playing)

3.2 The *C2F2* System

The *C2F2* system is built on top of AttentiveLearner. A new intervention technique, *C2F2*, is integrated into AttentiveLearner to address learner disengagement and improve learning outcomes.

3.2.1 *C2F2 Intervention*

The idea of *C2F2* intervention is based on two key assumptions in MOOC learning. First, if a learner becomes disengaged watching one video, she is likely to stay disengaged watching similar videos shortly. We made this assumption based on the temporal persistence nature of disengagement/boredom [2][7], e.g. "*Once a student is bored, it appears to be difficult to transition out of boredom*" [2]. 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, we propose *C2F2* to repair students' disengagement and help them maintain sustained engagement across multiple video sessions. The 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, *C2F2* takes into account both the learner's *cognitive states* and the *intrinsic importance* of the upcoming learning topic to determine the timing of intervention.

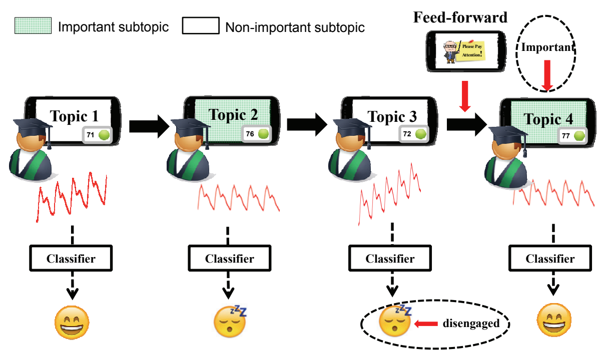

Figure 3. The working mechanism of *C2F2*. A feed-forward reminder is presented after Topic 3 because the learner is disengaged watching the Topic 3 video and the next topic (Topic 4) is important.

In the *C2F2* system, a feed-forward reminder appears 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 3, the feed-forward after Topic 3). If only one condition or no conditions are satisfied, the system directly presents the next subtopic. By considering both the content of the video, and the learner's real-time cognitive states, we hope to effectively regulate learners' disengagement/boredom state without frustrating them with too many feed-forward reminders.

3.2.2 *Disengagement Detection*

By monitoring and analyzing learners' PPG signals during MOOC learning on unmodified mobile phones in real time, we have the opportunity to infer important cognitive states such as mind wandering events [23] as well as boredom and confusion states [33] in learning. We focused on the detection of boredom/disengagement in our system.

When a learner is watching tutorial videos with *C2F2*, the system also implicitly captures her PPG signals at the same time. The raw PPG signals are processed by the LivePulse algorithm [13] to extract RR-intervals (the cardiac interval between two consecutive heart beats) and instant heart rates. Outliers of the RR-intervals are removed using the same heuristics in [33]. 24 dimensions of heart rate features are then extracted. Similar to [33], 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. While features 1-7 have already been used in [33], five new global features are added to achieve better prediction performance. The other 12 dimensions are corresponding local features extracted by averaging the same features in multiple fixed-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.

We 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, this prediction takes on average 1.51 seconds, which is hardly noticeable according to participants in the user study.

3.2.3 Feed-forward Reminder

The feed-forward reminder prompts a learner of upcoming important topic so as to redraw her attention back to the videos. Figure 4 is the design of feed-forward reminders after two rounds of pilot studies. The cartoon character acts as a learning companion who attracts the learner's attention. We 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 appears. The learner needs to explicitly acknowledge the feed-forward reminder by pressing the "*Learn*" button.

Through our pilot studies, we found that users preferred direct, concise messages more than indirect, polite messages. Meanwhile, we found it important to avoid using negative statements. Statements such as "*you should pay more attention*" or "*you are not paying enough attention*" might discourage learners, especially when they thought they have already paid enough attention to the video. Therefore, in the audio message, we attribute the occurrence of feed-forward to "*the next topic is very important*" to avoid eliciting negative emotions from the user.

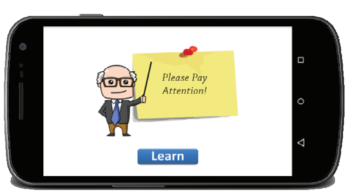

Figure 4. The feed-forward reminder presented to users.

Our feed-forward design also adopted a high-interruptive presentation. The learner has to explicitly acknowledge it before watching the next video. According to [8], the high-interruptive presentation is more effective than low-interruptive indicators when the learner is in a negative learning state.

4. EVALUATION

We conducted a lab-based study to further understand *C2F2*. We 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 investigate effects of *C2F2* on learning, we 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.

4.1 Experimental Design

We conducted a between-participant study in which we manipulated the presentation timing of feed-forward within a mobile MOOC learning system. The independent variable was the *type of feed-forward intervention* received: (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. This is a topic that participants were unlikely to have prior knowledge of, 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, we selected the second ("*Computer Virus and Worms*"), third ("*AIC Principles*"), fifth ("*Access Control*") and sixth ("*Session Hijacking*") topics as the important topics. These videos were important because they conveyed essential and relevant topics. In comparison, the first and the fourth video clips either contained some trivia, non-technical contents, or duplicate contents from previous videos.

To assess learning performance, participants were asked to answer eight multiple-choice questions for each subtopic. Unlike previous studies [7][8][27] which had the evaluation session after the whole learning session, we 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 differences in memory ability 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, we 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 [14]; however, the effect of quizzes on learner engagement is out of the scope of this paper.

We first conducted a pilot study to train and optimize the disengagement prediction classifier. We recruited 10 participants (4 females) between 23 to 33 years old (μ=27.8, σ=2.8) for this study. All participants were graduate students from a local university. Participants watched the six subtopic videos introduced earlier using AttentiveLearner (no feed-forward reminder 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 the performance of the classifiers. Of the 60 ratings (6 videos x 10 participants), 51.67% indicated disengagement during learning (rating <= 3).

We 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. In this study, we intentionally controlled the number of feed-forward reminders to avoid confounders. Otherwise if subjects receiving more reminders in one condition outperformed subjects receiving fewer reminders in another condition, it will be difficult to determine whether the difference was caused by the intervention or by the difference in the number of reminders. Therefore, 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, we 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 (0.8, 0.6, 0.5, 0.6, 0.5, 0.5). The adaptive thresholds are determined by participants' average engagement ratings reported for the six videos in the pilot study. For example, because no participant 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.

4.2 Procedure

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. Participants were randomly assigned to one of the four experimental conditions. 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*. Participants also took an 8-question, multiple choice quiz, which tested their understanding of the subtopic video they had just watched. After participants completed the questionnaire and quiz, they continued to learn the next subtopic.

During the MOOC learning and evaluation phase, our 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* of the whole learning session. Next, each participant took a post-experiment questionnaire to obtain their subjective evaluations of the mobile MOOC application.

4.3 Participants and Apparatus

Forty-eight subjects (28 males and 20 females) participated in our study (Figure 5). Each of the four conditions was gender balanced (seven males and five females). The average participant age was 23.4 (σ = 3.5) ranging from 18 to 32. All participants were undergraduate or graduate students recruited from a local university by fliers 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%). The average pre-lecture quiz scores for the four important topics were 11.8%, 9.7%, 11.8% and 16.0% respectively. Repeated Measures ANOVA ($F(3, 45) = 0.77$, $p = 0.516$) revealed no significant difference across the subtopics. The Condition Subtopic interaction for pre-lecture quiz scores was non-significant ($F(9, 132) = 0.523$, $p = 0.856$).

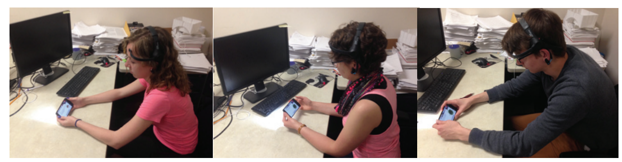

Figure 5. Sample participants in our experiments. Participants also wore an EEG headset during learning.

Our 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.

5. RESULTS

5.1 Signal Quality

The mobile MOOC systems collected PPG signals while participants were watching lecture videos and stored them on the mobile device. We have collected a total of 1305 minute PPG signals from the 48 participants (average 27 min 11s per participant). We analyzed the quality of collected PPG signals by investigating the RR-intervals in a 5-second moving window. We used the same signal quality metric as Xiao and Wang [33]. Using this metric, we calculated the percentage of high-quality signals for each video. We found that 73.26% of the 288 (48 x 6) video sessions, more than 80% of the signals were in high quality. This statistic was a little lower than the quality reported in [33], probably due to the shorter video clips and more diversity in the participants. In general, *C2F2* collected reliable PPG signals from learners' fingertips during video watching.

5.2 Feed-Forward Accuracy

To verify that our system was working correctly, we 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. We excluded participants whose ratings suggested consistent engagement throughout the whole learning session, as in this case, the position of feed-forward interventions was unlikely to make any difference. For the *context only* condition, 39.13% feed-forward were presented at the right position; for the *cognitive only* condition*,* 27.79% feed-forward were presented at the right position; and for the *C2F2* condition, 62.5% feed-forward were presented at the right position. If we did not consider presenting feed-forward before important videos as a constraint, in the *cognitive only* condition, 56.6% feed-forward were presented at the right position.

Each participant received exactly two feed-forward reminders in any feed-forward conditions. Some of these feed-forward reminders were extra because they were not triggered by the learner's cognitive state, but to balance the total number of feed-forward received. Ignoring such extra reminders for balancing purposes, only 12.25% feed-forward were triggered at the wrong place in the *C2F2* condition, 37.5% feed-forward were triggered at the wrong place in the *cognitive only* condition, and 41.67% feed-forward were triggered at the wrong place in the *context only* condition. If we did not impose the constraint of presenting feed-forward before important videos, only 12.5% feed-forward were triggered at the wrong place in the *cognitive only* condition. Therefore, in the *C2F2* condition, our algorithm generally presented feed-forward in the correct position.

5.3 Learning Performance

Our experiment was based on the concept that different feed-forward interventions would affect learning, thus we first utilized analysis of variance (ANOVA) to analyze the effect of feed-forward interventions on participants' learning performance. We 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. 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.

We 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 correction 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 we did not observe significant learning differences among 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, we divided participants in each condition into two groups. Based on participants' scores in the evaluation quizzes, we had the bottom half learners (six participants) who received the lower scores on the quizzes than the top half learners.

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. We observed a significant main effect of the 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 the bottom performers learned significantly better in the *C2F2* condition than in the *no feed-forward* condition (t(10) = 0.1829, p = 0.0018, d = 2.1501) and the *context only feed-forward* condition (t(10) = 0.1749, p = 0.0025, d = 1.9999). Although the sample size (N = 6) is small, the small p value and the large effect size suggested a high practical significance. However, we did not observe a significant main effect of the 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 6 shows major results of the study.

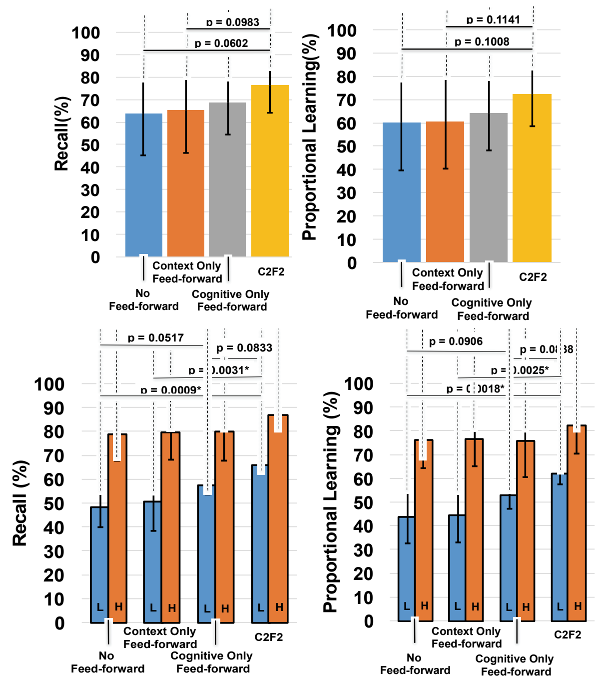

Figure 6. Recall and learning gains by conditions. Top: results for all participants, bottom: results for bottom performers (L) and top performers (H). (*) denotes significant differences.

The results suggested that *C2F2* was especially effective for the 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 impressions of the lectures, three of them reported consistent engagement and attention throughout the entire 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.

Our 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 entire learning session.

5.4 Subjective Feedback

Participants reported an average rating of 3.67 (σ = 0.89), 4 (σ = 0.60), 4.17 (σ = 0.72), and 4.17 (σ = 0.58) of the overall experience of using the mobile MOOC system in the *no feed-forward*, *context only feed-forward*, *cognitive only feed-forward* and *C2F2* conditions on a five-point Likert Scale (1 - strongly dislike, 5 – strongly like).

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'm learning, extrinsic motivation often is not helpful for me. If I do not find a topic interesting, it's 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, "*I could tell I preferred the one video about the computer virus. It was more interesting for me and also easier to follow.*"

5.5 Comparison with EEG

Another goal of the study was to compare the performance of our 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 [27][28] from physiological signals. As a result, all participants were required to wear a Neurosky Mindwave EEG headset (see Figure 5) during the learning session. This setup was similar to previous studies investigating the use of EEG-monitored attention during learning [27][28]. Among the 48 participants in the experiment, the EEG signals from 11 participants were either incomplete (due to a device problem) or partially unusable (highly corrupted by noises). Therefore, we compared the performance of the PPG-based method and EEG-based method using data from the remaining 37 subjects.

We performed off-line analysis and used the EEG-based engagement monitoring algorithm in [27][28] 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. We used participants' self-reported engagement ratings for each video as the ground truth. We 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 performance of our PPG-based engagement prediction method. We 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, we were also able to predict the video(s) with the lowest engagement. The same set of PPG features as well as signal processing methods presented in section 3 were used. The leave-one-subject-out evaluation was utilized to evaluate 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, we could correctly identify the video during which a learner showed the least engagement with 55.56% accuracy. On the other hand, our 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*).

Although we achieved a generally better performance using our specifically trained machine learning models, the EEG-based engagement monitoring method does not require any learning phase. Moreover, the attention index is updated every one second, thus it is able to capture finer-grained attention changes. On the other hand, we used PPG signal sequences of a few minutes to predict learners' general attention and engagement over a period of time.

One problem we observed during our 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 our study, participants were instructed to wear the device before the learning session and take off the device after the entire 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 that we were not able to collect complete good quality EEG data from the 11 participants excluded from this analysis.

5.6 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 mind wandering at a moment using the implicitly captured PPG signal is moderate at best (highest precision 40% and highest recall 65% in [23]). 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 is still disengaged *before* the *C2F2* reminder. To address this problem, *C2F2* 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.

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.[26] 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 our study, most participants commented that they could stay focused for 3 to 5 minutes after seeing the feed-forward reminder. For longer videos, we may use brief in-video alert in the middle of the video.

6. CONCLUSIONS

We present Context and Cognitive State triggered Feed-Forward (*C2F2*), an intelligent tutoring system and algorithm, to improve both student engagement and efficacy in mobile MOOC learning. Through a 48-participant user study, we found that *C2F2* on average improved learning gains by 20.2% when compared to a baseline system. *C2F2* was especially effective for the bottom performers and improved their learning gains by 41.6%. We also showed that using PPG signals implicitly captured by the system to predict disengagement, we achieved better performance than the state-of-the-art EEG-based method.

7. REFERENCES

[1] Cross. A. et al. VidWiki: Enabling the crowd to improve the legibility of online educational videos. In Proc. *CSCW* 2014.

[2] Baker, R., et al. 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, In *IJHCS* 2010.

[3] Breslow,L., et al. Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research & Practice in Assessment 8*, (2013), 13–25.

[4] Coetzee, D., et al. Chatrooms in MOOCs: All talk and no action. In Proc. L@S 2014.

[5] Coetzee, D., et al. Should your MOOC forum use a reputation system? In Proc. *CSCW* 2014.

[6] Craig, S., et al. Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. *Journal of educational media*, 29.3 (2004): 241-250.

[7] D'Mello, S., et al. Gaze tutor: A gaze-reactive intelligent tutoring system. In Proc. *IJHCS* Vol 70.5 (2012).

[8] D'Mello, S., et al. A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning. In *Proc. ITS 2010*.

[9] Fowler, G., An early report card on massive open online courses. *The Wall Street Journal*, October 8, 2013.

[10] Glassman, E., et al. Mudslide: A spatially anchored census of student confusion for online lecture videos. In *Proc. CHI 2015*.

[11] Grawemeyer, B., et al. Light-Bulb Moment?: Towards Adaptive Presentation of Feedback based on Students' Affective State. In *Proc. IUI 2015*.

[12] Guo, P. J., et al. How video production affects student engagement: An empirical study of MOOC videos. In *Proc. L@S 2014*.

[13] Han, T., et al. Balancing accuracy and fun: designing camera based mobile games for implicit heart rate monitoring. In *Proc CHI 2015*.

[14] Kim, J., et al. RIMES: embedding interactive multimedia exercises in lecture videos. In *Proc. CHI 2015*.

[15] Kim, J., et al. Data-driven interaction techniques for improving navigation of educational videos. In *Proc. UIST 2014*.

[16] Kim, J., et al. Understanding in-video dropouts and interaction peaks in online lecture videos. In *Proc. L@S 2014*.

[17] Kizilcec, R., et al. Showing face in video instruction: effects on information retention, visual attention, and affect. In *Proc CHI 2014*.

[18] Kumar, V. Enhancing video lectures with digital footnotes. In Proc IEEE *FIE* 2014.

[19] Lee, Y., et al. Using time-anchored peer comments to enhance social interaction in online educational videos. In *Proc. CHI 2015*.

[20] Monserrat, T., et al. L.IVE: an integrated interactive video-based learning environment. In *Proc. CHI 2014*.

[21] Parr, C., Not Staying the Course, *Times Higher Education*, May 10, 2013.

[22] Pavel, A., et al. Video digests: A browsable, skimmable format for informational lecture videos. In *Proc. UIST 2014*.

[23] Pham, P., et al. AttentiveLearner: improving mobile MOOC learning via implicit heart rate tracking. In *Proc. AIED 2015*.

[24] Risko, E., et al. Everyday Attention: Mind Wandering and Computer Use during Lectures. *Computers & Education* 68, 2013.

[25] Shah, D. MOOCs in 2015: Breaking Down the Numbers. *Edsurge.com*, December 28, 2015.

[26] Smallwood, J., et al. The restless mind. *Psychological bulletin*, 132(6), 946 2006.

[27] Szafir, D., et al. ARTFul: adaptive review technology for flipped learning. In *Proc. CHI 2013*.

[28] Szafir, D., et al. Pay attention!: designing adaptive agents that monitor and improve user engagement. *In Proc. CHI 2012*.

[29] Vogel-Walcutt, J., et al. Boredom in learning. *Encyclopedia of the Sciences of Learning*. Springer US, 2012. 477-479.

[30] Weir, S., et al. Learnersourcing subgoal labels for how-to videos. In Proc *CSCW* 2015.

[31] Woolf, B., et al. Affect-aware tutors' recognising and responding to student affect. *Int. J. Learn. Technol*. 4, 3⁄4, Oct. 2009, pp129–164.

[32] Wu, T. Y., et al. Mobile e-learning for next generation communication environment. *International Journal of Distance Education Technologies* 6.4 2008.

[33] Xiao, X., et al. Towards Attentive, Bi-directional MOOC Learning on Mobile Devices. In *Proc. ICMI 2015*.

[34] Xiao, X., et al. LensGesture: augmenting mobile interactions with back-of-device finger gestures. In *Proc. ICMI 2013*.

[35] Yang, D., et al. Exploring the effect of confusion in discussion forums of massive open online courses. In *Proc. L@S 2015*.