AttentiveLearner^2: A Multimodal Approach for Improving MOOC Learning on Mobile Devices

Phuong Pham and Jingtao Wang(✉)

Computer Science and LRDC, University of Pittsburgh, Pittsburgh, PA, USA
{phuongpham, jingtaow}@cs.pitt.edu

Abstract. We propose AttentiveLearner^2, a multimodal mobile learning system for MOOCs running on unmodified smartphones. AttentiveLearner^2 uses both the front and back cameras of a smartphone as two complementary and fine-grained feedback channels in real time: the back camera monitors learners' photoplethysmography (PPG) signals and the front camera tracks their facial expressions during MOOC learning. AttentiveLearner^2 implicitly infers learners' affective and cognitive states during learning by analyzing learners' PPG signals and facial expressions. In a 26-participant user study, we found that it is feasible to detect 6 types of emotion during learning via collected PPG signals and facial expressions and these modalities are complement with each other.

Keywords: Mobile learning · Intelligent tutoring systems · Massive open online courses · Multimodal interaction

1 Introduction

By 2016, Massive Open Online Courses (MOOCs) have attracted over 700 universities and 58 million registered learners worldwide. To facilitate learning on-the-go, major MOOC providers, e.g. Coursera, edX, and Udacity, have launched their mobile apps. Despite the popularity and rapid growth, today's MOOCs still suffer from low completion rates (e.g. 5.5% as reported in [1]), low engagement, and little personalization. These challenges are caused, at least in part, by the limited interactions between instructors and learners in MOOCs. Other than activity logs [4] and surveys, there is little information from students to instructors representing the learning progress.

We propose AttentiveLearner^2 (Fig. 1), an emotion-aware multimodal intelligent learning system for MOOCs running on unmodified smartphones. AttentiveLearner^2 builds upon and extends AttentiveLearner [6] by Pham and Wang. Similar to AttentiveLearner, AttentiveLearner^2 uses on-lens finger gestures to control video playback (i.e. covering and holding the back camera lens to play a tutorial video, while uncovering the lens to pause the video) and implicitly sense learners' photoplethysmogram (PPG) signals. Going beyond AttentiveLearner, AttentiveLearner^2 leverages the front camera for real-time facial expressions analysis (FEA). By using a combination of PPG signals and facial expressions, AttentiveLearner^2 infers learners' affective and cognitive states during learning. We intentionally choose a superscripted

E. André et al. (Eds.): AIED 2017, LNAI 10331, pp. 561–564, 2017.
DOI: 10.1007/978-3-319-61425-0_64

2 (pronounced as “*square*”) in project name to emphasize that: (a) AttentiveLearner^2 is a major upgrade of AttentiveLearner [6]; and (b) it leverages two independent channels of signals, i.e. PPG and FEA, to model and understand learners. AttentiveVideo [5] is another relevant research project. In comparison, AttentiveVideo focuses on detection emotional responses to mobile ads, which is around 30 s long while MOOC videos usually last 3 to 20 min. Although previous research explored the use of PPG [6] and FEA [2] in learning environments, to the best of our knowledge, AttentiveLearner^2 is the first mobile learning system that supports both real-time PPG sensing and FEA on unmodified smartphones for MOOCs.

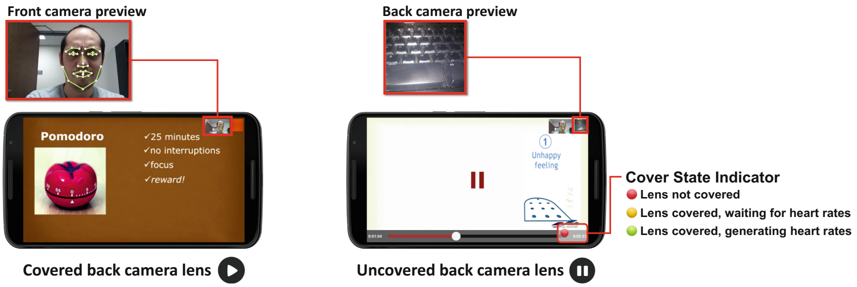

Fig. 1. AttentiveLearner^2 uses both the front and the back cameras as feedback channels: back camera for PPG sensing and front camera for facial expression analysis (FEA).

2 Design of AttentiveLearner^2

2.1 On-Lens Video Control

AttentiveLearner^2 uses tangible, on-lens finger gestures for video control: i.e. covering and holding the back camera lens play a lecture video while uncovering the lens pauses the video (Fig. 1). We utilize the *Static LensGesture* [7] for lens-covering detection.

2.2 Double-Camera Tracking System

AttentiveLearner^2 implicitly senses the PPG signals from a learner’s fingertip while she is watching a tutorial video. The underlining working mechanism is: the come and withdrawal of fresh blood in every cardiac cycle change the learner’s skin transparency color. These transparency changes (PPG signals) are highly correlated to heart beat cycles (NN intervals) and can be detected by the back camera. AttentiveLearner^2 uses *LivePulse* [3] to extract NN intervals from the detected PPG signals.

AttentiveLearner^2 also uses the front camera to monitor the learner’s facial expressions during the tutorial video. In this paper, we employ *Affdex* (http://affectiva.com) as the facial expression analysis library.

2.3 Emotion Inference Algorithms

AttentiveLearner^2 uses SVM with RBF kernels to detect learner's emotions while watching tutorial videos. We use leave-one-participant-out cross validation method.

AttentiveLearner^2 uses both global and sliding local windows to extract PPG signals and facial expression features (Fig. 2). In this project, we evaluate three different feature sets: PPG, FEA, and fusion feature set. The PPG feature set contains 8 dimensions of Heart Rate Variability (HRV): (1) AVNN (average NN intervals); (2) SDNN (standard deviations of NN intervals); (3) pNN10; (4) rMSSD; (5) SDANN; (6) SDNNIDX; (7) SDNNIDX/rMSSD; (8) MAD (median absolute deviation). In total, there are 16 PPG features (8 global features and 8 local features). With FEA features, we propose Action Unit Variability (AUV) capturing the dynamic of each facial expression output value from *Affdex*: (1) AVAU (average action unit value); (2) SDAU (3) MAXAU (the maximum value of action unit value during the video); (4) rMSSD; (5) SDAAU; (6) SDAUIDX; (7) SDNNIDX/rMSSD; (8) MAD. To balance with the PPG feature set, the FEA feature set selects the top 16 AUV features. Lastly, the feature fusion set selects 8 top PPG features and 8 top FEA features. All feature selections were done using univariate ANOVA as in [2].

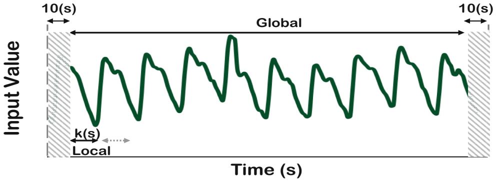

Fig. 2. PPG and FEA features are extracted from each video using global and local windows.

3 User Study

We conducted a within-subject user study to evaluate the feasibility of detecting emotions from PPG signals and facial expressions via unmodified smartphones. There were 26 participants (8 females) joining the user study. Each participant watches three tutorial videos (6 min/video) about Astronomy, Learning Science, and Programming. After each video, participants answered an emotion survey (7-point Likert scale format) about 6 different emotions: *boredom*, *confusion*, *curiosity*, *frustration*, *happiness*, and *self-efficacy*. We turned each emotion into a binary classification problem using thresholding, i.e. ratings smaller than 4 are negatives, otherwise positives.

Table 1 shows the detection performance of three feature sets and the majority vote baseline. AttentiveLearner^2 achieved high performance as all our models outperformed the baseline. Moreover, we found PPG signals and facial expressions are complement each other. If FEA features can win in 3 emotions (Confusion, Happiness, and Self-efficacy), PPG features are the best solution for Curiosity, and feature fusion can improve detection performance for Boredom and Frustration.

Table 1. Accuracy (Acc) and Kappa of prediction models.

Emotion	Majority	PPG		FEA		Feature fusion	
	Acc	Kappa	Acc	Kappa	Acc	Kappa	Acc
Boredom	70.51%	0.35	78.21%	0.56	84.62%	0.57	83.33%
Confusion	74.36%	0.30	78.21%	0.65	88.46%	0.54	84.62%
Curiosity	56.41%	0.46	74.36%	0.41	71.79%	0.43	73.08%
Frustration	78.21%	0.22	80.77%	0.69	91.03%	0.71	91.03%
Happiness	52.56%	0.41	70.51%	0.61	80.77%	0.61	80.77%
Self-efficacy	70.51%	0.38	79.49%	0.70	88.46%	0.67	87.18%

4 Conclusions and Future Work

We introduced AttentiveLearner^2, a multimodal emotion-aware interface for mobile MOOC learning on unmodified smartphones. In a 26-participant user study, we found that by taking advantages from two modalities, AttentiveLearner^2 achieved higher detection accuracy than models using only one modality across 6 different emotions. More importantly, these results were achieved on unmodified smartphones which supports the scalable deployment of AttentiveLearner^2. In the future, we plan use the inferred emotions to improve learner's outcomes and engagement.

References

1. Chuang, I., Ho, A.D.: HarvardX and MITx: Four Years of Open Online Courses – Fall 2012-Summer (2016). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2889436
2. D'Mello, S.K., Graesser, A.: Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. User Model. User-Adap. Inter. **20**(2), 147–187 (2010)
3. Han, T., Xiao, X., Shi, L., Canny, J., Wang, J.: 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 (2015)
4. Kim, J., Guo, P.J., Seaton, D.T., Mitros, P., Gajos, K.Z., Miller, R.C.: Understanding invideo dropouts and interaction peaks in online lecture videos. In: Proceedings of the First ACM Conference on Learning@ Scale Conference, pp. 31–40. ACM (2014)
5. Pham, P., Wang, J.: Understanding emotional responses to mobile video advertisements via physiological signal sensing and facial expression analysis. In: Proceedings of the 22nd International Conference on Intelligent User Interfaces, pp. 67–78. ACM (2017)
6. Pham, P., Wang, J.: AttentiveLearner: improving mobile MOOC learning via implicit heart rate tracking. In: Conati, C., Heffernan, N., Mitrovic, A., Verdejo, M.F. (eds.) AIED 2015. LNCS, vol. 9112, pp. 367–376. Springer, Cham (2015). doi:10.1007/978-3-319-19773-9_37
7. Xiao, X., Han, T., Wang, J.: 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 (2013)