AttentiveLearner: Improving Mobile MOOC Learning via Implicit Heart Rate Tracking

Phuong Pham and Jingtao Wang(✉)

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

Abstract. We present AttentiveLearner, an intelligent mobile learning system optimized for consuming lecture videos in both Massive Open Online Courses (MOOCs) and flipped classrooms. AttentiveLearner uses on-lens finger gestures as an intuitive control channel for video playback. More importantly, AttentiveLearner implicitly extracts learners' heart rates and infers their attention by analyzing learners' fingertip transparency changes during learning on today's unmodified smart phones. In a 24-participant study, we found heart rates extracted from noisy image frames via mobile cameras can be used to predict both learners' "mind wandering" events in MOOC sessions and their performance in follow-up quizzes. The prediction performance of AttentiveLearner (accuracy = 71.22%, kappa = 0.22) is comparable with existing research using dedicated sensors. AttentiveLearner has the potential to improve mobile learning by reducing the sensing equipment required by many state-of-the-art intelligent tutoring algorithms.

Keywords: Heart rate · Attention-aware interfaces · Mind Wandering · MOOC · Intelligent tutoring system · Affective computing · Zoning out · Mobile device

1 Introduction

With the rapid growth in recent years, Massive Open Online Courses (MOOCs) provide both opportunities and obstacles to learning at scale. On one hand, MOOCs allow learners to get access to diversified high quality learning materials at low cost, and "*to control where, what, how and with whom they learn*" [12]. As a result, there were around 16.8 million registered MOOC learners by the end of 2014 [1]. On the other hand, educators and researchers have raised concerns on the low completion rates (10% in [6], less than 7% in [15]), high in-session interruptions [7], and lack of interactions among students and instructors. In current MOOCs, pre-recorded lecture videos, split into 3 – 15 minutes pieces, is the dominant format for knowledge dissemination. In fact, major MOOC providers such as Coursera, edX, and Udacity, have released mobile apps to allow learners to consume video materials "on the move".

Unfortunately, MOOCs today face at least three major challenges. First, learners are more prone to "*mind wandering*" (MW, or zoning out) in non-classroom environments [16]. This is in part due to external distractions and the lack of *sustained motivation* when studying alone. The second problem is the current design of MOOCs is

C. Conati et al. (Eds.): AIED 2015, LNAI 9112, pp. 367–376, 2015.
DOI: 10.1007/978-3-319-19773-9_37

primarily *uni-directional*, i.e. from instructors to students. Although feedback forms and learner activity logs (e.g. log-in frequency, in-page dwell time, click-through rates) can be used to infer learning efficacy [11], such measurements are only *indirect measurement* of the cognitive states in learning. As a result, instructors have little information on how well lectures are received by the learners. Finally, there is little personalization of instruction. It is hard for the instructors in MOOCs to cater learning materials for individual learners' need and learning process. Different from traditional classrooms, the instructors can no longer rely on facial cues and in-class activities to discover learners who are struggling or MW.

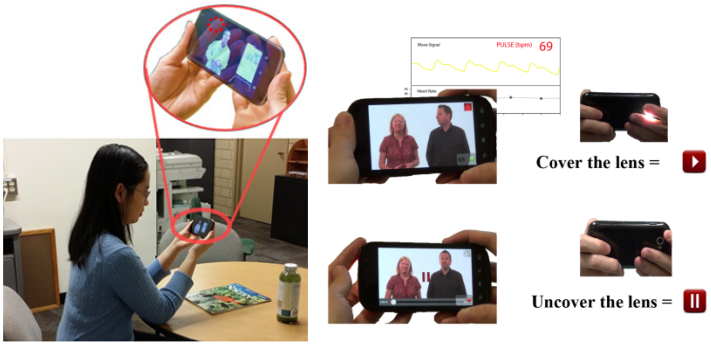

Fig. 1. AttentiveLearner uses the back camera as both a video play control channel in MOOC and an implicit heart rate sensing channel in learning

In response to these challenges, we propose AttentiveLearner (Fig. 1), an intelligent mobile learning system which supports attentive and bi-directional learning on *unmodified mobile phones*. AttentiveLearner uses on-lens finger gestures as an intuitive control mechanism for video playback (i.e. covering and holding the camera lens to play an instructional video, uncovering the lens to pause the video, Fig. 1 right). More importantly, AttentiveLearner *implicitly* extracts learners' heart rates and infers "zoning out" events by analyzing fingertip transparency changes captured by the built-in cameras. With MW information from learners, AttentiveLearner has the potential to enable adaptive tutoring features on today's mobile phones (e.g., alerting learners when zoning out, providing more relevant review exercises). AttentiveLearner can also help instructors to improve their syllabus and teaching style by providing an aggregated timeline view of learners' attention levels synchronized with the learning material.

Our main contributions are two folds. First, we discuss the design, implementation, and evaluation of AttentiveLearner. To our knowledge, AttentiveLearner is the first mobile MOOC learning system that infers learners' cognitive states during video watching via implicit heart rate tracking on today's unmodified mobile phones. Second, our 24-subject experiment shows AttentiveLearner can predict learners' MW states and their quiz performance in a user-independent fashion via heart rate signals implicitly captured from today's commodity mobile cameras. The accuracy and kappa of AttentiveLearner are comparable with existing technologies that rely on dedicated sensors.

2 Related Work

Various techniques have been explored to enrich both the output and feedback of MOOCs. For example, L.IVE by Monserrat et al [14] allows learners to comment, annotate, and complete assessment questions directly on top of MOOC videos. Kim et al [11] mined mouse-click logs on edX to infer drop-out patterns in MOOCs. In comparison, AttentiveLearner enables a new heart rate sensing channel directly correlated with learners' attention and cognitive states on mobile devices without hardware modification. AttentiveLearner can bring new opportunities to enrich large scale learning analytics and adaptive learning in MOOCs.

Existing research on using physiological signals to infer learners' cognitive states, affective states, and attention levels can be a promising direction complimentary to MOOCs. Researchers demonstrated the feasibility of using heart rates [20], galvanic skin response [3], facial expressions [3], mouse mounted with pressure sensor [20], and Electroencephalography (EEG) [18] to infer learners' attention and affective states. However, all of these approaches require dedicated sensors for signal collection. The cost, availability and portability of sensors may prevent the wide adoption of such technologies in large scale in the near future.

Mind Wandering (MW), or zoning out, is ubiquitous in both learning and everyday activities. In a large scale study involving 2240 adults, Killingsworth et al [10] discovered that MW occurred in 46.9% of the everyday random samples. Researchers have attempted to automatically detect MW in learning environments using various signals, such as pitch features in speech dialogues [4], eye fixation time and locations [1], skin conductance and skin temperature features [2]. In this paper, we show that it is possible to design an easy to learn and intuitive to use camera-based interface on today's mobile phones, to capture learner's heart rates and detect their MW states implicitly during MOOC learning without any hardware modification.

3 The Design of AttentiveLearner

The AttentiveLearner mobile client has four unique components when compared with today's MOOC mobile apps: 1) a tangible video control channel; 2) an implicit heart rate sensing module, 3) an on-screen AttentiveWidget visualizing real-time states of video control and heart rate sensing; and 4) an algorithm that infers learners' attention states (MW or not) from heart rate signals captured.

3.1 Tangible Video Control

In AttentiveLearner, the camera lens on the back of mobile phones is used as the "play" button for video/media control (Fig. 1 right). A learner uses his/her finger to cover and hold the camera lens to play an instructional video. Uncovering the lens will pause the video. We used the *Static LensGesture* detection algorithm in [21] to detect lens covering actions (sensitivity parameters can be adjusted to accommodate inadvertent finger jittery). The user independent detection algorithm can achieve an accuracy of 97.9% in

different illumination conditions at the speed of 2.3ms per estimate [21]. Our benchmarking results and informal tests also show that the algorithm is accurate and responsive as a video control channel. Anecdotally, users reported that on-lens gesture based video control is easier to use than traditional on-screen touch widgets for two reasons: 1) the edge/bezel of the camera optical assembly can provide natural tactile feedback to users' index finger; 2) In landscape mode, which is common in video watching, users can play or pause the lecture video when holding a mobile phone with both hands (Fig. 1, left). To overcome inadvertent "finger jittery", we keep playing video for 4.5 seconds and then pause even if the lens is uncovered.

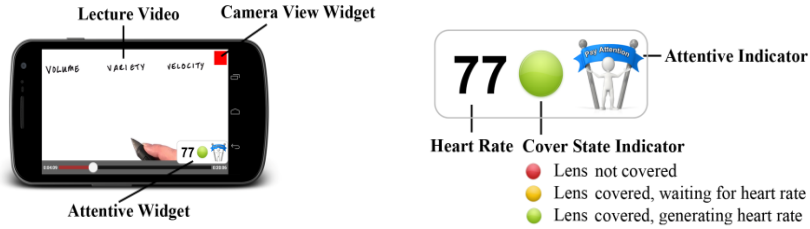

Fig. 2. The AttentiveWidget Interface

3.2 Implicit Heart Rate Sensing

In addition to video play back control, AttentiveLearner also captures learners' heart rates *implicitly* during learning by monitoring the fingertip transparency changes captured by the back camera. This technique is essentially commodity camera based Photoplethysmography (PPG) sensing. The underlining theory of PPG sensing is: in each cardiac cycle, the heart pumps blood to capillary vessels and changes the transparency of the corresponding human body parts, including the lens covering fingertip. These transparency changes correlate directly with heart beats and can be detected by the built-in camera lens when it is covered by the learner's fingertip. We used *LivePulse* [9], a heuristic based peak counting algorithm, to measure heart rates. The *LivePulse* algorithm is accurate (+/- 2 beat per minute when compared with a medical grade oximeter), robust, and can run efficiently in mobile device in real time.

3.3 AttentiveWidget

We designed an on-screen widget (Fig. 2) to visualize finger covering states, real time heart rates, and attention states (MW or not). The AttentiveWidget disappears if the system detects that the learner is in a no MW state for three minutes. The widget can be dragged and dropped around the screen or explicitly toggled by double tapping.

3.4 Mind Wandering Detection

By extracting learners' heart rates during MOOC learning on unmodified mobile phones in real time, we have the opportunity to infer important cognitive states such as stress levels, affective states, and attention states in learning. We focus on the de-

tection of MW in this paper and plan to infer and incorporate other cognitive states in our learning system in the future. Although existing research exists that uses heart rate and heart rate variability signals to improve learning, AttentiveLearner is the first to achieve heart rate enhanced learning on unmodified mobile devices.

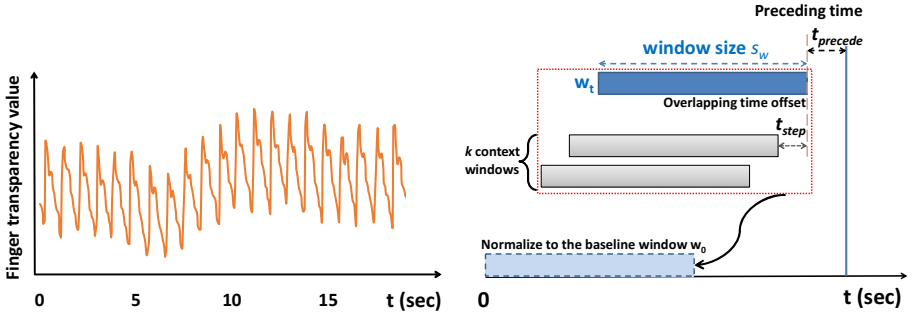

Fig. 3. Feature extraction in PPG signals (**left**: 20 seconds of PPG signal captured from mobile camera during video watching; **right**: using multiple moving windows for feature extraction)

We extracted two types of features, *heart rate features* and *lecture content features* for MW prediction. *Heart rate features* were extracted from multiple, overlapping *context windows* imposed on real time PPG readings (Fig. 3 left) before the time of prediction (Fig. 3 right). We extracted 12 dimensions of heart rate feature from each context window. The 12 dimension of features are: 1) AVNN (average heart rate); 2) SDNN (temporal standard deviations of heart beats); 3) pNN50 (percentage of adjacent heart beats with a difference longer than 50 ms); 4) rMSSD; 5) SDANN; 6) SDNNIDX; 7) SDNNIDX / rMSSD; 8) LF: low frequency (0.04 – 0.15 Hz); 9) HF: high frequency (0.15 – 0.4 Hz); 10) LF / HF; 11) totalPSD (total power spectral density); 12) MAD (median absolute deviation). The detailed definitions of features 2 to 7 can be found in [19], and the detailed definitions of features 8 to 12 are in [8], [19]. All of these features (except MAD) are based on heart rate variability features which are used by many researchers in heart rate signal related studies. We have tried multiple context window numbers, sizes, overlapping time, and preceding time offsets when training the classifier and we defer the details to the evaluation session.

We split lecture videos into equal-length, non-overlapping *content windows.* We extracted 7 dimensions of *lecture content features* from each *content window*. The 7 features are: 1) Lecture style (pure slide[1] or Khan-style[2]); 2) Duration of the current page; 3) Duration of the previous page; 4) Speech rate (words/min); 5) speech rate (words/min) of the previous page; 6) Cosine similarity between current and the previous page (Bag-of-words representation of the transcribed lecture text); 7) Cosine similarity previous two pages. A page is a slide (slide-style) or a small clip (Khan-style).

We applied feature rescaling and feature selection techniques to raw features above before training the classifiers. We report the detailed parameter selection and experimental results in the next section.

[1] Slide-style: slides are shown full screen and instructors' voice is played in background.
[2] Khan-style: instructors are facing front with handwriting notes as transparent overlays.

3.5 Implementation

AttentiveLearner was written in Java for Android 4.1. We used the LensGeture algorithm for lens-covering detection and LivePulse algorithm for extracting heart rates from fingertip transparency images. We used WEKA to train and optimize the classifiers. The final prediction algorithm (KNN) can run in real time on mobile devices.

4 Evaluation

4.1 Participants and Procedure

We have recruited 24 participants (5 females) between 22 and 31 years old (μ=25.2, σ=2.3) in our study. All the participants were graduate students in a local university. We use a within-subjects design and all the participants learned two MOOC lectures in the study. One was a 21-minute lecture on Hadoop (Khan-style) with 24 quiz questions; the other was a 23-minute lecture on R programming (slide-style) with 19 quiz questions. All subjects had little or no knowledge about the two topics used. The order of the two lectures was randomized. We used a Google Nexus Galaxy smartphone running Android 4.1 in the experiment.

We ran a tutorial session and collected a background questionnaire at the beginning of each session and then followed by presenting two MOOC lectures. Participants were required to complete corresponding quizzes after each lecture and there was a 5-minute break between lectures. Finally, the participants took an exit survey.

During learning, we used auditory probes [1,2] to figure out whether the participant was MW. After hearing an audio beep, the subjects report verbally "Yes" or "No" to indicate whether they were MW the moment before the probe. Auditory probes were triggered randomly at a 3 minute mean interval, and at the end of each page.

Table 1. Number of PPG sampling and frames/second of each subject

Signal	Average	SD	Max	Min
# of PPG samples	21,267.4	1,916.5	23,845	17,840
Sampling rate (fps)	16.1	0.5	16.6	14.7

In total, we collected 991 responses to auditory probes and 227 (22.9%) responses were MW. This ratio is similar to previous study (24.4%) in comprehensive reading [2]. The average accuracy of quiz questions was 78.3%. The average sampling rate 16.1 Hz (Table 1) was lower than the 30Hz normal camera frame rate, we attribute this to the extra CPU cycles used in video decoding and play back. Participants covered the lens of the camera 99.2% of the time during MOOC learning (min = 94.1%, max = 100%, σ = 1.4%).

4.2 Classifier Training

Five supervised machine learning algorithms were used in this study. The classifiers were K nearest neighbors (*KNN*), Gaussian mixture model (*GMM*), support vector machine with linear kernel (*SVM*), logistic regression with lasso regularization (*LogReg*), and local outlier factor (*LOF*). *LogReg* was trained by LibLinear and all other models were trained by WEKA. We also tested *SVM* with nonlinear kernels, but preliminary results showed their performances were worse than the linear kernel.

We used the leave-one-participant-out method to ensure that data from each participant was exclusive to either the training or testing set. As a result, all the results reported were user-independent.

Feature selection was performed to remove correlated features and those did not have sufficient discrimination power. This technique can restrict the model complexity and ensure sufficient speed when running on mobile devices.

We tried to use different parameter combinations for both feature extraction and classifiers. The optimal combination was the one giving the best average Kappa over all subjects. To be specific, we have tried 3 different context window numbers (1, 3, 5) × 4 context window widths (30s, 60s, 90s, 120s) × 4 context window overlaps (5s, 10s, 30s, 60s) × 3 preceding time values (1s, 2s, 5s) × model specific parameters. The model specific parameters are: *KNN* (number of nearest neighbors: 1, 3, 5), *GMM* (number of clusters: 2), *SVM* (feature weight for the MW class: 1, 3, 5), *LogReg* (feature weights for the MW class: 3, 5, 10), and *LOF* (number of neighbors: 7, 10, 20).

In summary, we extract $7 + 12\ (k + 1)$ dimensions of features where k is the number of context windows. We used information gain based feature selection to select the top 5 features to train classifiers.

5 Results and Discussions

Table 2 shows the MW prediction performance, i.e. predicting whether a participant was MW at a moment or not. The KNN classifier (K=5) led to the best overall accuracy (71.22%) and kappa (0.22). This performance is comparable with existing systems that rely on acoustic-prosodic features by Drummond and Litman [4] (learner dependent model, accuracy = 64.3%), eye gaze fixation features by Bixler and D'Mello [1] (learner independent model, accuracy=72%, kappa =0.28), and skin conductance and skin temperature features by Blanchard et al. [2] (learner dependent model, kappa=0.22). It is worth noting that our performance was achieved on today's mobile phones *without any hardware modifications*.

We also explored the feasibility of predicting learners' question-answering performance, i.e. determining whether a participant will make an error in the follow-up quiz based on heart rate signals when the topic was first mentioned in the lecture video (Table 3). The GMM classifier achieved the best kappa (0.22) with an accuracy of 65.14%. Although such accuracy can be considered to be moderate at best, it can be used to provide adaptive reviewing exercises to encourage learners practice on topics they didn't pay enough attention to during learning [18]. E.g., when using the LogReg model (highest recall = 74.69% in Table 3), AttentiveLearner can recommend learners to review around 58.59% of the lesson (rather than the whole lecture) in order to cov-

er all topics learners may make mistakes. In other words, AttentiveLearner has the potential to save around 41.41% of reviewing time when compared with a full review.

Table 2. Mind Wandering detection performance. Standard deviation in parenthesis.

Model	Precision	Recall	Accuracy	Kappa
LOF	30.06 (24.8)	23.45 (21.1)	70.51 (18.6)	0.08 (0.18)
GMM	33.51 (13.4)	**65.00 (21.7)**	60.15 (12.9)	0.18 (0.15)
KNN	**40.00 (24.3)**	40.99 (22.0)	**71.22 (10.8)**	**0.22 (0.22)**
LogReg	28.80 (15.3)	42.18 (13.4)	64.08 (09.4)	0.11 (0.13)
SVM	29.55 (12.9)	47.14 (18.6)	62.73 (09.5)	0.12 (0.13)

Table 3. Quiz error prediction performance. Standard deviation in parenthesis.

Model	Precision	Recall	Accuracy	Kappa
LOF	37.25 (29.1)	20.77 (16.6)	66.02 (14.0)	0.07 (0.2)
GMM	44.35 (20.2)	52.88 (22.3)	65.14 (10.0)	**0.22 (0.16)**
KNN	**44.80 (31.0)**	32.80 (18.3)	**68.13 (9.6)**	0.17 (0.13)
LogReg	36.47 (18.3)	**74.69 (16.0)**	55.05 (14.6)	0.17 (0.16)
SVM	37.06 (18.7)	74.32 (18.7)	54.79 (16.7)	0.16 (0.17)

Fig. 4 shows aggregated MW histogram of 24 subjects over two lectures. We have normalized cross-bin MW events to avoid biases. For example, if a learner has 2 MW events at the 6th and the 20th minute respectively. Each MW event will contribute ½ counts for each moment in the histogram. In the Hadoop lecture, the MW events peaked at around the 6th minute when discussing several open question. The second peak was around the 14th minute when the instructor was teaching the 2nd longest page (3.2 min) in this lecture. The three most frequent MW moments in R programming (the 6th, 13th-16th and 20th minute) were the three longest pages of the lecture, discussing Input (2.4 min), Matrices (2.7 min) and Factors (4.6 min) respectively.

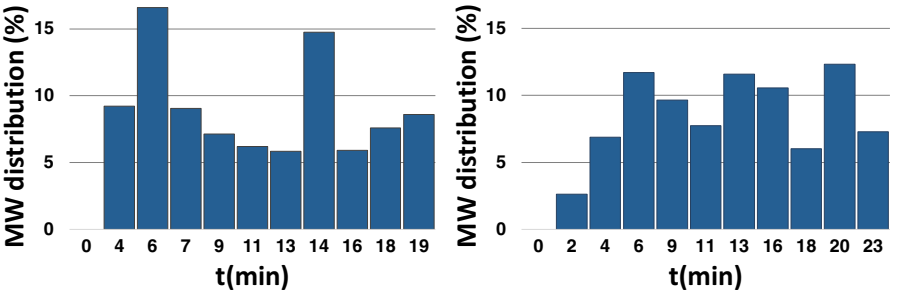

Fig. 4. MW histogram of the Hadoop lecture (left) and the R lecture (right)

6 Conclusions and Future Work

We presented AttentiveLearner, an intelligent mobile learning system optimized for consuming lecture videos in Massive Open Online Courses (MOOCs) on today's smartphones without any hardware modifications. In a 24-participant study, we found that AttentiveLearner can extract heart rates reliably from noisy image frames captured by mobile cameras and that it can be used to predict both learners' "mind wandering" events in MOOC sessions and their performance in follow-up quizzes.

Given the scale and scope of the current study, our current efforts should be treated as a "proof-of-concept" towards follow-up research work in the future. First of all, we plan to increase both the number of participants and the number of learning sessions in the follow-up studies. Second, the prediction performance reported here was based on offline benchmarking rather than live measurement. Third, we plan to use the prediction model to provide intelligent learning interventions on mobile devices. We plan to focus on features that can tolerate reasonable levels of false predictions, such as adaptive reviewing, non-intrusive MW alerting, etc. Fourth, we also plan to explore instructor side visualization interfaces. We hope that an instructor side interface could answer questions like a) Did most students keep up when I was teaching concept X? b) Did my joke "wake up" the students? or c) Were students bored by the end of the lecture? Considering that only PPG compatible live body parts such as fingers can be used to operate AttentiveLearner, AttentiveLeaner may take "virtual attendance" for instructors, addressing in part one of the major concerns in flipping a course.

Acknowledgements. We thank Andrew Head, Chris Schunn, Chris Thomas, Wenchen Wang, Zhijie Wang and Xiang Xiao for the help and support. We also thank anonymous reviewers for the constructive feedback.

References

1. Bixler, R., D'Mello, S.: Toward fully automated person-independent detection of mind wandering. In: Dimitrova, V., Kuflik, T., Chin, D., Ricci, F., Dolog, P., Houben, G.-J. (eds.) UMAP 2014. LNCS, vol. 8538, pp. 37–48. Springer, Heidelberg (2014)
2. Blanchard, N., Bixler, R., Joyce, T., D'Mello, S.: Automated physiological-based detection of mind wandering during learning. In: Trausan-Matu, S., Boyer, K.E., Crosby, M., Panourgia, K. (eds.) ITS 2014. LNCS, vol. 8474, pp. 55–60. Springer, Heidelberg (2014)
3. Calvo, R.A., D'Mello, S.: Affect detection: an interdisciplinary review of models, methods, and their applications. In: IEEE Transactions on Affective Computing, vol 1, pp 18–37. IEEE Press, New York (2010)
4. Drummond, J., Litman, D.: In the zone: towards detecting student zoning out using supervised machine learning. In: Aleven, V., Kay, J., Mostow, J. (eds.) ITS 2010, Part II. LNCS, vol. 6095, pp. 306–308. Springer, Heidelberg (2010)
5. Fisher, D.: Warming up to MOOCs. http://chronicle.com/blogs/profhacker/warming-up-to-moocs/44022
6. Fowler, G.A.: An Early Report Card on Massive Open Online Courses. The Wall Street Journal (2013)

7. Guo, P.J., Kim, J., Rubin, R.: 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, New York (2014)
8. Haapalainen, E., Kim, S., Forlizzi, F.J., Dey, K.A.: Psycho-physiological measures for assessing cognitive load. In: Proceedings of the 12th ACM International Conference on Ubiquitous Computing, pp. 301–310. ACM, New York (2010)
9. Han, T., Xiao, X., Shi, L., Canny, J., Wang, J.: Balancing accuracy and fun: designing engaging camera based mobile games for implicit heart rate monitoring. In: CHI 2015 Human Factors in Computing Systems. ACM, New York (2015)
10. Killingsworth, M.A., Gilbert, D.T.: A Wandering Mind is an Unhappy Mind. Science **330**(6006), 932 (2010)
11. Kim, J., Guo, P.J., Seaton, D.T., Mitros, P., Gajos, K.Z., Miller, R.C.: 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, New York (2014)
12. Kop, R., Fournier, H.: New Dimensions to Self-Directed Learning in an Open Networked Learning Environment. International Journal of Self-Directed Learning **7**(2), 2–20 (2011)
13. Malik, M., Bigger, J.T., Camm, A.J., Kleiger, R.E., Malliani, A., Moss, A.J., Schwartz, P.J.: Heart rate variability: standards of measurement, physiological interpretation, and clinical use. European heart journal **17**(3), 354–381 (1996)
14. Monserrat, T., Zhao, S., Li, Y., Cao, X.: 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, New York (2014)
15. Parr, C.: Not Staying the Course, Times Higher Education (2013)
16. Risko, E., Buchanan, D., Medimorec, S., Kingstone, A.: Everyday attention: Mind wandering and computer use during lectures. Computers & Education **68**, 275–283 (2013)
17. Smallwood, J., Schooler, J.W.: The restless mind. Psychological Bulletin **132**(6), 946–958 (2006)
18. Szafir, D., Mutlu, B.: Artful: adaptive review technology for flipped learning. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1001–1010. ACM, New York (2013)
19. Task Force of the European Society of Cardiology, & Task Force of the European Society of Cardiology: Heart rate variability: standards of measurement, physiological interpretation and clinical use. Circulation **93**(5), 1043–1065 (1996)
20. Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., Picard, R.: Affect-aware tutors recognising and responding to student affect. International Journal of Learning Technology **4**(3), 129–164 (2009)
21. 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, New York (2013)