How Presenters Perceive and React to Audience Flow Prediction In-situ: An Explorative Study of Live Online Lectures

WEI SUN*, Institute of Software, Chinese Academy of Sciences, China
YUNZHI LI*, University of Chinese Academy of Sciences, China
FENG TIAN†, Institute of Software, Chinese Academy of Sciences, China
XIANGMIN FAN†, Institute of Software, Chinese Academy of Sciences, China
HONGAN WANG, Institute of Software, Chinese Academy of Sciences, China

The degree and quality of instructor-student interactions are crucial for students' engagement, retention, and learning outcomes. However, such interactions are limited in live online lectures, where instructors no longer have access to important cues such as raised hands or facial expressions at the time of teaching. As a result, instructors cannot fully understand students' learning progresses. This paper presents an explorative study investigating how presenters perceive and react to audience flow prediction when giving live-stream lectures, which has not been examined yet. The study was conducted with an experimental system that can predict audience's psychological states (e.g., anxiety, flow, boredom) through real-time facial expression analysis, and can provide aggregated views illustrating the flow experience of the whole group. Through evaluation with 8 online lectures ($N_{instructors} = 8, N_{learners} = 21$), we found such real-time flow prediction and visualization can provide value to presenters. This paper contributes a set of useful findings regarding their perception and reaction of such flow prediction, as well as lessons learned in the study, which can be inspirational for building future AI-powered system to assist people in delivering live online presentations.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**; • **Applied computing → Computer-assisted instruction**.

Additional Key Words and Phrases: flow, facial expression analysis, live online lectures

ACM Reference Format:
Wei Sun, Yunzhi Li, Feng Tian, Xiangmin Fan, and Hongan Wang. 2019. How Presenters Perceive and React to Audience Flow Prediction In-situ: An Explorative Study of Live Online Lectures. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 162 (November 2019), 19 pages. https://doi.org/10.1145/3359264

1 INTRODUCTION

Technologies are reshaping modern education around the world. One good example is the rapid growth of online learning, which transcends the traditional barriers of physical location access

*Both authors contributed equally to the paper.
†Contact authors

Authors' addresses: Wei Sun, Institute of Software, Chinese Academy of Sciences, Beijing, China, sunwei2017@iscas.ac.cn; Yunzhi Li, University of Chinese Academy of Sciences, Beijing, China, liyunzhi15@mails.ucas.ac.cn; Feng Tian, Institute of Software, Chinese Academy of Sciences, Beijing, China, tianfeng@iscas.ac.cn; Xiangmin Fan, Institute of Software, Chinese Academy of Sciences, Beijing, China, xiangmin@iscas.ac.cn; Hongan Wang, Institute of Software, Chinese Academy of Sciences, Beijing, China, tianfeng@iscas.ac.cn.

2573-0142/2019/11-ART162 $15.00
https://doi.org/10.1145/3359264

and makes the high-quality learning materials and environments available at relatively low cost. Massive Open Online Courses (MOOCs) has attracted 81 million registered learners by year 2017 [2]. In addition to MOOCs which are usually asynchronous, synchronous methods such as live online lectures consist another trend in online learning, especially with the proliferation of high-definition cameras and high-speed internet during recent years. For example, over one million students in China participate in live online lectures provided by English learning platforms such as KMF [4] and New Oriental [5].

Different from traditional classrooms, instructors of live online lectures no longer have access to important cues of audience feedback such as raised hands or facial expressions at the time of teaching. As a result, the fine-grained monitoring and modeling of the learning process, which is crucial to ensure information is delivered properly and to keep the audience engaged, is insufficient in such settings. Therefore, it is even more challenging for instructors to dynamically adjust their instructional discourse to achieve and maintain high levels of student involvement.

To address this challenge, researchers proposed a set of methods to help instructors gather feedback from students in both asynchronous and synchronous online learning settings. Nota Bene [44], WebAnn [10], and Mudslide [17] are interactive annotation tools to help instructors understand and locate learners' confusion points in learning materials. While useful, the requirement for learners' active annotations may lead to discrete and coarse-grained feedback. With the aim of introducing an implicit way of collecting real-time and continuous feedback, EngageMeter [18] used physiological sensors to collect electroencephalography (EEG) signals of the audience and provide presenters with feedback regarding audience's engagement and workload. However, the cost and availability of the dedicated sensors have become a major obstacle preventing the wide adoption of such technologies in real world scenarios. Besides, the study was conducted in physical conference rooms where presenters leveraged such information as a complement to face-to-face observations of their audience's behaviors. Recently, researchers also proposed approaches to automatically predict audience feedback through facial expression analysis, e.g., during TV [20] and movie [25] watching, as well as in online learning contexts [30, 31]. Our work was inspired and built upon such prior works. At the same time, we go beyond investigating *personal tracking* to exploring the usage of *aggregated information* from multiple users with the goal of obtaining real-time feedback from the whole audience.

The concept of *flow* [13–15] in positive psychology describes the optimal experience where skillful and successful action seems effortless, even when physical or mental energy is greatly involved. The theory of flow is inherently related to learning and it has been studied in numerous learning settings reported by the literature of educational psychology [32, 33]. Flow experience occurs simultaneously with high concentration, enjoyment, and interest in learning activities, thus is the ideal state for learners [32]. The most central condition for flow to occur is that the challenge of the activity is well matched to the individual's skills, otherwise the learner will be in either anxiety (resulting from high challenge but low skills) or boredom (resulting from low challenge but high skills). Previous studies have demonstrated that teachers' instructional discourse can greatly influence the occurrence of students' flow experience [32, 36], and we hypothesize that knowing about the flow experience of the audience can help instructors cultivate better learning environment when delivering live online lectures.

In this paper, we present an explorative study investigating how presenters leverage audience flow prediction when giving live-stream lectures, which has not been examined yet. The study was conducted with an experimental system that can predict audience's flow-related psychological states (e.g., anxiety, flow, boredom) through real-time facial expression analysis, and can provide aggregated views illustrating the flow experience of the whole group. Our study consisted of 8 real-world live-stream lectures. Findings stem from the analysis of the video recorded presenting behaviors and the interviews with presenter members. Our study shows how presenters perceive and react in real time to the fluctuations in audience flow experience predicted by the AI system. We contribute a set of useful findings regarding their usage patterns, as well as lessons learned in the study, which are inspirational for researchers as well as practitioners building such AI-powered system to assist people in delivering live online presentations.

2 RELATED WORK

2.1 Flow and Learning

In Csikszentmihalyi's theory, flow is an optimal psychological state that people experience when they are engaged in certain activities which would later bring high levels of mental satisfaction to the experiencer [13, 15]. While flow has been examined in the context of both short and long time intervals, it has often been used to describe an individual's state in a certain moment. In terms of the measuring method, the study of flow has been mainly conducted using the Experience Sampling Method (ESM) [19] since the past two decades. The ESM, which is a method that records the signal of participants at specific moments and then asks them to finish a brief questionnaire which includes some scaled questions about their experiences. When it comes to the area of education, flow typically refers to the optimal learning state where learners feel a great sense of motivation and engagement and can consequently achieve better learning outcomes [11]. One key characteristic of flow in learning is that the the skill level determines the challenge one faces, the learner being in anxiety state or boredom state when the task is too difficult or too easy, respectively [6, 9]. Prior studies [7, 42] have shown that knowing the relationship between the task challenges and students skills is helpful for instructors to develop appropriate strategies for the teaching process. Researchers have brought up several adaptive learning systems which aim at improving learning quality by customizing educational materials according to learner's particular needs and skills [22]. Inspired by prior findings, we hypothesize that knowing about the flow experience of the audience can help instructors cultivate better learning environment when delivering live online lectures. Our study results fill the knowledge gap on how presenters perceive and react to flow predictions generated by AI approaches in-situ, where face-to-face observation on the audience is no longer available.

2.2 Emotion Detection

Emotion detection is a popular research topic which is in the intersection of computer science, psychology, cognitive science, and it has become one of the most important aspects in Affective Computing [43]. Researchers in this field focus on designing and implementing algorithms and systems that can detect, process and interpret human emotions. For example, facial expressions are considered to be a critical and informative source for emotion detection [38] which have already been used to automatically predict audience interest in movies and television [20]. Besides, researchers in both Computer Vision community and Human-Computer Interaction (HCI) community have built Deep Convolutional Neural Network (DCNN) for image processing to do emotion classification and achieved remarkable performance [23]. Human speech is considered to be another valuable source for emotion detection. Researchers made several breakthroughs by applying modern machine learning and signal processing techniques to fully investigate the representative emotional information in human speech, such as the speed, tone and word usage [35]. In addition, physiological signals, such as skin conductance and heart rate variability (HRV) have been leveraged as informative inputs to detect human emotions [28, 39]. Multimodal emotion detection is also another increasingly popular way for emotion recognition by representing human emotion with facial expressions, body gestures, voice and physiological signals. For example, Pham [30, 31] employed a combination of implicit photoplethysmography (PPG) sensing and facial expression analysis (FEA) to predict viewers' attention, engagement, and sentiment when watching video advertisements on smartphones.

2.3 Obtaining Student Feedback in Online Education

It is widely believed that online education has many advantages, such as the low costs, convenient repetition of lectures, and flexible learning schedules and environments. However, when learners study online, gathering feedback about whether the learning materials were understood and how to improve them for future learners is not as straightforward as the traditional classroom environment. Both research community and online education application designers brought up several methods to provide post-hoc feedback for online instructors [21, 31, 34]. The application like InstFeedback [1] used explicit polling to help instructors understand learners' confusions about the learning materials. Alternatively, instructors can gather feedback by viewing discussion forums which are almost a standard feature of online learning platforms (e.g. Coursera, Udacity). Researchers also designed and implemented some online interactive annotation tools to help instructors understand the learner's confusion distribution over online reading materials and slides, such as WebAnn [10], nb [44], and Mudslide [17]. There are also some research [41] about developing helpful interactive tools for online knowledge transfer between instructors and novices. Prior research indicates that explicit feedback from learners can help instructors understand more about their learner's learning status. However, such method may lead to coarse-grained feedback. Live chatroom can be a valuable information source for instructors to observe learners' learning status, but it still needs students' active participation. Moreover, by only observing live chat records, it is almost impossible for instructors to estimate how many learners benefit from instructors answering a particular question and learners who are in the boredom are much easier to be overlooked for their passive participation.

Similar to the learning environment of live online lecture, knowledge-sharing live streaming's emerging and popularity has been noticed by many researchers. According to the previous research [16], although live streaming augment traditional online learning in many aspects, the existence of problems, like difficulty in keeping the learner engaged, could impede a highly-efficient and interactive online learning environment. StreamWiki [24], a tool which is developed to help knowledge-sharing live streaming viewers produce real-time text-based archives, Di [12] investigates how incorporating multimedia tools in live streaming affected interactive experiences between the teacher and students, forms both more interactive and post-hoc learning experience. While useful, the need of active participation of stream viewers reveals that implicitly gathering feedback from the audiences could be a potential direction of improvement.

3 THE EXPERIMENTAL SYSTEM

Before we describe the study, we would like to introduce the experimental system utilized in this work. Our system is composed of three parts, namely, learning state identification, web server processing, and feedback visualization.

As shown in Figure 1, the workflow of our system is:

- `Part 1`: Based on the user's facial expressions, which are implicitly collected by the camera, the learning state classifier identifies the user's learning state as one of the Boredom, Flow and Anxiety. Then each learner's state will be sent to the web server from the Learner's end.
- `Part 2`: The web server aggregates each learner's learning state and sends the result to Instructor's end.

- Part 3: Instructor's end receives the result of learners' learning states and plots them into real-time visual feedback.

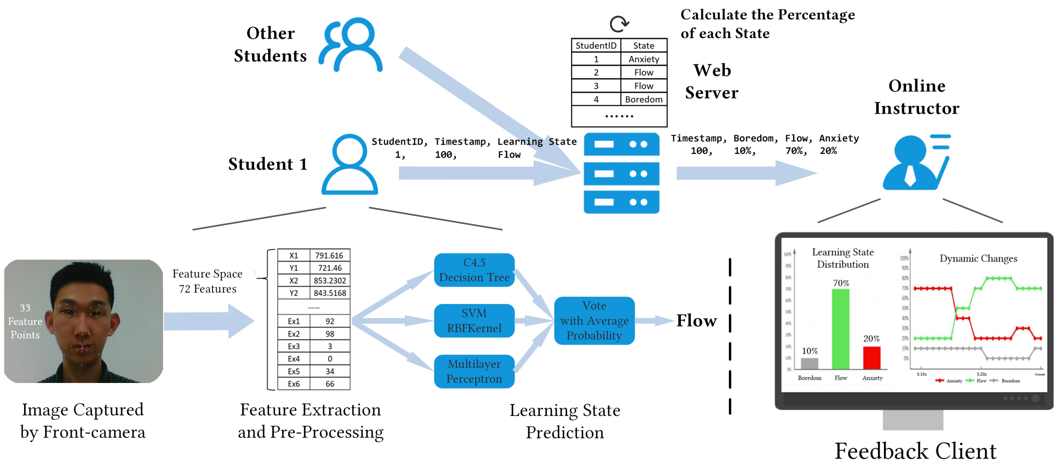

Fig. 1. Workflow of the experimental system in this study: 1) Based on the user's facial expressions, which are implicitly collected by the camera, a classification result will be generated after processing the data and sent to web server. 2) Web server will further analyze the result. 3) Instructor will then receive the visualized feedback.

3.1 Learning State Sensing Component

Based on the theory of Flow and previous research about emotion recognition [7, 22, 38, 42], we implement the learning state sensing component accordingly so that it can predict the learner's state.

Before the formal experiments, we conducted 2 preliminary studies to establish the learning state predictor, during which two 40-minute live online lectures were held and totally 15 learners participated. In each lecture, the video of each learner's facial expressions was recorded by the front camera on his/her laptop and the participants were asked to label their learning state as one of Anxiety, Flow or Boredom in every two minutes. After that, we used Affdex SDK [3] for video stream analysing in 30fps to extract 33 facial feature points' coordinates, 4 output emotion scores (Engagement, Attention, Joy and Disgust) and 2 action unit scores (Inner Brow Raise and Brow raise) as our feature space (totally 72 features: 6 scores and 66 feature point coordinates) for model training. Later in data pre-processing stage, we changed the coordinates into head-centred coordinates which is similar to previous Bailenson's work [8] for normalization.

For emotion classification, we were inspired a lot by previous research. For example, SVM with RBF kernel [29, 30] and Multilayer Perceptron [8, 37] with 2 hidden layers were used for emotion classification, by analyzing facial feature points. C4.5 algorithm was also used for emotion classification, by analysing multiple physiological signals [40] and achieved desirable performance. Therefore, we tried all these three methods in WEKA software package and further decided to use ensemble learning method by voting with the average possibility of all three algorithms, which was inspired by the idea of combining SVM and Decision Tree algorithm for emotion classification [26].

Throughout the whole procedure of model training and evaluation, we consistently built user-independent models and utilized the leave-one-subject-out cross-validation method for evaluation. The final results and comparison show that our ensemble method is indeed effective which achieves the best accuracy of 64.09% over all four methods. , as shown in Table 1. Hence, we built our final learning state prediction model for formal experiments based on the ensemble method of voting. To summarize, the pipeline of our learning state sensing component contains three main steps. (1) the student's facial image will be captured by the front-camera on his/her own laptop, then (2) the features will be extracted and pre-processed as we mentioned above, finally (3) the student's learning state will be predicted by our model and sent to the administration component for further analyzing and visualization, as shown in Figure 1.

Table 1. Accuracy (Acc) and Kappa of learning state prediction models, * indicates that compared with that model, our ensemble learning method achieves a significant improvement ($p \leqslant 0.011$).

	Decision Tree C4.5	SVM with RBFKernel	Multilayer Perceptron with 2 hidden layers	Ensemble Learning Vote with average Probability
Accuracy	60.97%*	62.21%*	63.52%	**64.09%**
Kappa	0.3079	0.1136	0.3174	0.3022

3.2 Administration Component

Composed of a web server and a database, the administration component undertakes the task of receiving, analyzing and saving the learning state data from each learner client, as well as presenting the result to online instructors via Feedback Client. To be more specific, the web server will receive every identified learning state from the Learning State Sensing Component and calculate the percentage of each state.

When the instructor starts the live online lectures, he/she inputs related information about the lecture and chooses a session type. We offer two different session types: 1) normal feedback session in which the feedback diagram consists of a bar graph and a moving line graph. 2) intervention session in which the Feedback Client will provide active intervention for instructors when the percentage of Boredom state or Anxiety state is equal to or higher than 50%. The database stores the raw data sent from each Learning State Sensing Component, every learning state distribution, active interventions and their timestamps.

3.3 Feedback Client

The Feedback Client refers to a web-based data visualization component of our whole system, which is used to provide the online instructor with the feedback of the learner's learning state in real-time. The calculated percentages, together with their timestamps, are sent to the front-end web Feedback Client from the Administration Component. The online instructor can get access to the real-time feedback diagrams by simply typing in the correct IP address through the browser.

The normal feedback session displays current percentages of each learning state in a three-column bar graph (grey: boredom, green: flow, red: anxiety) and a moving line graph with the percentage of each learning state over time, which are shown in Figure 2. The bar graph provides instructors with a quick sense of the distribution of learners' learning states, which can be interpreted by instructors in short glances. The moving line graph is intended to help instructors understand the dynamic change of learner's learning states in a holistic view. The intervention session will provide instructors with interventions when the percentage of Boredom state or Anxiety state is equal to or higher than 50%, with the aim of helping instructors reduce the expecting extra cognitive workload caused by the glances of feedback diagrams. The active interventions contain both a sound alarm and text prompt.

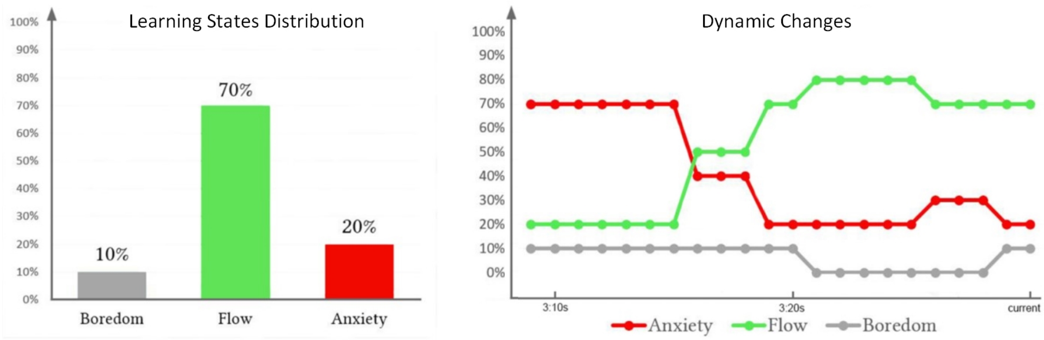

Fig. 2. Presenter's dashboard of our system. The number 70% means that 70% learners are in the Flow condition.

4 THE EXPLORATIVE STUDY

Throughout the process of our explorative study, totally 8 multi-condition lab-based live online lecture experiments were conducted with the aims of answering the following three questions under the scenario of live online lectures:

Q1: How often and when do instructors refer to learning state feedback?

Q2: If yes, what are the motivations behind and what kind of usage pattern do they have?

Q3: How does perceiving the feedback affect the instructors in their online teaching process? Specifically, we are interested in exploring the effect from the following three aspects: instructor's teaching pedagogy, emotion and workload.

4.1 Participants

We recruited 8 academic researchers (6 males, 2 females) from our research lab to give live online lectures as instructor participants. We consider 4 participants out of them as experienced instructors for the reason that they have an average 2-year's teaching experience in traditional classroom environment, and rich experience in giving online academic presentations. The rest of them are considered to be inexperienced instructors for they seldom teach or deliver presentations. We recruited 21 (15 males, 6 females) learner participants (some of them took lectures two or three times) who were all undergraduate students (Median age = 20.8, SD = 0.745) with diverse major backgrounds, such as computer science, biology, physics, material science, chemistry and math.

4.2 Procedure

We informed the instructor of the distribution of their learners' majors (e.g. 1 computer science, 2 biology, 2 chemistry and 1 math) three days before the lecture and asked the instructor to carefully prepare a 40 minutes' lecture in his/her area of expertise and divided it into 4 conditions (each condition lasts for about 10 minutes). In addition, to evaluate the effectiveness of Learner-to-Instructor feedback in live online lectures, before the lectures, the instructor was required to mark which slides were the possible difficult or boring points in his/her lectures.

The instructor was required to review his/her teaching slides half hour before the lecture started. Moreover, we briefed the workflow of our system and carefully introduced every element displayed in the presenter's dashboard and its meaning to the instructor. We offered two screens to the instructor. One was for displaying the teaching materials, such as the lecture slides and related videos, the other served as a secondary screen to display the feedback charts and text prompts.

After the instructor was familiar with the feedback charts, we introduced the whole evaluation procedure to the online instructor as follows:

There were 4 conditions in each lecture and each condition lasted for about 10 minutes.

- `condition 1`: In this condition, the online instructor will not receive any feedback from the learners.
- `condition 2`: In this condition, the secondary screen will display the three-column bar graph and the moving line graph to the online instructor.
- `condition 3`: In this condition, the online instructor not only has the access to the three-column bar graph and the moving line graph but also will receive active interventions (include the alarm sound and text prompts) from our system when the percentage of Boredom state or Anxiety state is equal to or higher than 50%.
- `condition 4`: In this condition, we deliberately exchange the percentage of Boredom state and Anxiety state which will be displayed on the secondary screen. In other words, the instructor will be provided with the wrong distribution and dynamics of these two states and receive wrong interventions. In addition, we will also send extra interrupting interventions to the instructor even when most of the learners are in the Flow state. However, condition 4's existence is **not informed** to the participants throughout the whole experimental procedure which means condition 3 and condition 4 are identical from their perspectives.

Particularly, we used the Latin Square Design to reduce the order effect in our experiments. More specifically, we conducted experiments with conditions shuffled to 4 different orders (1234, 2341, 3421, 4123). For one experiment of a certain order, one experienced instructor and one inexperienced instructor were engaged.

In order to eliminate the impact of different proficiency, all these instructors are first-time users of our system. During each lecture, we kept a video recording of the instructor's teaching process and recorded each timestamp of his/her glance at the presenter's dashboard. After the lecture finished, We told the instructor participants that what was the actual participant's feedback in condition 4. We also immediately confirmed the validity of each glancing at the presenter's dashboard with the instructor and investigated about the motivation and corresponding effect of receiving feedback to his/her emotion, cognitive workload and decision making on the teaching pedagogy. No information needed to be provided when they felt vague. Then, we asked the instructor to provide more details about other noticeable situations or anything related to the effect of the feedback. Finally, a survey was completed by each participated instructor to provide some subjective scores of the effect of Learner-to-Instructor feedback in live online lectures. Our experimental protocol was approved by the institutional IRB before the study.

5 RESULTS

In order to have a better understanding of the user experience, researchers would watch the lecture video together with the teachers upon finishing a class, annotating the point when the teachers refer to the system-provided information of feedback as well as the students' status. Meanwhile, discussion regarding the instant thoughts and improvement of teaching method would be conducted among teachers and researchers. Afterwards, two researchers categorized each response independently by conducting open-source coding and resolved any conflicting terms through mutual review. Figure 3 describes the observed behavior of a participant during a specific phase. By analyzing the video records and the post-hoc interview results, we summarized the instructor's usage pattern of Learner-to-Instructor feedback and gained some quantitative findings and qualitative insights about the effect of the feedback in live online teaching. Based on the questions we brought in the previous section, we divide this section into 3 main parts as follows. For brevity, we will refer lecturer X as TX in the following content.

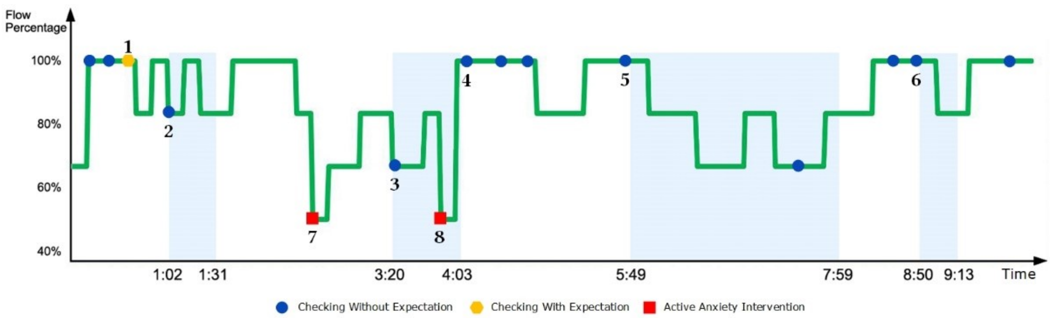

Fig. 3. Point 1 serves as a Checking With Expectation point. At this point, T1 raised a question to the whole class and actively check the feedback charts with the expectation that all learners are in the optimal Flow state. Point 2-6 all serve as Checking Without Expectation points within the Turn Page Window. At point 7 and 8, T1 receives active anxiety intervention which are caused by that half of the learners are in the Anxiety state. Therefore, the feedback checking actions which come after receiving the interventions are considered as passive checking behavior.

To summarize about the instructor's usage pattern of our learning state feedback system, especially from the perspectives of timing and motivation, we plot the recorded behavior log and the corresponding feedback data of T1's condition 3 in Figure 3 as an example to help illustrate the timing and motivation of instructor's feedback checking behavior in a holistic view.

5.1 Do the instructors actually refer to our system for learning state feedback?

This section serves as the answer for **Q1:** Do the participated instructors refer to our system for learning state feedback? **Feedback checking** refers to the instructor's behavior of checking the learner's learning states in the presenter's dashboard of our feedback system. Through our system evaluation, we find that the participated instructors indeed check our system for learning state feedback. More specifically, throughout our experiments of 8 instructors, totally 282 feedback checking actions are confirmed by the participated instructors, as shown in Figure 4, and their average feedback checking frequency is 1.175 times per minute (SD = 0.528).

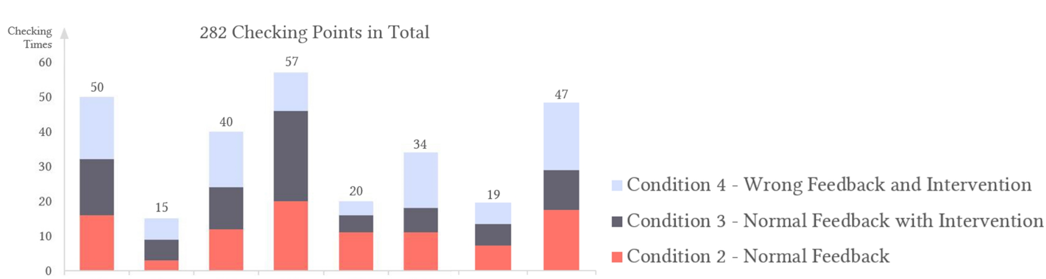

Fig. 4. The statistical result of feedback checking.

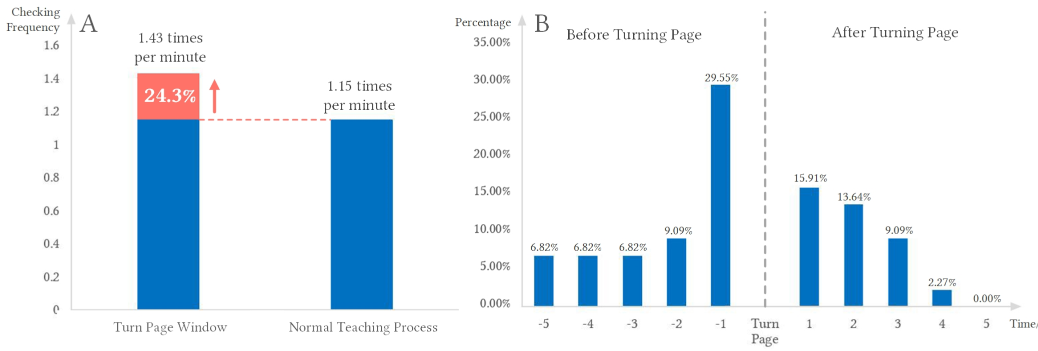

Fig. 5. A: The differences between the feedback checking frequency in Turn Page Window and Normal Teaching Process; B: The distribution of feedback checking action over the Turn Page Window

5.2 Motivation and Timing of Feedback Checking

After the conclusion that instructors indeed refer to our system for learner's feedback, we further investigate on the motivation and timing behind the instructor's feedback checking behaviour. This section serves as the answer for **Q2:** If yes, what are the motivations behind and what kind of usage pattern do they have?

5.2.1 Timing of Feedback Checking. By analyzing the timestamp of each feedback checking action, we find that instructors are highly likely to check the presenter's dashboard before or after they turn pages of their slides. Here, we define a 10-second time period including 5 seconds before page turning and 5 seconds after page turning as the Turn Page Window. As the Figure 5.A shows, we further compare the differences between the feedback checking frequency in Turn Page Window and Normal Teaching Process (the rest of time in the whole teaching process excluding the Turn Page Window), the higher checking frequency can also illustrate that instructors tend to check the feedback dashboard more frequently in the Turn Page Window. Additionally, the distribution of checking behaviour over the Turn Page Window shows that nearly 50%'s related feedback checking behavior happened in one second before or after the page turning, as shown in Figure 5.B.

5.2.2 Motivation of Feedback Checking. According the motivations behind each feedback checking behaviour, we divide feedback checking behaviour into two categories: passive checking and active checking. The passive feedback checking action refers to the checking behavior which is caused by the active intervention. Throughout 8 experiments, totally 30 interventions were sent to the instructors. Except one of them was overlooked by T1, the rest of them all caused further passive feedback checking of the instructor. Then according to the self-reported motivations which we gained through our post-hoc interviews, we further divide the active checking behavior into two sub-categories: (i) checking with expectation and (ii) checking without expectation, as follows.

Checking with expectation refers to the active checking action done by the instructor with the expectation that the distribution of the learners' current learning states is the same as or similar to his/her own judgment. Through the post-hoc interviews, we found that their judgments are usually established on:

- **Basis 1:** The teaching experience, such as the learner's own judgment of the matching degree between the learning materials and the learner group.
  Here are some typical examples with the motivations given by the participated instructors.

`Expectation for Boredom`: In lecture 5 when T5 was on page 15 of the slides, he was illustrating a knowledge diagram to the learners and he actively checked the feedback charts at 3:48s in condition 3. He stated "When designing the slides, I could imagine that learners are not willing to listen carefully to a complicated diagram full of codes and formulas. However, there was no other better way to illustrate all those knowledge points together. Therefore, I checked the charts to help me decide whether to continue illustrating more details or go on to the next slide."
`Expectation for Flow`: In lecture 4 when T4 was on page 17 of the slides, he was talking about an example about the semantic segmentation in natural language processing, and he actively checked the feedback charts at 8:26 in condition 3. He stated "This is a very interesting example for the funny misunderstanding caused by the wrong semantic segmentation. It should be attractive to all learners. Therefore, I checked the charts to see whether they were engaged in my lecture or not.)
`Expectation for Anxiety:` In lecture 2 when T2 was illustrating an algorithm on page 5 of the slides, he actively checked the feedback charts at 4:12 in condition 3. He stated "This is a very difficult algorithm, especially for students who do not major in computer science. I checked the feedback charts to see whether I needed to further elaborate this algorithm or quickly move on to the next page."

- **Basis 2**: The expected change caused by his/her previous real-time adjustment of teaching. For example, in lecture 8 when T8 was on page 28 of the slides, she received an active intervention when most students were in the Boredom state at 4:25s in condition 4. After 30 seconds, T8 actively checked the feedback charts at 4:55s in condition 4. She stated "I was alarmed by the system that most students felt boredom before. I then switched to another more complicated and difficult way of expression to see whether this adjustment is effective or not."

Checking without expectation refers to the active feedback checking behavior without any expectation for the possible distribution of the learners' learning states. By summarizing the interview results of the participated instructors, we found that there are two representative motivations behind, as follows:

- **Motivation 1:** Understanding learner's learning state after finishing teaching one certain knowledge point or a whole slide.
  All of the participated instructors showed this active checking action during his/her lecture. For example, on page 1 of the lecture 3's slides, T3 actively checked the presenter's dashboard before moving to the next slide. He stated "The background knowledge was about to be finished, therefore I referred to the feedback charts to see whether they (the learners) are still listening or not." T1 also clearly illustrated about his motivation on checking the presenter's dashboard at 3:02s in condition 2 before turning the slide, he stated " I just finished this whole slide and naturally hoped to understand my students learning state at that moment."
- **Motivation 2:** Evaluating the effect of his/her own poor teaching performance through learner's status feedback.
  For example, on lecture 7's page 10, T7 actively checked the feedback charts when she got stuck on explaining one certain knowledge point. She stated "I was really nervous about getting stuck on this slide, so I hope to get feedback from them (the learners)." On lecture 1's page 42, T1 actively checked the feedback charts when explaining one certain knowledge point at 2:46s in condition 4. He stated "This part (of the lecture) was hard to illustrate clearly and I felt my expression was not coherent at that moment, so I checked their (the learners)

status to see how they reacted to this (poor teaching performance)." On lecture 2's condition 2, T2 actively checked the feedback charts at 5:28. He later stated "I felt a little bit tired at that time and I thought my teaching quality was not satisfactory. Therefore, I hoped to see whether they (the learners) were affected by this (unsatisfying teaching performance) or not."

5.3 Effect of Learner-to-Instructor Feedback

The objective of this section is to gain both qualitative and quantitative findings about the effect of perceiving the feedback to live online instructors from three main aspects: instructor's teaching behavior, emotion and workload. Table 2 describes the reported pros and cons of in-video prompting from instructors' perspective.

Table 2. Main pros and cons perceiving the feedback

	Pros	**Cons**
Teaching	Support real-time pedagogy adjustment	None
	Support post-hoc teaching adjustment	
Emotion	Bring Confidence and Engagement	Cause anxiety
Workload	None	Introduce extra workload

5.3.1 Effect on Teaching Behavior. After receiving the feedback of the learner's learning states, the instructors might repeat, emphasize or skip some of their teaching content, or adjust their way of expression accordingly which are all considered as Teaching Adjustment behavior. We found that:

Finding 1: Learner-to-Instructor feedback is useful for real-time teaching adjustment. All eight instructors found the real-time feedback was useful and informative for their real-time teaching adjustment. In condition 1 (no feedback), no significant adjustment behavior was observed or self-reported. However, in the conditions with feedback, the average adjusting frequency of all participated instructors was about 2.89 times per condition. More specifically, the average adjusting frequency of experienced instructors was about 7 times per lecture and the one of inexperienced instructors was about 10 times per lecture which raised by 42.9%. For example, T1 decided to bring up more examples to help him illustrate one notion when he found some learners were in the boredom state. T3 decided to quickly illustrate certain knowledge and move on to another slide when he found that many learners were in the anxiety state. Although different instructors have different ways and preferences for adjustment, 8 out of 8 participated instructors gained the ability to make real-time teaching adjustment through our experiments with the help of the Learner-to-Instructor feedback.

Finding 2: Learner-to-Instructor feedback plays two different roles in helping instructors make real-time adjustments to their teaching.

- **Role 1:** the feedback can be used to help instructors find unrealized problems. Recall to Section 4.2, before the experiments, we asked each participated instructor to think in advance about the possible difficult or boring points in his/her lectures. Throughout our 8 experiments, all instructors (8 out of 8) found unexpected Anxiety or Boredom points in his/her lectures. For example, T1 got an Anxiety intervention in condition 3 and he mentioned "I didn't think this slide will cause any confusion to learners since it just simply contains some conceptual knowledge. However, I was alarmed by the system (most learners were in the anxiety state) and considered that this slide was not very important, I then decided to quickly skip it."
- **Role 2:** the feedback can be used as a support for instructors' real-time decision-making. To be more specific, in traditional classroom environment, experienced instructors usually have alternative teaching strategies and they could flexibly change their teaching content

according to their learner's feedback. By providing instructors with the feedback of learner's learning states, online instructors are able to achieve similar actions. For example, T2 decided to explain more details about one knowledge point when he found that all learners were in the Flow state. He mentioned in post-hoc interviews "I was about to move on to the next slide. However, after seeing that all students are in the optimal states, I decided to illustrate more details of this point." Another participated instructor, T5, changed her teaching content by illustrating some more interesting examples when she received a boredom intervention at 3:27s in condition 4. In the post-hoc interview, she stated "Before I received the intervention, I was hesitant about whether to make the lecture more interesting. This intervention just arrived in time and helped me to make the decision!"

Finding 3: Learner-to-Instructor feedback is useful for post-hoc adjustment.
During the interview, all instructors stated that they were willing to make post-hoc adjustment after the lecture by aligning the feedback data with their slides. Particularly, during the teaching process of T1, he marked the slide and the corresponding distribution of learners' learning state after he received each intervention. He later stated "At that moment (when he received the intervention), I was not capable of making an optimal real-time adjustment. Therefore, I decided to mark down that point and carefully reflect on it after the lecture." Another instructor T7 brought up her unique point. She stated that the real-time feedback was also benefit for helping the instructor improve his/her own teaching skill. "Every time I found my adjustment is effective, I sort of gained some experience for teaching. This would help me perform better in the future." She stated.

5.3.2 Effect on Emotion. After the live online lecture had been finished, as we confirmed each feedback checking point with the instructor, we also asked the instructor to provide his/her emotional status after knowing each feedback of learner's learning state. Considering the forgetting phenomenon of the emotional memory, especially for neutral emotion [27], only the positive and negative emotional memory with high confidence was marked. Therefore, in the totally 282 confirmed feedback checking points, 48 points were marked as positive emotion point which means the instructor felt positive after he/she checked the feedback charts. Another 44 points were marked as negative emotion points which meant the instructor was in a negative mood after the feedback checking. By analyzing the results, we got several insightful findings as follows.

Finding 1: Learner-to-Instructor feedback can make the instructor feel more confident or engaged during live online lectures.
Totally 6 out of 8 instructors mentioned that the feedback has made them more confident when they are teaching. For example, when T1 talked about one positive emotion point, he stated "This part is newly added content and this is also my first time to teach related knowledge. Therefore, I am eager to know their (the learners) learning states. After I saw that all students are in the optimal flow state, I felt more confident about my teaching performance and the feedback did promote the teaching process." Particularly, 2 instructors mentioned that with the help of feedback teaching online became more interactive and they felt more engaged in teaching. For example, T5 stated "Without the feedback, online teaching is more likely to be individual work. However, the feedback could not only support me to make decisions but also encourage me to find the teaching content or way of expressions that students are interested in."

Finding 2: Learner-to-Instructor feedback could make the instructor feel nervous or disappointed during live online lectures.
4 out of 8 instructors mentioned that the emergence of negative states (Boredom and Anxiety) would make them feel nervous and anxious about their teaching. For example, T2 reported that he was very nervous after he saw that one-third of the learners were in the anxiety state, he stated

"This negative feedback information interfered with my normal pace of teaching. I then felt anxious about what to say and I was also worried about the students' engagement in the remaining lecture." Besides, 2 instructors reported that they felt upset when the learner's learning states were not consistent with their expectations. T6 said that she felt extremely upset when she found the many learners were still in the boredom state after she repeated emphasis several times. T8 also stated "I felt disappointed because I found that they (the learners) were not interested in the 'shining points' in my research at all."

Finding 3: Boredom state is more likely to bring negative emotions to instructors.

When we were studying about the causes of negative emotions, we found that the Boredom state is more likely to bring negative emotions to instructors than the Anxiety state, albeit they are all negative learning state in the notion of Flow. More specifically, the negative emotion rate of Boredom state is about 19.7% which is more than twice of the Anxiety state's rate which is 8.3%. T7's statement may help us understand this phenomenon. She stated that "To some extent, I considered the anxiety state to be a positive state and it may show that students are listening carefully to my lecture."

5.3.3 Effect on Workload. During the post-hoc interviews, we also asked the instructors to give objective opinions and scores about the potentially extra workload which is introduced by perceiving and reacting to the learning state feedback. By analyzing the instructors' behavior records and the interview results, we found that:

Finding1: Our feedback system will introduce extra workload to users.

After the lecture, the instructor rated about the extra workload on a 5-point Likert item (1 = no extra workload at all just as same as traditional classroom , 5 = very significant extra workload even affected the quality of lectures). The average score is 2.63 (SD = 0.992) which means our feedback system did introduce some extra workload to the users. Besides, the average checking frequency is 1.08 times per minute which could also be served as a piece of evidence that instructors need to pay extra attention to receiving the feedback information. Moreover, 4 out of 8 instructors mentioned that interpreting the feedback information will burden extra workload on them. T1 mentioned "If students are anxious, boring contents will distract student even more than when there's no anxiety. But this system does not provide enough information regarding the points where student get lost, and I can only try to guess which knowledge point is it that cause the confusion."

Finding 2: The bar graph is more helpful and intuitive than moving line graph in live online lectures.

We also asked the instructor to rate about the helpfulness of two feedback charts on a 5-point Likert item (1 = not helpful at all, 5 = extremely helpful). The average score of the bar graph is 4.33 and the average score of the moving line graph is 1.11. One-way ANOVA showed that (F =85.784, $p<0.001$) instructor are more inclined to check the results of the bar graph. By further asking the reason, T3 answered "I don't think I have mastered the ability to quickly interpret the meaning of the line graph. Therefore, I usually refer to the bar graph to simply understand the current distribution of students' learning state" Another instructor T1 mentioned "The line graph would be more helpful for post-hoc review because it will show us the change in the distribution of the students' learning states over time. And the bar graph is more intuitive and helpful than the line graph for real-time adjustments."

Finding 3: Active intervention could reduce the extra workload to some extent.

In total, we got 282 feedback checking in condition 2, condition 3 and condition 4. One-way ANOVA showed that active intervention did not affect the times of check by teachers in conditions 3 (F=0.061, p=.752) or 4 (F=0.061, p=.968). However, in post-hoc interview, 6 of the 8 instructors think the active intervention could reduce the extra workload while 2 out of them not. T3 stated "I do not have to double check the student status after knowing the mechanism of active reminders." We further compared the average checking frequency in these conditions. We found that for those instructors who think the intervention is helpful, their average checking frequency dropped about 23.3%, which means active intervention could help them focus more on their online teaching rather than constantly checking the feedback charts. However, for the other instructors who think the intervention would not help them reduce the extra workload, their average checking frequency raised about 65%. For this phenomenon, T4 stated " The appearance of intervention would cause additional anxiety to me. Thus, I tended to check the feedback charts more frequently after every time I received an intervention notice." The experimental results in this experiment did not show a significant decrease. It may be that our sample size is not large enough, and we will continue to explore this issue in future work.

5.4 Trust Issue between the instructors and our Human-AI collaborative teaching-support system

After each lecture, we asked the instructor to rate on a 5-point Likert scale about the consistency between the students' learning states indicated by each intervention (Boredom or Anxiety) and their expectation of learners' learning states (1 = highly inconsistent, 5 = highly consistent). As shown in Table 3, the instructors think condition 3's Correct Intervention is more similar to their expectation than condition 4's Wrong Intervention. According to our experiment procedure, two participated instructor's condition 4 (provide wrong feedback and interrupting interventions) were at the first ten-minute. During the interview, they all doubted the accuracy of our system and their average Likert score in condition 4 is 2 out of 5. T6 specially mentioned that she was highly questioned about the second intervention and therefore decided to ignore it. She stated " It is inconsistent with my experience. I have taught this part several times in classroom environment, personally speaking, I think nobody would consider this part to be a difficult one." However, for all other instructors whose condition 4 came after one or two conditions which provided the instructor with correct feedback, although their Likert score in condition 4 (2.67 out of 5) is significantly lower than the one in condition 3 (4 out of 5), they still kept making real-time adjustment according to the feedback and only 2 instructors reported they once doubted about the accuracy of the interventions they received. This phenomenon may suggest that instructors tend to be skeptical of one decision-support system and mainly dependent on their own judgment when they just start to use the system. However, when the instructor finds that the system is working well and the suggestions given by the system are reasonable or similar to their expectation for a period of time, they tend to trust the system and follow the suggestions even when they conflict with their own judgments.

Table 3. Average Likert scores of the eight instructors for two different intervention modes

	condition 3 Correct Intervention	**condition 4 Wrong Intervention**
condition sequence of 4123	4.33	2
condition sequence of 1234, 2341 and 3412	4	2.67

Here are some examples: when T4 was explaining about page 29 of his slides, the feedback chart showed that two students were in the anxiety state and T4 decided to add more explanation about the content. During the interview, T4 said "I am confused about the feedback result. According to my teaching experience, I feel that this should not be the case. However, as this course will finish soon, so I still explain an example according to the feedback." T8 also mentioned once that "The feedback chart showed that two students were bored, but this slide is the most interesting and also difficult part of my entire content. At that time, I started to doubt the correctness of the system.

But I still talked about some more complicated content. Also, I was wondering if it was a problem that my explanation was not clear enough to make students having the states."

Apart from the finding above, we also noticed that all instructors who has mentioned his/her doubt of our system are all only doubting about the correctness and accuracy about our learning state prediction model. No one mentioned about the doubt of the visualization format, our training metadata or the student participants.

6 DISCUSSION

6.1 Usability and Effect of the Learning State Feedback

Through our system evaluation and result analyzing, we are happy to find that the real-time feedback of learners' learning states is indeed helpful for live online lectures. Live online lecturers can not only use the real-time feedback to help them find unrealized problems in their lectures but also support their decision-making on teaching adjustment in both real-time and post-hoc reviews. The evaluation also shows that the feedback will introduce both positive and negative emotional effect to instructors. During the live online lectures, instructors can either be encouraged by the learner's active participation or be disappointed by the learner's negative learning state, especially the Boredom state. At the same time, instructors have to pay extra attention to check the feedback charts and interpret their meanings which will also introduce extra workload to them and burden their teaching process. For the active intervention, some instructors think that it could reduce the extra workload, but others think it doesn't.However, with the help of the feedback, live online teaching becomes more similar to traditional real-world teaching and more interactive than its previous form. Besides, this implicit feedback gathering method can provide instructors with more fine-grained and prolonged feedback information than other explicit ways which will put significant effort on the learner to achieve the same goal. We can also expect a much more interactive online teaching environment will be introduced to real-world use by combing our system with other excellent interactive tools which are mentioned in previous sections.

6.2 Feedback Format

We find that the active intervention indeed could reduce the need for extra cognitive workload for most of the instructors. With the help of the intervention, they don't need to constantly check the feedback charts and therefore can simply concentrate on their lectures. In our design, the threshold for intervention is 50%. However, we can customize this number for different teachers according to their preference and also send active intervention to them when other predefined situations occur such as all learners are in the optimal Flow state. At the same time, for another small part of the instructors, the active intervention would cause extra anxiety and cognitive workload to them. We believe this situation could be properly solved by designing other more appropriate intervention formats, such as gentler text prompts and better alarm sounds.

We also find that, compared with the Moving Line Graph, the Bar Graph is more helpful in live online teaching. It is more intuitive and can easily been interpreted in a relatively short time, albeit it just contains the current distribution of three different learning states. At the same time, the moving live graph is more complicated, especially when the proportion of each learning state is significantly fluctuating. However, the line graph is much more helpful for post-hoc reviews. By aligning with lecture's slides, the moving line graph can clearly display the dynamic changes of each learning state over the whole lecture and help the instructor understand their performance on different pages.

6.3 Trust Issue between Users and Human-AI Collaboration Systems

The Trust Problem between human and AI has been much discussed and studied in the past decade. Although the recent breakthroughs show the great power of AI systems, many people still seem to lack confidence in AI predictions. In this paper, we define our feedback system as a new kind of Human-AI collaborative teaching-support system because the learning state is implicitly recognized through learner's facial expressions by Machine Learning Algorithms. Therefore, we are interested in exploring the Trust Problem between the live online instructors and our feedback system under this scenario of live online teaching. The experiment results suggest that the instructors indeed have the tendency to trust our feedback system and begin relying on the feedback result to make decisions on their teaching. Although, they might doubt the accuracy of our system at the beginning, they start to believe the result after they finish the evaluation of the system by themselves. We also find that for those instructors whose condition 1 was coming after other conditions with feedback, they tend to unconsciously check the feedback charts and ask for the support of feedback data in that condition.In the future, we are interested in exploring whether this kind of teaching-support system will cause some feedback dependence to our system and further introduce negative effects, such as anxiety, to the long-time user of our system when they have to teach without the feedback in live online lectures. Apart from the aspect we have discussed in this paper, there are also many other important and interesting aspects we can focus on in the future, such as whether the user trusts the machine learning component, data visualization component or even the system designer of our system. If not, what kind of doubts they have and why would they have such doubts on our system. We believe that we could gain insights about how to build a more trustworthy AI system and Human-AI interface by investigating the question above in the future.

7 CONCLUSION

In this paper, we present an explorative study investigating how presenters perceive and react to audience flow prediction when giving live-stream lectures. The study was conducted with an experimental system that can predict audience's psychological states (e.g., anxiety, flow, boredom) through real-time facial expression analysis, and can provide aggregated views illustrating the flow experience of the whole group. Through evaluation with 8 online lectures ($N_{instructors} = 8, N_{learners} = 21$), we found such real-time flow prediction and visualization can provide value to presenters. This paper contributes a set of useful findings regarding their perception and reaction of such flow prediction, as well as lessons learned in the study, which can be inspirational for building future AI-powered system to assist people in delivering live online presentations.

ACKNOWLEDGMENTS

This work was supported by National Key RD Program of China (Grant No. 2016YFB1001405), National Natural Science Foundation of China (Grant No. 61872349), Key Research Program of Frontier Sciences, CAS (Grant No. QYZDY-SSW-JSC041), and the CAS Pioneer Hundred Talents Program. We would like to thank all the reviewers for their detailed feedback and the participants in the user study.

REFERENCES

[1] 2012. InstFeedback. http://instfeedback.com.
[2] 2018. By The Numbers: MOOCS in 2017. https://www.classcentral.com/report/mooc-stats-2017/.
[3] 2019. Affectiva. https://www.affectiva.com/.
[4] 2019. KMF online learning platform. https://www.kmf.com/.
[5] 2019. New Oriental Education and Technology Group. http://www.neworiental.org/english/.

[6] Daniel Afergan, Evan M Peck, Erin T Solovey, Andrew Jenkins, Samuel W Hincks, Eli T Brown, Remco Chang, and Robert JK Jacob. 2014. Dynamic difficulty using brain metrics of workload. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3797–3806.

[7] Daniel Afergan, Tomoki Shibata, Samuel W Hincks, Evan M Peck, Beste F Yuksel, Remco Chang, and Robert JK Jacob. 2014. Brain-based target expansion. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. ACM, 583–593.

[8] Jeremy N Bailenson, Emmanuel D Pontikakis, Iris B Mauss, James J Gross, Maria E Jabon, Cendri AC Hutcherson, Clifford Nass, and Oliver John. 2008. Real-time classification of evoked emotions using facial feature tracking and physiological responses. *International journal of human-computer studies* 66, 5 (2008), 303–317.

[9] Ashok R Basawapatna, Alexander Repenning, Kyu Han Koh, and Hilarie Nickerson. 2013. The zones of proximal flow: guiding students through a space of computational thinking skills and challenges. In *Proceedings of the ninth annual international ACM conference on International computing education research*. ACM, 67–74.

[10] AJ Brush, David Bargeron, Jonathan Grudin, Alan Borning, and Anoop Gupta. 2002. Supporting interaction outside of class: anchored discussions vs. discussion boards. In *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community*. International Society of the Learning Sciences, 425–434.

[11] Isabel Buil, Sara Catalán, and Eva Martínez. 2019. The influence of flow on learning outcomes: An empirical study on the use of clickers. *British Journal of Educational Technology* 50, 1 (2019), 428–439.

[12] Di Laura Chen, Dustin Freeman, and Ravin Balakrishnan. 2019. Integrating Multimedia Tools to Enrich Interactions in Live Streaming for Language Learning. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 438.

[13] Mihalyi Csikszentmihalyi. 2014. Flow and the foundations of positive psychology. *Recuperado de http://www. springer. com/us/book/9789401790871* (2014).

[14] Mihaly Csikszentmihalyi and Isabella Csikszentmihalyi. 1975. *Beyond boredom and anxiety*. Vol. 721. Jossey-Bass San Francisco.

[15] Mihaly Csikszentmihalyi and Isabella Selega Csikszentmihalyi. 1992. *Optimal experience: Psychological studies of flow in consciousness*. Cambridge university press.

[16] Travis Faas, Lynn Dombrowski, Alyson Young, and Andrew D Miller. 2018. Watch me code: programming mentorship communities on Twitch. tv. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 50.

[17] Elena L Glassman, Juho Kim, Andrés Monroy-Hernández, and Meredith Ringel Morris. 2015. 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*. ACM, 1555–1564.

[18] Mariam Hassib, Stefan Schneegass, Philipp Eiglsperger, Niels Henze, Albrecht Schmidt, and Florian Alt. 2017. EngageMeter: A system for implicit audience engagement sensing using electroencephalography. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 5114–5119.

[19] Joel Hektner, Jennifer Schmidt, and Mihaly Csikszentmihalyi. 2019. *Experience Sampling Method*. Thousand Oaks, California. https://doi.org/10.4135/9781412984201

[20] Javier Hernandez, Zicheng Liu, Geoff Hulten, Dave DeBarr, Kyle Krum, and Zhengyou Zhang. 2013. Measuring the engagement level of TV viewers. In *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. IEEE, 1–7.

[21] Juho Kim, Philip J Guo, Daniel T Seaton, Piotr Mitros, Krzysztof Z Gajos, and Robert C Miller. 2014. Understanding in-video dropouts and interaction peaks inonline lecture videos. In *Proceedings of the first ACM conference on Learning@ scale conference*. ACM, 31–40.

[22] Kyu Han Koh, Ashok Basawapatna, Hilarie Nickerson, and Alexander Repenning. 2014. Real time assessment of computational thinking. In *2014 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE, 49–52.

[23] Yang Li, Samy Bengio, and Gilles Bailly. 2018. Predicting human performance in vertical menu selection using deep learning. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 29.

[24] Zhicong Lu, Seongkook Heo, and Daniel J Wigdor. 2018. StreamWiki: Enabling Viewers of Knowledge Sharing Live Streams to Collaboratively Generate Archival Documentation for Effective In-Stream and Post Hoc Learning. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 112.

[25] Rajitha Navarathna, Patrick Lucey, Peter Carr, Elizabeth Carter, Sridha Sridharan, and Iain Matthews. 2014. Predicting movie ratings from audience behaviors. In *IEEE Winter Conference on Applications of Computer Vision*. IEEE, 1058–1065.

[26] Thao Nguyen, Iris Bass, Mingkun Li, and Ishwar K Sethi. 2005. Investigation of combining SVM and decision tree for emotion classification. In *Seventh IEEE International Symposium on Multimedia (ISM'05)*. IEEE, 5–pp.

[27] Anna Nowicka, Artur Marchewka, Katarzyna Jednorog, Pawel Tacikowski, and Andre Brechmann. 2010. Forgetting of emotional information is hard: an fMRI study of directed forgetting. *Cerebral Cortex* 21, 3 (2010), 539–549.

[28] Phuong Pham and Jingtao Wang. 2015. AttentiveLearner: improving mobile MOOC learning via implicit heart rate tracking. In *International Conference on Artificial Intelligence in Education*. Springer, 367–376.

[29] Phuong Pham and Jingtao Wang. 2017. AttentiveLearner 2: a multimodal approach for improving MOOC learning on mobile devices. In *International Conference on Artificial Intelligence in Education*. Springer, 561–564.

[30] Phuong Pham and Jingtao Wang. 2017. 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*. ACM, 67–78.

[31] Phuong Pham and Jingtao Wang. 2018. Adaptive Review for Mobile MOOC Learning via Multimodal Physiological Signal Sensing-A Longitudinal Study. In *Proceedings of the 2018 on International Conference on Multimodal Interaction*. ACM, 63–72.

[32] David J Shernoff and Mihaly Csikszentmihalyi. [n. d.]. Cultivating Engaged Learners and Optimal Learning Environments. ([n. d.]), 16.

[33] David J. Shernoff, Mihaly Csikszentmihalyi, Barbara Shneider, and Elisa Steele Shernoff. 2003. Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly* 18, 2 (2003), 158–176. https://doi.org/10.1521/scpq.18.2.158.21860

[34] Hyungyu Shin, Eun-Young Ko, Joseph Jay Williams, and Juho Kim. 2018. Understanding the Effect of In-Video Prompting on Learners and Instructors. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 319.

[35] M Iftekhar Tanveer, Samiha Samrose, Raiyan Abdul Baten, and M Ehsan Hoque. 2018. Awe the Audience: How the Narrative Trajectories Affect Audience Perception in Public Speaking. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 24.

[36] Julianne C Turner, Kathleen E Cox, Matthew DiCintio, Debra K Meyer, Candice Logan, and Cynthia T Thomas. [n. d.]. Creating Contexts for Involvement in Mathematics. ([n. d.]), 16.

[37] Johannes Wagner, Jonghwa Kim, and Elisabeth André. 2005. From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. In *2005 IEEE international conference on multimedia and expo*. IEEE, 940–943.

[38] Jacob Whitehill, Marian Bartlett, and Javier Movellan. 2008. Measuring the perceived difficulty of a lecture using automatic facial expression recognition. In *International Conference on Intelligent Tutoring Systems*. Springer, 668–670.

[39] Xiang Xiao and Jingtao Wang. 2016. Context and cognitive state triggered interventions for mobile MOOC learning. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. ACM, 378–385.

[40] Chun yan Nie, Ju Wang, Fang He, and Reika Sato. 2015. Application of J48 decision tree classifier in emotion recognition based on chaos characteristics. In *2015 International Conference on Automation, Mechanical Control and Computational Engineering*. Atlantis Press.

[41] Chi-Lan Yang, Chien Wen (Tina) Yuan, Tzu-Yang Wang, and Hao-Chuan Wang. 2018. Knowing That You Know What I Know Helps?: Understanding the Effects of Knowledge Transparency in Online Knowledge Transfer. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 189 (Nov. 2018), 21 pages. https://doi.org/10.1145/3274458

[42] Beste F Yuksel, Kurt B Oleson, Lane Harrison, Evan M Peck, Daniel Afergan, Remco Chang, and Robert JK Jacob. 2016. 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*. ACM, 5372–5384.

[43] Wei-Long Zheng, Wei Liu, Yifei Lu, Bao-Liang Lu, and Andrzej Cichocki. 2018. Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE transactions on cybernetics* 99 (2018), 1–13.

[44] Sacha Zyto, David Karger, Mark Ackerman, and Sanjoy Mahajan. 2012. Successful classroom deployment of a social document annotation system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1883–1892.

Received April 2019; revised June 2019; accepted August 2019