International Journal of Learning, Teaching and Educational Research
Vol. 15, No. 3, pp. 204-217, March 2016

Student Attentive State Data Accumulation for Attention Tracker Development

Chi-Jen Lin
Fo Guang University
Yilan 26247, Taiwan

Abstract. Attention is vital to learning, and attention trackers are potentially powerful tools for education practitioners. Herein, the promising technologies and relevant studies on attention are reviewed. In order to realize the goals of attention trackers, this study aimed to accumulate initial attentive state data, and to explore potential problems in the use of the accumulated data. It was found that the gaze location was a good estimator of the student's attentive state. It was also discovered that real-time applications of attention trackers may find that previously obtained student attentive states must be altered at a later time. More studies are required for the development of attention trackers with desired characteristics. However, published results are promising.

Keywords: Attention tracker; gaze estimation; webcam

Introduction

Attention is an indispensable factor of successful learning. Without attention, poor learning outcomes are expected (Nissen & Bullemer, 1987). The impact of inattention on math and reading achievement of elementary children is both concurrent and longitudinal (Grills-Taquechel, et al., 2013; Gray, et. Al., 2015). Therefore, student attentive state data (attentive or inattentive) are valuable to investigate the reasons of poor student learning outcomes. In fact, student attentive state data are also informative for investigations of how instruction influences learning. For example, if a specific group of students in a class lose their attention at a specific time, this may indicate that the lecture was boring or too difficult for them at that time. Therefore, student attentive state data are valuable for the instructor's assessment and for instructional improvement studies.

In an ordinary classroom, it is generally not feasible to track and record the variation of attentive states of each student during the class. Even if it is economically feasible to track students' attentive states manually, it is expected there will be increased interference during the class under such educational settings. Hence, student attentive state data should be better obtained through automated tools, or using the attention trackers proposed in this study. In addition to applications in classrooms, attention trackers may also increase the bandwidth of intelligent tutoring systems in student modeling and allow implementation of more effective learning experiences. They may also be used to create new types of e-learning systems.

Recent advancements in the field of computer vision, along with other technological advancements, have facilitated the implementation of numerous practical applications, such as body motion sensing (Zhang, 2012), eye gaze control of computing devices (Lopez-Basterretxea, Mendez-Zorrill, & Garcia-Zapirain, 2015), Google glasses, and self-driving cars (Greenblatt, 2016). Most of these applications focused on commodities or entertainment. Unfortunately, development of similar applications in the field of education has received much lesser attention. However, based on the trend of development of these applications, this study envisions and argues that the development of attention trackers using webcams is promising with existing computer vision techniques. The use of webcams is emphasized because they are currently available on most smart phones, pad and laptop computers. Thus, all these devices can be converted into attention trackers with the installation of dedicated software. Relevant computer vision techniques will be reviewed in this paper to justify this argument.

When attention trackers are eventually created, their accuracy in attention tracking must be quantified to evaluate their usability. To quantify the accuracy of attention trackers, a video database of student learning with the associated attentive state labels assigned by human beings is required. The video data will be input into attention trackers to produce attention tracking data that will be compared with the attentive state labels stored in the database. Subsequently, the accuracy of attention trackers will be calculated based on the comparative results. Given that this attention tracking database does not exist currently, it is imperative that is manually created. In addition to accumulating data for this database, it is also significant to investigate whether consistency problems exist in the attentive state labels assigned by different persons. This consistency study is significant for verification of the effectiveness of the accuracy measure.

In the remaining parts of this paper, we will first review some prior literature studies on the topic of attention, and the promising techniques for attention tracker development. An experiment on how to accumulate student attention data, and the generated results, are subsequently described. Discussion and conclusions are also outlined.

Published studies on attention

Attention was extensively studied in many academic fields, including psychology, cognitive science, special education, human-computer interface, computer vision, etc. It was reported that more than 40,000 studies existed in a survey of attention studies (Lin & Chou, 2010). However, concerns raised by the published studies differed among academic fields or even within the same field in some occasions. In the field of psychology, it was stated that the word "attention" could refer to different phenomena, such as focused attention, selective attention, attention switching, divided attention, and sustained attention (Wickens & McCarley, 2007). Nonetheless, psychological studies mainly focused on articulating the mental mechanisms of attention of human beings. On the other hand, human-computer interface studies manage user attention in order to optimize information displayed to users (Bulling, 2016), while special education studies might focus on training of children with attention deficit disorders (Barkley, 2013; DuPaul & Stoner, 2014; Smith, et al., 2015). However, among the various concerns of attention studies, the topic of visual attention intrigued the largest number of researchers, including those in the fields of psychology, cognitive science, computer vision, and education. A vast amount of work was conducted to study visual attention during reading, scene perception, and visual search (Rayner, 2009; Borji & Itti, 2013). Such studies were generally conducted with eye movement data.

This study investigates attention tracking, the mechanisms to recognize immediately on whether students are attentive during learning, especially during lectures. This scientific concern of attention is novel and few similar study is dedicated to it. It is significant to discover how much information is needed to recognized effectively students' attentive states. It was indicated in the literature that eye behavioral information, such as saccadic velocities, fixation durations, blink rates, and pupil diameters, were beneficial for inferring the emotional states of students (Porta, Ricotti, & Perez, 2012), which might be also beneficial for inferring the attentive states of students. Therefore, the studies of visual attention and eye behaviors were valuable for the study of attention tracking. It was also known that attention could be classified into overt and covert attention (Wickens & McCarley, 2007; Rayner, 2009; Bulling, 2016). Behavioral traits of overt attention showed alignment between gaze position and the object of interest, while covert attention did not. However, covert attention was difficulty to be estimated (Bulling, 2016). Furthermore, it was claimed that covert attention was not easy to achieve for tasks such as reading, scene perception, and visual search (Rayner, 2009). Thus, it was practical to neglect covert attention while inferring student attentive states for attention tracker development.

Potential techniques and strategy for attention tracker development

Attention tracking might involve estimation of visual attention, eye behaviors, facial expressions, and body gestures. Therefore, the task of attention tracking was supposed to be complex and difficult. However, the techniques used to estimate most of the aforementioned human behaviors were extensively studied in the computer vision literature. Particularly, gaze estimation (estimation of gaze position) with webcam data is now maturing (Li, Li, Qin, 2014; Wood & Bulling, 2014), although the main stream studies typically use cameras with improved specifications and an additional infrared light source to increase the accuracy of estimation (Chennamma & Yuan, 2013; Al-rahayefh & Faezipour, 2013). Products also exist for facial expression recognition, such as the FaceReader (for recognition of the emotion expressed by a facial expression), and for body gestures recognition, such as the Xbox Kinect. Therefore, in order to develop an attention tracker, it is important to learn how to discern whether students are attentive during learning by combining all these information.

To reduce the complexity in the development of attention trackers, techniques that could be potentially beneficial will be introduced incrementally, with a priority placed on gaze estimation. Gaze position is believed to account for most of the attentive states of students. If the gaze position of a student is targeted on a reasonable area, such as the lecturer or learning material, it is plausible to assume that the student is attentive. This hypothesis will be preliminarily explored in the experiment reported in next section. Information obtained from facial expressions and body gestures may be used to vindicate or override the assumptions posed for gaze estimation. Comparing the estimation of attention trackers with accumulated data in the attention tracking database might be beneficial to reveal how to combine all these information. Furthermore, studies of consistent attentive state label assignments by different people may also help discover the knowledge possessed by humans. Therefore, the task of accumulating student attentive state data is essential for the development of attention trackers.

Experiment on attentive state data accumulation

As stated in the introduction section, two types of information have to be accumulated in the attention tracking database, namely, video data of student learning, and their attentive state labels. However, the duration of video clips must be decided before commencing the data labeling task. In this study, video data of student learning was logically divided into units of five-second video clips to facilitate subsequent work. The video clips of the same student were actually stored in the same file, but a blank frame was added to separate one video clip from another. Each logical video clip was reviewed, and was then independently labeled with its attentive state by two students. This video analysis approach was suggested in the literature (Wu, Sung, & Chien, 2010). Hence, there were two attentive state labels for each logical video clip in order to facilitate the conduct of the consistency study. At this stage of the attention tracker development, the best place to use the webcam for this experiment was in PC rooms. Therefore, the experiment was conducted with students who attended classes in PC rooms. The details and the results of the experiment are given in the following subsections.

Experimental tools

Some software applications were developed to facilitate the experimental tasks and to reduce the error rates in video data labeling, and labeling data transcription. The functions of each software application are explained below.

(1) Video recording software: The main task of this software was to record video data of student learning. Since the video data was planned to be divided into five-second video clips, a blank frame was automatically added every five seconds by the software. Moreover, before recording, a calibration process was conducted to facilitate subsequent work. Firstly, a student had to ensure that the webcam setup was able to fully capture the

user. Therefore, after launching the software, a window appeared, and displayed what the webcam captured, as shown in Fig. 1a, in order to facilitate adjustment of the webcam setup. Secondly, in order to facilitate the task of attentive state labeling, some reference shots of each student were taken when the student was looking at the teacher and at the four corners of the screen, respectively, as shown in Fig. 1b to Fig. 1f. These reference shots were supposed to be compared with the video clip data in order to determine if the student was attentive or not. In Fig. 2b, the user was asked to press the spacebar while looking at the teacher. In the meantime, while the spacebar was pressed, a reference shot of the student was taken while looking at the teacher. Similarly, the student was asked to click at the disk located at the four corners of the screen. Reference shots were taken while the student was looking at the four corners of the screen.

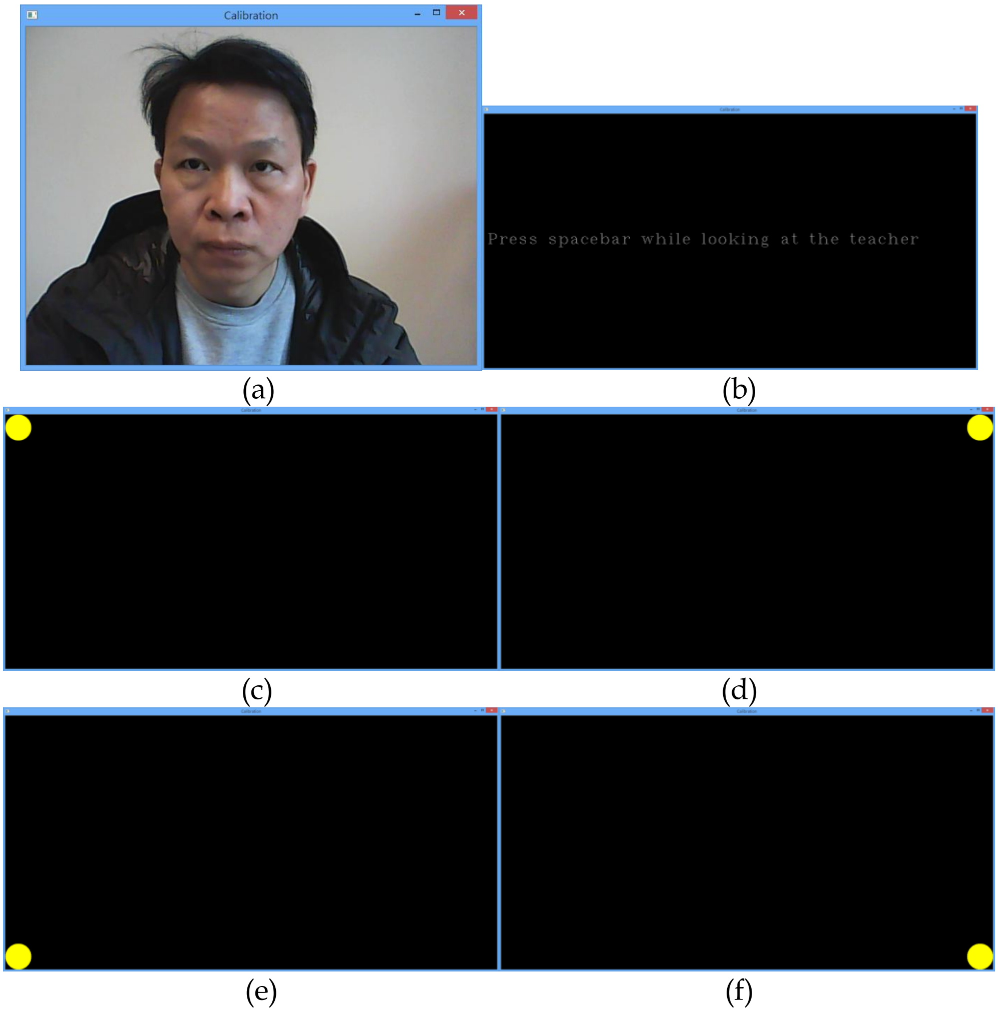

Figure 1: Calibration process before video recording.

(2) Attentive state labeling software: This software was used to display video clips stored in video files created by the video recording software, and to assign the associated attentive state labels. After loading a video file, the software automatically read the five reference shots and displayed them in designated tabs, as shown in Fig. 2b to Fig. 2f. The video itself was shown in

the first tab, as indicated in Fig. 2a. The user could switch from one tab to another freely, even during the video play. At the end of a video clip, the video display was automatically stopped. The play button was pressed to advance to the next video clip. However, to enable a fast search of video clips, the number of video clips could be keyed in to allow easy navigation. The results of attentive state labels were saved whenever the user clicked on the save button.

Figure 2: Interface for attentive state labeling (a), and sample reference shots (b-f).

(3) Gaze location labeling software: This software was used to label student gaze location in each video clip whenever it focused on reasonable targets, namely, on the computer screen, or on the teacher. These gaze location data were used for comparison with attentive state data to explore how well the gaze location data accounted for attentive state data. The interface of this software was similar to attentive state labeling software, and only some descriptive text labels in the interface differed.

Participants

Eleven students in total participated in the video recording experiment. Two students failed to properly set up their webcams, and their videos did not always include their entire heads. Unknown technical problems occurred in the video recording for one student leading to a non-useful video file. Therefore, the video data of these three students were not used in subsequent studies. All students recorded their own videos while attending a class in a PC room. Four students were hired to conduct the video labeling tasks, including the attentive state labeling, and the gaze location labeling.

Procedures

Initially, four student workers were hired to conduct the video recording experiment in order to identify potential problems of video recording in a PC class. These four students positioned webcams on the top of the computer monitors, at a position in front of the user faces. The frontal position of the webcams facilitate video reviews.

After the success of the video recording experiment, seven volunteers were recruited from two classes in Fo Guang University in Taiwan. No rewards were given to them. The intended academic use of their videos, and instructions for the use of the video recording software, were explained during their recruitment. The participants signed their agreements before they entered the experiments. Each participant received a webcam and a USB drive containing the video recording software. Student video files were stored in the USB drives during recording. All devices were returned after video recordings. The volunteers positioned webcams on top of the desks. This position of the webcam resulted in a skewed frontal view of student faces, because it was only approximately 50 cm away. At a later stage, it was found that this skewed frontal view resulted in difficulties in video reviews. The student workers reported that it was harder to discern the eye targets of the students in the skewed frontal views compared to those in frontal views. However, in order to develop an attention tracker, we need videos at different view angles to test the capability and limitations of attention trackers.

The number of recruited volunteers was much lower than our expectation. Most students did not enter the experiment because they hesitated to record their own videos. Even if the volunteers decided to enter the experiment, some of them behaved unnaturally before the cameras for a period of time. Furthermore, two volunteers failed to produce usable videos for our study, because their heads were not always in view. Another volunteer encountered unknown software problems. Finally, only of the four volunteer datasets was kept for subsequent labeling tasks.

After receiving the student videos, the four student workers served as reviewers of the conduct of the attentive state labeling tasks, with each video reviewed by two reviewers. Subsequently, only the videos with frontal views were used. Gaze location labeling tasks were conducted first, and inconsistent labels were then discussed by the reviewers to assess on whether they could reach consensuses.

Results

The statistical data of attentive state labels of frontal view videos are listed in Table 1. The data associated with each student video (denoted by s1 to s4) provided by the two reviewers are shown in each row. There were 2145 five-second (7.15 h) video clips in total in this category. Reviewers were allowed to assign three types of labels: attentive, inattentive, and undiscernible (a label to be given when the reviewer was not able to tell whether the student was attentive or not). However, there was a small percentage of video clips (0.09%) in which the reviewers neglected to assign labels. Thus, there were four possible attentive states in a video clip in total. Indiscernible video clips only occupied a small percentage (6.43%) of the total recordings on average, indicating that the attentive state of most video clips were discernible.

Table 1: Attentive state labels of frontal view videos.

Video	Attentive state label				Total number of video clips	Label consistency
	Unlabeled	Attentive	Inattentive	Indiscernible		
s1	0 (0%)	294 (50.43%)	242 (41.51%)	47 (8.06%)	583	317 (54.37%)
	1 (0.17%)	88 (15.09%)	494 (84.73%)	0 (0%)		
s2	3 (0.52%)	189 (32.64%)	373 (64.42%)	14 (2.42%)	579	399 (68.91%)
	0 (0%)	40 (6.91%)	539 (93.09%)	0 (0 %)		
s3	0 (0 %)	407 (55.68%)	117 (16.01%)	207 (28.32%)	731	304 (41.59%)
	0 (0 %)	250 (34.20%)	473 (64.71%)	8 (1.09%)		
s4	0 (0 %)	205 (81.35%)	47 (18.65%)	0 (0 %)	252	124 (49.21%)
	0 (0 %)	87 (34.52%)	165 (65.48%)	0 (0 %)		
Total	4 (0.09%)	1560 (36.36%)	2450 (57.11%)	276 (6.43%)	2145	1144 (53.33%)

According to Table 1, on average, only 53.33% of the video clips were assigned consistent labels by the two reviewers, indicating that human beings tended to be inconsistent about their views of student attentive states. However, this inconsistency was eliminated after the reviewers discussed the reasons of their decisions while watching the video clips together. From time to time, the attentive state of the student in the five-second video clips might have been attentive for a part of time, and inattentive for the rest. This finding was influential on the problem of attentive state inconsistency. When this finding was discovered, a criterion was established to assign to video clips the attentive state that dominated it. Nonetheless, sometimes it was still difficult for reviewers to estimate which attentive state dominated. In this case, reviewers would assign an indiscernible label. There were also some cases for which reviewers needed to also consider the subsequent video clip to assess which decision was more sensible. In addition to partial attentive state problems, misinterpretation of student behavior was another significant reason of inconsistency in attentive state labels. No matter which reason led to the assignment of inconsistent attentive state labels, the discussion phase was significant to ensure the quality of the final attentive state data.

Table 2: Consistency analysis of self-reviews and peer reviews (consensus).

Video	Attentive state label				Total number of video clips	Consistency
	Unlabeled	Attentive	Inattentive	Indiscernible		
s1 self	0 (0%)	92 (15.78%)	491 (84.22%)	0 (0%)	583	488 (83.7%)
peer	0 (0%)	155 (26.59%)	428 (73.41%)	0 (0%)		
s2 self	0 (0%)	516 (89.12%)	63 (10.88%)	0 (0%)	579	262 (45.25%)
peer	0 (0%)	199 (34.37%)	380 (65.63%)	0 (0%)		
s3 self	0 (0 %)	344 (47.06%)	384 (52.53%)	3 (0.41%)	731	608 (83.17%)
peer	0 (0 %)	362 (49.52%)	363 (49.66%)	6 (0.82%)		
s4 self	1 (0.40%)	150 (59.52%)	97 (38.49%)	0 (0%)	252	178 (70.63%)
peer	0 (0 %)	185 (73.41%)	67 (26.59%)	4 (1.59%)		
Total self	1 (0.05%)	1102 (51.38%)	1035 (48.25%)	3 (0.14%)	2145	1536 (71.61%)
Total peer	0 (0.00%)	901 (42.00%)	1238 (57.72%)	10 (0.47%)		

The discussion phase produced the consensual results of the two reviewers of each video. In this experiment, no disputation between the four pairs of reviewers was found during the discussion phase. In order to investigate how the consensual results of peer reviews of each video relate to student self-reviews, the students in the frontal view group were asked to do a self-review of their own videos. Consistency analysis between self-reviews and peer reviews was depicted in Table 2. Note that the rate of indiscernible labels was greatly reduced after peer discussion, as revealed by comparing Table 1 and Table 2.

The discussion phase helped the reviewers to become more capable of discerning the attentive states of students.

In consideration of the statistical data of skewed frontal view videos shown in Table 3, it was found that the average label consistency was not much different from those of the frontal view videos. In fact, the average label consistency of skewed frontal view videos was slightly better than those of the frontal view videos, although the reviewers reported that it was more difficult to review the videos in this category. It is possible that the skewed frontal view only increased the mental efforts of the reviewers, but did not influence other aspects of the review. There were 4480 five-second (14.93 h) video clips in total in this category.

Table 3: Attentive state labels of skewed frontal view video.

Video	Attentive state label				Total number of video clips	Label consistency
	Unlabeled	Attentive	Inattentive	Indiscernible		
s5	0 (0 %)	487 (50.84%)	467 (48.75%)	4 (0.42%)	958	733 (76.51%)
	5 (0.52%)	308 (32.15%)	645 (67.33%)	0 (0 %)		
s6	0 (0 %)	560 (47.95%)	164 (14.04%)	444 (38.01%)	1168	665 (56.93%)
	0 (0 %)	984 (84.25%)	118 (10.10%)	66 (5.65%)		
s7	0 (0 %)	1098 (93.77%)	35 (2.99%)	38 (3.25%)	1171	794 (67.81%)
	4 (0.34%)	768 (65.58%)	399 (34.07%)	0 (0 %)		
s8	0 (0 %)	649 (54.86%)	239 (20.20%)	295 (24.94%)	1183	534 (45.14%)
	3 (0.25%)	212 (17.92%)	845 (71.43%)	123 (10.40%)		
Total	9 (0.10%)	4854 (54.17%)	2067 (23.07%)	552 (6.16%)	4480	2726 (60.85%)

In addition to attentive state labels, each reviewer also assigned gaze location labels to the same set of video clips. Thus, the relationship between these two types of information of the same reviewers would reveal how well gaze location labels predicted attentive state labels. Table 4 showed the results of this relationship. The numbers in Table 4 denoted numbers of video clips.

The value of a gaze location label was either positive, indicating that the gaze location of the student was targeted on the computer screen or the teacher, or negative, indicating the opposite case. If a gaze location label was positive and the associated attentive state label was attentive, then the gaze location label was an accurate positive predictor of the attentive state label. Similarly, if a gaze location label was negative, and the associated attentive state label was inattentive, then the gaze location label was an accurate negative predictor of the attentive state label. According to Table 4, the rate of accurate positive predictions was approximately the same as the rate of accurate negative predictions, and they reached 80.7% on average. Therefore, the gaze location was a good approximation for student attentive states.

Table 4: Accurate prediction of gaze location on attentive states.

Review	Accurate positive prediction	Accurate negative prediction	Total accurate prediction	Inaccurate prediction	Total number of video clips
R1 (s1)	165 (79.33%)	224 (59.73%)	389 (66.72%)	194 (33.28%)	583
R2 (s1)	86 (93.48%)	489 (99.8%)	575 (98.63%)	8 (1.372%)	583
R3 (s2)	175 (46.67%)	185 (90.69%)	360 (62.18%)	219 (37.82%)	579
R4 (s2)	24 (54.55%)	519 (97.01%)	543 (93.78%)	36 (6.218%)	579
R5 (s3)	394 (98.01%)	117 (35.56%)	511 (69.9%)	220 (30.1%)	731
R6 (s3)	330 (100%)	387 (97.24%)	717 (98.08%)	14 (1.915%)	731
R7 (s4)	134 (67.68%)	36 (66.67%)	170 (67.46%)	82 (32.54%)	252
R8 (s4)	73 (64.04%)	124 (89.86%)	197 (78.17%)	55 (21.83%)	252
Total	1381 (78.33%)	2081 (82.48%)	3462 (80.70%)	828 (19.30%)	4290

In order to preliminarily investigate the effects of student ages, and to explore other potential factors that may influence the accuracy of attention tracking and the effectiveness of the data in the attentive state database, two grade-four students were recruited to conduct the same experiment. However, the learning activity was modified. The two elementary students were studying an online geometric learning material during the experiment. No human teacher gave the two students lectures during the experiment, but a teacher was around them during the experiment. Four teachers with more than ten-years teaching experience were recruited as reviewers. The results of label consistency analysis were listed in Table 5. According to Table 5, the label consistency of the reviewers of the two elementary student videos was much higher than that of Table 1 and Table 3, indicating that the teaching experience of reviewers do impact the effectiveness of the reviews.

Table 5: Label consistency analysis of elementary student data.

<table>
<tr><th rowspan="2">Video</th><th colspan="4">Attentive state label</th><th rowspan="2">Total number of video clips</th><th rowspan="2">Label consistency</th></tr>
<tr><th>Unlabeled</th><th>Attentive</th><th>Inattentive</th><th>Indiscernible</th></tr>
<tr><td rowspan="2">e1</td><td>0 (0 %)</td><td>117 (77.48%)</td><td>31 (20.53%)</td><td>3 (1.99%)</td><td rowspan="2">151</td><td rowspan="2">119 (78.81%)</td></tr>
<tr><td>0 (0%)</td><td>116 (76.82%)</td><td>22 (14.57%)</td><td>13 (8.61 %)</td></tr>
<tr><td rowspan="2">e2</td><td>0 (0 %)</td><td>253 (93.36%)</td><td>18 (6.64%)</td><td>0 (0%)</td><td rowspan="2">271</td><td rowspan="2">251 (92.62%)</td></tr>
<tr><td>10 (3.69%)</td><td>251 (92.62%)</td><td>10 (3.69%)</td><td>0 (0%)</td></tr>
<tr><td>Total</td><td>10 (1.18%)</td><td>737 (87.32%)</td><td>81 (9.60%)</td><td>16 (1.90%)</td><td>422</td><td>370 (87.68%)</td></tr>
</table>

Since the two students were generally concentrated on the learning tasks, the rate of inattentive states was 9.60% only, according to Table 5. The low number inattentive states may have influenced the rate of negative prediction of gaze location on attentive states, as depicted in Table 6. The negative prediction rates greatly fluctuated between 100% and around 48%, with an average of 68.97%. The positive prediction rates were more stables, with an average of 93.78%. The overall prediction rate of gaze location on attentive states was 85.19%, indicated again that gaze location was a good approximation of attentive states.

Table 6: Accurate prediction of gaze location on attentive states (elementary students).

Review	Accurate positive prediction	Accurate negative prediction	Total accurate prediction	Inaccurate prediction	Total number of video clips
r1 (e1)	117 (84.78%)	7 (100%)	124 (82.12%)	27 (17.88%)	151
r2 (e1)	116 (94.31%)	17 (100%)	133 (88.08%)	18 (11.92%)	151
r3 (e2)	206 (96.71%)	7 (53.85%)	213 (78.6%)	58 (21.4%)	271
r4 (e2)	240 (96%)	9 (42.86%)	249 (91.88%)	22 (8.118%)	271
Total	679 (93.78%)	40 (68.97%)	719 (85.19%)	125 (14.81%)	844

Discussion

It was difficult to solve the partial attention problem in student video clips at this stage. Ideally, the boundaries of video clips should be best placed at the transition points of changing attentive states in the video rather than be placed at the fixed 5-seconds periods. However, such boundary decisions would require knowledge on how to discern on whether a student was attentive or not, which was one of the goals pursued by segmenting the entire video into multiple clips. Although it was possible to determine the video clip onsets and ends manually, such a decision-making process was time consuming, and it was still unclear on whether such boundary decisions would be accurate, given that currently, appropriate theories for boundary decisions of video clips for attentive state labeling are still lacking.

The experience gained through the data accumulation experiment in this study might be worthy of consideration in attention tracker development, in that, sometimes reviewers may require more observation time to decide on whether a student was attentive or not. This experience indicated that attention trackers might be unable to produce attentive state data until a later time. Another possibility was that attention trackers may need to change their previously produced data at a later stage. Real-time applications of attention trackers should also take these features into consideration.

A noteworthy finding revealed in Tables 1, 2, and 3 was the low percentage of student attentive time. Students in the frontal view group were attentive for only 36.36% (42.00% according to the consensual results after peer discussion, 51.38% according to the results of self-reviews) of class time, while students in the skewed frontal view group were attentive for only 54.17% of class time. This finding was surprising for the class teachers. From this perspective, if attention trackers are available, teachers might use them to obtain the distribution of student attentive time of the entire class. Such information would be based on evidence rather than perception for teachers. In the study of the 2 sigma problem that is well known in the intelligent tutoring system research community (Bloom 1984), it was claimed that teachers were generally under the impression that all students in their classes were given equal opportunity for learning, while in fact they provided more favorable conditions for top students, and ignored average students. Attention trackers would be enabling tools to investigate the details of Bloom's claim (Bloom 1984). Attention trackers might also be used to estimate the impact of new pedagogical tools on student motivation. In most occasions, learning attention was also an indicator of learning motivation. Bored students would lose their learning attention eventually, while motivated students were usually highly focused on learning. Therefore, comparing student attention time would be an objective measurement for evaluating which pedagogical approach must be preferred.

Conclusions and future work

Attention trackers are potentially powerful tools for instructional improvements and educational tool enhancements. The realization of attention tracker development is promising, but more studies are required to investigate how to immediately recognize whether a student is attentive or not. Gaze location was shown to be a good approximation of the student attentive state. The use of other information to recognize student attentive state, such as eye behavior and body gesture, requires more investigations. In fact, the studies on how primary school teachers discern the attentive state of students are in progress. The goal is to obtain a set of reliable rules to recognize student attentive state. Such a set of rules will be useful for training video reviewers, for producing database with high quality data, and for developing attention trackers. It is expected that the realization of attention trackers will result in beneficial influences in the field of education.

Acknowledgements

The author would like to thank for the grant support of Fo Guang University and Ministry of Science and Technology of Taiwan under the contract number MOST 104-2511-S-431-001. Special thanks were given to Grace Gao for her efforts in conducting the elementary school experiment.

References

Al-Rahayfeh, A., & Faezipour, M. (2013). Eye tracking and head movement detection: A state-of-art survey. *IEEE Journal of Translational Engineering in Health and Medicine*, 1, 2100212-2100212. DOI: 10.1109/JTEHM.2013.2289879.

Barkley, R. A. (2013). *Defiant children: A clinician's manual for assessment and parent training*. Guilford Press.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4–16.

Bulling, A. (2016). Pervasive Attentive User Interfaces. *IEEE Computer*, 49(1), 94–98.

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1), 185–207.

Chennamma, H. R., & Yuan, X. (2013). A Survey on Eye-Gaze Tracking Techniques. *Indian Journal of Computer Science and Engineering*, 4(3), 388–393.

Conati, C., Merten, C., Muldner, K., & Ternes, D. (2005). Exploring eye tracking to increase bandwidth in user modeling. In L. Ardissono, P. Brna, and A. Mitrovic (Eds.): *User Modeling 2005, Lecture Notes in Computer Science 3538*, (pp. 357–366). Heidelberg: Springer, Berlin.

DuPaul, G. J., & Stoner, G. (2014). *ADHD in the schools: Assessment and intervention strategies*. Guilford Publications.

Gray, S. A., Rogers, M., Martinussen, R., & Tannock, R. (2015). Longitudinal relations among inattention, working memory, and academic achievement: testing mediation and the moderating role of gender. *PeerJ*, 3:e939. DOI 10.7717/peerj.939.

Greenblatt, N.A. (2016). Self-driving cars and the law. *IEEE Spectrum*, 53(2), 46-51.

Grills-Taquechel, A.E., Fletcher, J.M. Vaughn, S.R., Denton, C.A., & Taylor, P. (2013). Anxiety and inattention as predicators of achievement in early elementary school children. *Anxiety, Strees, & Coping*, 26(4), 391-410.

Li, X., Li, Z.L., & Qin, J.L. (2014). An improved gaze tracking technique based on eye model. *Proceedings of the 33rd Chinese Control Conference*, (pp. 7286-7291). Nanjing, China. IEEE.

Lin, H.Y., & Chou, T.J. (2010). The development of an attention test for elementary school children. *Bulletin of Special Education*, 35(2), 29-53 (In Chinese).

Lopez-Basterretxea, A., Mendez-Zorrilla, A., & Garcia-Zapirain, B. (2015). Eye/head tracking technology to improve HCI with iPad applications. *Sensors*, 15(2), 2244-2264.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: evidence from performance measures. *Cognitive Psychology*, 19(1), 1–32.

Porta, M., Ricotti, S., & Perez, C. J. (2012). Emotional e-learning through eye tracking. *IEEE Global Engineering Education Conference (EDUCON)*, (pp. 1–6). Marrakech, Morocco. IEEE.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506.

Smith, T. E., Polloway, E. A., Patton, J. R., Dowdy, C. A., & Doughty, T. T. (2015). *Teaching students with special needs in inclusive settings*. Pearson.

Wickens, C. D., & McCarley, J.S. (2007). *Applied Attention Theory*. CRC Press, Taylor & Francis Group.

Wood, E., & Bulling, A. (2014). Eyetab: Model-based gaze estimation on unmodified tablet computers. *Proceedings of the Symposium on Eye Tracking Research and Applications,* (pp. 207–210). Safty Harbor, Florida, USA. ACM.

Wu, H.K., Sung, Y.T., & Chien, H.Y. (2010). Conducting video analysis in educational research. *Journal of Research in Education Sciences*, 55(4), 1–37 (In Chinese).

Zhang, Z. (2012) Microsoft kinect sensor and its effect. *IEEE MultiMedia*, 19(2), 4-10.