Towards Automatic Recognition of Perceived Level of Understanding on Online Lectures using Earables

Dongwoo Kim*
NAVER
Seongnam, Republic of Korea
woo.kim@navercorp.com

Chulhong Min†
Nokia Bell Labs
Cambridge, UK
chulhong.min@nokia-bell-labs.com

Seungwoo Kang
KOREATECH
Cheonan, Republic of Korea
swkang@koreatech.ac.kr

ABSTRACT

The COVID-19 pandemic has seriously impacted education and forced the whole education system to shift to online learning. Such a transition has been readily made by virtue of today's Internet technology and infrastructure, but online learning also has limitations compared to traditional face-to-face lectures. One of the biggest hurdles is that it is challenging for teachers to instantly keep track of students' learning status. In this paper, we envision earables as an opportunity to automatically estimate learner's understanding of learning material for effective learning and teaching, e.g., to pinpoint the part for which learners need to put more effort to understand. To this end, we conduct a small-scale exploratory study with 8 participants for 24 lectures in total and investigate learner's behavioral characteristics that indicate the level of understanding. We demonstrate that those behaviors can be captured from a motion signal on earables. We discuss challenges that need to be further addressed to realize our vision.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; • **Applied computing** → *E-learning*.

KEYWORDS

Online Learning, Understanding Level, Earable, Automatic Recognition

ACM Reference Format:
Dongwoo Kim, Chulhong Min, and Seungwoo Kang. 2021. Towards Automatic Recognition of Perceived Level of Understanding on Online Lectures using Earables. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp-ISWC '21 Adjunct), September 21–26, 2021, Virtual, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3460418.3479323

*This work was done while the first author was with KOREATECH.
†Corresponding author.

UbiComp-ISWC '21 Adjunct, September 21–26, 2021, Virtual, USA

ACM ISBN 978-1-4503-8461-2/21/09…$15.00
https://doi.org/10.1145/3460418.3479323

1 INTRODUCTION

The COVID-19 pandemic has impacted every aspect of our lives. Among others, education has been seriously affected. Schools in a lot of countries had to close or reduce their face-to-face classes. According to some reports [1, 2], more than 1 billion students worldwide could not use their classrooms at the peak of the crisis. Alternatively, schools are providing access to education using online learning technology and numerous students are currently educated remotely all over the world. The pandemic would accelerate the educational innovation far beyond the advance of online learning witnessed over the last decade, e.g., Coursera and Udacity.

While the transition to online learning has been readily made by virtue of today's Internet technology and infrastructure, online learning also has limitations compared to traditional face-to-face lectures. One of the biggest hurdles is that teachers could not instantly keep track of students' learning status. In pre-recorded video lectures such as Coursera, teachers can neither observe how students engage in lectures, e.g., nonverbal and behavioral cues, nor interact with them. In live lectures using online conferencing tools such as Zoom, such observation and interaction can be possible if students have a camera and a microphone, but it imposes significant burdens on teachers, especially when there are a number of students in a lecture. These limitations hinder teachers from adapting their lecture materials or teaching methods when necessary. One typical method is to give a quiz after/during a lecture, but it is also burdensome to teachers due to the time and effort required.

In this paper, we envision *earables* (also known as smart earbuds) as an opportunity to automatically estimate learner's understanding status of learning materials. Such functionality would enable effective learning and teaching even in online lectures, e.g., to pinpoint the part for which learners need to put more effort to understand. From an explorative study, we uncover that a learner's postures and head motions can be a clue to represent their understanding of lectures. We present the capability and feasibility of identifying learners' understanding levels based on such behavioral patterns. Then, we discuss opportunities and challenges for realizing *the automatic estimation of understanding* in the wild.

2 RELATED WORK

Previous studies have demonstrated the capability of detecting online learners' diverse status (e.g., inattention, engagement, frustration, mind wandering) based on their behavioral patterns [3, 7, 13–15]. For example, Mota et al. presented a technique to recognize naturally occurring postures of learners and detect their interest level based on pressure sensors mounted on a chair [13]. Pham et al. proposed a multimodal approach to infer learners' affective and cognitive states such as boredom, confusion, and frustration

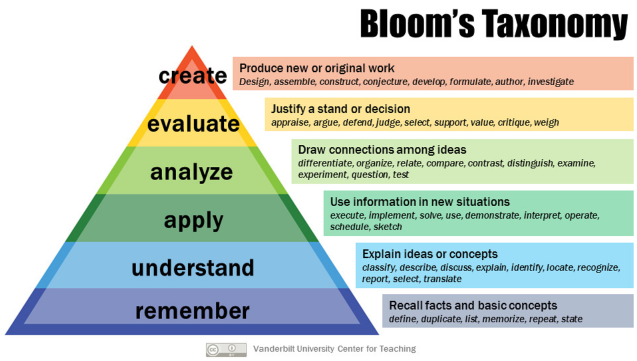


Figure 1: Bloom's Taxonomy (the image by the Vanderbilt University Center for Teaching / CC-BY [1])

by analyzing facial expressions and PPG signals captured from the front and back cameras of a smartphone [14]. Robal et al. presented an IntelliEye system that tracks face and mouth use of online learners to detect in-attention state [15]. It also provides alerts to them when they are in-attentive to return learners' attention to a lecture video. Grafsgaard et al. proposed to predict the engagement and frustration of students during computer-mediated tutoring based on fined-grained facial movements [7]. Bosch et al. proposed an automatic mind wandering detector by analyzing a set of facial features from face videos of students [3].

These works commonly imply that online learners' behaviors during the lecture carry meaningful information about their learning status. However, they mostly rely on physiological sensors and/or computer vision, thereby significantly limiting their practicality in real-life situations, due to the need for using additional, dedicated devices and privacy concerns. We investigate how learners' behaviors, more specifically postures and motion gestures, can be interpreted to represent their understanding while taking an online lecture. For example, a student might often tilt her head or look at a monitor vacantly when she does not understand what she listens to. Similarly, some students might nod their head when they well follow what is taught. Our initial idea and preliminary study was presented in [9].

3 APPROACH TO UNDERSTANDING ESTIMATION

3.1 Modelling Understanding

Measuring how well students learn and understand lectures is essential for teachers to provide effective teaching. Bloom's Taxonomy is a widely-adopted educational framework, which was developed to assist teachers to plan classes and design valid assessment strategies [4, 10]. The revised version of Bloom's Taxonomy introduces six levels of cognitive learning; remembering, understanding, applying, analyzing, evaluating, and creating [10]. As shown in Figure 1, each level represents different cognitive skills and learning behaviors from the most basic to the more complex levels. For example, *remembering* is related to retrieving, recalling, and recognizing factual information and relevant knowledge, and *understanding* is related to interpreting, summarizing, and explaining main ideas and concepts of learning material. Moving up the levels, they refer to higher cognitive thoughts and skills. In this study, we adopt the first two levels of Bloom's Taxonomy, remembering and understanding^{+}, to model online student's understanding* of learning material [2].

[1]https://www.flickr.com/photos/vandycft/29428436431

Table 1: Questionnaire for the level of understanding.

1	I could tell the important keywords/concept of the lecture
2	I could briefly explain the important keywords/concept of the lecture
3	I could tell what I newly learned
4	I could explain the summary of the lecture content
5	I could explain the lecture content so that others can understand it

3.2 Quantifying Understanding Level

To quantify the level of learner's understanding, we design a questionnaire by adopting Bloom's Taxonomy. We note that teachers are often encouraged to use different types of questions in class and on assignments and tests based on Bloom's Taxonomy to stimulate and assess students' cognitive thinking. Example questions that can be used are as follows.

- How would you define ... ?
- What was the main idea ... ?
- Can you write a brief outline ... ?
- Can you provide an example of ... ?

Inspired by these, we design the questionnaire with five statements as shown in Table 1. We construct the first two statements in the table for the concept of remembering and the rest three for the concept of understanding^{+}. For each statement, respondents are asked to rate how much they agree with the statement using a 5-point Likert scale (1 to 5) (5 - "strongly agree", 4 – "agree", 3 – "neutral", 2 – "disagree", 1 – "strongly disagree").

We use Likert scale answers that can be easily collected and quantified regardless of lecture types and contents. Asking detailed answers specific to the lecture content might be better to assess respondent's understanding level more accurately, but it imposes much burden on respondents to answer and assessors. Investigating the effect of different types of questions will be our future work.

As a granularity of understanding level estimation, we target *a lecture slide* as a unit, i.e., estimating the student's understanding level for each lecture slide. A slide often conveys a single topic and content, and thus it is naturally expected by teachers to be mapped the level of learner's understanding. Other lecture units, e.g., an explanation of a specific term in the slide or a whole lecture, can also be considered for different purposes. We leave it future work.

3.3 Why Motion Sensing on Earables?

A key decision to be made for the design of a sensing solution is to determine devices and sensors to be used. Many existing methods often rely on computer vision using a user-facing camera to detect learner's behaviors, thereby significantly limiting their applicability in real-life situations and raising privacy concerns. Unlike them, we focus on motion signals (accelerometer and gyroscope)

[2]Understanding^{+} refers to the specific level of Bloom's Taxonomy. We use understanding* as the term representing both *remembering* and *understanding* of the taxonomy. In the rest of the paper, understanding refers to understanding*, unless otherwise noted.

on earable devices. Our choice offers several benefits. First, according to our study in the following section, postures and gestures relevant to understanding are mostly made around the head and upper body, which could be captured by earable devices. Second, processing motion signal is computationally efficient and privacy preserving compared to other methods, especially computer vision. Last but not least, earables are widely used when students take online lectures and inertial measurement unit (IMU) for motion signal is already employed on most smart earbuds. Thus, an earbud-integrated motion-sensing solution would be easily adopted without requiring additional devices.

4 BEHAVIORAL CUES FOR UNDERSTANDING ESTIMATION

4.1 DATA COLLECTION

For an in-depth study, we exploit a dataset including 8 participants and 24 online lectures in total. The participants (4 males and 4 females) were recruited from a university campus, and they were undergraduate students in Computer Science and Engineering. Their ages were between 23-26 (mean: 24.13, SD: 0.99). Each participant was compensated with a gift card equivalent to USD 18. The study was approved by the Institutional Review Board of KOREATECH (No. 20022502).

Table 2 shows the online lectures we used for the study. We choose four lectures on the course of Artificial Intelligence Basic, provided by K-MOOC, a Korean MOOC established in 2015. All the participants have a general interest in the topic of AI, but did not take lectures with the same content as the lectures in our study.

Each participant was invited to the lab for data collection. We explained the purpose and procedure of the study and obtained informed consent. Based on each participant's prior knowledge and level on AI and ML, we chose three different lectures that cover a range of difficulty levels. During the lecture, the participants were asked to wear the eSense earbuds [8, 12] for sensor data collection. They were also provided with printed lecture materials and a pen to make them feel that they take lectures as usual.

During the lecture, we collected three types of data from the participants: (1) 3-axis accelerometer and 3-axis gyroscope data sampled at 32 Hz from eSense to analyze their behaviors while taking online lectures, (2) a video stream using a participant-facing webcam as ground truth of their behaviors, and (3) questionnaire (Table 1) answers on every slide as ground truth of their understanding. The participants completed the questionnaire after finishing each lecture. Between lectures, they took a break of 5 minutes. From the responses to the questionnaire, we obtain a final understanding score between 5 and 25 for a lecture slide, by summing all the scores from five answers, with 25 indicating the highest possible understanding score.

4.2 Natural Behaviors during Online Lectures

We extract a set of behaviors that learners naturally make during online lectures. For the analysis, the researchers transcribed and coded the recorded videos with observable, repetitive behaviors that the participants performed while they were watching the lectures.

Figure 2 shows the list of behaviors we observed in the videos (See the left side of the figure); we excluded infrequently observed behaviors. We group them into three categories; *posture*, *body/head motion*, and *facial motion*. As expected, learners' macroscopic, whole-body movements were quite limited during online lectures, mostly sitting and watching the video. However, interestingly, a variety of their microscopic movements were observed, especially around the head and upper body. From this behavioral characteristic, we believe that our choice of earbuds as a sensing device has a great potential to capture learners' behaviors on online lectures.

4.3 Understanding-relevant Behaviors

As a next step, we identify key behaviors that can be used as clues to estimate understanding level. To understand the impact of our observed behaviors, we assess the statistical association between a) the statistic of observed behaviors and b) the reported scores of understanding in the questionnaire. For the statistic of the observed behaviors, we annotate the start and end time of behavior events and compute the statistical features on every slide. For the behaviors that last for a certain time duration, we measure the duration of every segments of behaviors and normalize it by dividing by the slide duration. Note that we additionally compute the total, average and standard deviation for the posture behaviors that last longer than 1 second. For the behaviors involving a brief motion, we count the number of events in every slide and also normalize it with the slide duration. We use Spearman's rank correlation for the correlation analysis, which is used to assess the relationship between two variables measured on at least an ordinal scale [11].

Figure 2 shows the correlation coefficients of the behavior statistics and the reported understanding scores (See the right side). The results show two important findings. First, there are several behavioral patterns that imply a moderate relationship to the level of understanding, e.g., over 0.3 or under -0.3. This implies the potential of estimating a learner's *understanding* from the observation of behavioral patterns. Interestingly, after the survey, we heard from our participants some comments that support our analysis. For example, P7 reported that he frequently looked at the lecture material when he could not understand well. Accordingly, he frequently lowered and raised the head. Such patterns are observed through the coefficient analysis, e.g., the negative correlation of the count of *briefly gazing at a monitor*, *lowering the head*, and *raising the head*. P6 mentioned that he often did neck rolls and changed his sitting posture when he could not understand well. These behavioral patterns are revealed in the negative correlation of the count of *moving the neck* and *briefly moving the body*.

Second, the relationship between behavioral patterns and the level of understanding differs depending on the individual, showing the need for personalized estimation. For example, five behaviors, i.e., *keeping looking down a desk*, *keeping gazing at a monitor*, *briefly gazing at a monitor*, *lowering the head*, and *nodding*, show correlation coefficients larger than 0.3 or smaller than −0.3 for P7, but they do not for P4. P4 has only one behavior, *moving the neck*, that shows a correlation coefficient larger than 0.3.

4.4 Feasibility of Understanding Estimation

To study the feasibility of understanding level estimation, we build regression models to predict the understanding score using the aforementioned behavioral features. Since the impact of behavioral

Table 2: Online lectures used in data collection.

Course	Topics	Duration	# of slides	# of participants
Artificial Intelligence Basic	Introduction to Reinforcement Learning	34 min.	13	8
	Markov Process	22 min.	10	8
	Markov Decision Process	39 min.	13	7
	Heuristic Search	22 min.	17	1

Category	Behaviors	Features (every slide)	P1	P2	P3	P4	P5	P6	P7	P8
Posture	Keeping looking down a desk (>= 1 sec)	Total duration	-0.06	-0.01	-0.08	0.04	0.22	0.37*	0.01	0.15
		Average duration	-0.11	-0.03	-0.18	0.00	0.47*	0.47*	0.33*	0.25
		Std of duration	-0.23	-0.16	0.02	-0.19	0.13	0.39*	0.21	0.12
	Briefly looking down a desk briefly (<1 sec)	Count of events	0.13	0.14	-0.25	0.15	-0.15	-0.15	-0.24	-0.01
	Keeping gazing at a monitor (>= 1 sec)	Total duration	0.02	0.03	0.14	-0.01	-0.20	-0.35*	0.11	-0.14
		Average duration	-0.01	0.09	0.15	-0.14	0.13	-0.05	0.38*	0.16
		Std of duration	-0.08	-0.02	0.36*	-0.09	-0.11	-0.24	0.15	-0.17
	Briefly gazing at a monitor (< 1 sec)	Count of events	0.00	-0.07	-0.31	-0.02	-0.04	-0.02	-0.36*	-0.26
Body/head motion	Lowering the head	Count of events	0.16	-0.19	-0.22	0.05	0.18	0.16	-0.45*	0.02
	Raising the head	Count of events	0.37*	0.10	-0.23	0.07	0.15	0.14	-0.23	0.05
	Leaning the body forward	Count of events	0.21	0.12	-0.09	0.14	-0.13	0.01	-0.01	-0.19
	Leaning the body backward	Count of events	0.03	0.30	-0.21	0.14	-0.10	-0.29	0.02	-0.24
	Leaning the body to the right	Count of events	-0.16	-0.09	-0.14	-0.13	0.00	-0.06	0.17	-0.24
	Leaning the body to the left	Count of events	0.23	0.09	-0.33*	-0.05	0.00	-0.08	0.17	-0.14
	Nodding	Total duration	0.20	0.18	0.05	-0.13	-0.02	-0.14	-0.33*	-0.18
	Moving the neck	Total duration	-0.15	0.02	0.20	0.42*	-0.29	-0.25	-0.15	-0.01
	Shaking legs	Total duration	0.00	0.01	0.00	0.00	0.27	-0.12	0.20	0.00
	Touching the face	Count of events	-0.08	0.02	0.05	-0.10	-0.02	-0.31	0.23	-0.19
	Briefly moving the body	Total duration	0.56*	-0.00	0.03	-0.04	-0.26	-0.25	0.21	-0.39*
Facial motion	Grimacing	Count of events	0.02	0.00	0.00	0.00	0.00	-0.21	0.00	-0.06
	Yawning	Count of events	-0.03	0.13	-0.02	0.21	-0.06	-0.39*	0.16	0.00
	Sighing	Count of events	0.19	-0.28	-0.00	0.00	0.00	-0.00	0.00	0.00
	Laughing	Count of events	0.00	0.00	-0.16	0.00	0.00	-0.09	0.00	0.00
	Dozing	Total duration	0.00	0.00	-0.30	0.00	0.06	0.00	0.00	0.00

Figure 2: Behavior list and correlation coefficients of behavior features and understanding score (* indicates p-value < 0.05)

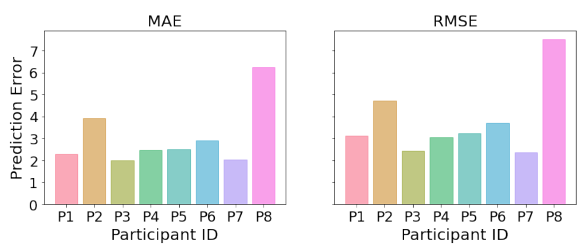


Figure 3: Prediction error for overall understanding scores

patterns is different depending on the individual, we train a separate regression model for each participant. We adopt linear regression. To train the model, we use the features with top-3 correlation coefficient values for each participant. We apply leave-one-slide-out cross validation.

Figure 3 shows the prediction errors with two metrics, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Our estimation model achieves 3.04 of the average MAE and 3.77 of the average RMSE across the participants; the target value is the total understanding level score and its range is from 5 to 25. For comparison, we use a baseline that takes an average understanding score as a prediction. The average MAE and RMSE of the baseline are 3.35 and 3.96, respectively. Our model shows the better estimation results for both metrics. We additionally compare the results of SVM regression with different types of kernels, i.e., linear and RBF, using top-3 features. They show slightly larger errors for both metrics, around 3.2 and 4 of MAE and RMSE, respectively.

We can also observe the variation of prediction errors depending on the participant. For example, P1, P3 and P7 show around 2 of MAE while P2 and P8 show 3.9 and 6.2 of MAE, respectively. We find that those who show relatively high errors have the smaller number of features with larger coefficient values. Also, their range of understanding scores is relatively larger across the lecture slides than others. We discuss this issue in the following section.

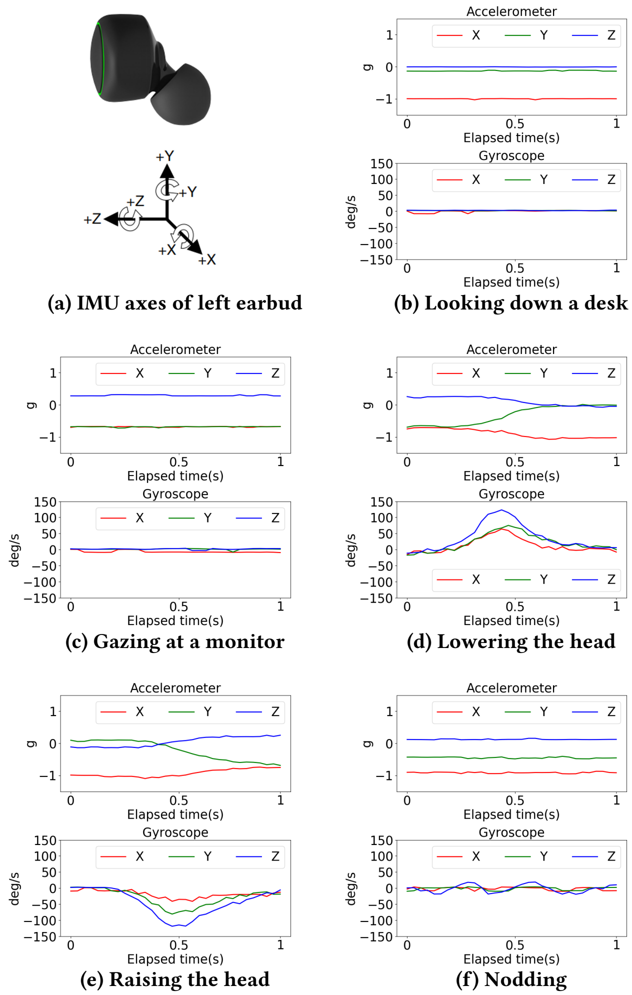
5 OPPORTUNITIES AND CHALLENGES

Our explorative study shows that the automatic estimation of learner's understanding is promising from identifying understanding-relevant behaviors and mapping them to the understanding score. In this section, we discuss opportunities and challenges for realizing the automatic estimation of understanding in the wild.

5.1 Behavior Detection using Earables

A key for the automatic estimation of understanding is to detect understanding-relevant behaviors at runtime. Here, we explore detection techniques for these behaviors using earable devices and their performance.

As an initial attempt, we choose five behaviors as primitive contexts for automatic estimation of learner's understanding level: two postures (*looking down a desk* and *gazing at a monitor*) and three motion gestures (*lowering the head*, *raising the head*, and *nodding*). These behaviors show a meaningful relationship with the understanding score, i.e., correlation coefficients larger than 0.3 or smaller than -0.3 in many participants and also observed relatively more frequently than other behaviors throughout the lectures. Note that, by detecting two postures, we can derive all of the posture category features in Figure 2.



(a) IMU axes of left earbud

(b) Looking down a desk

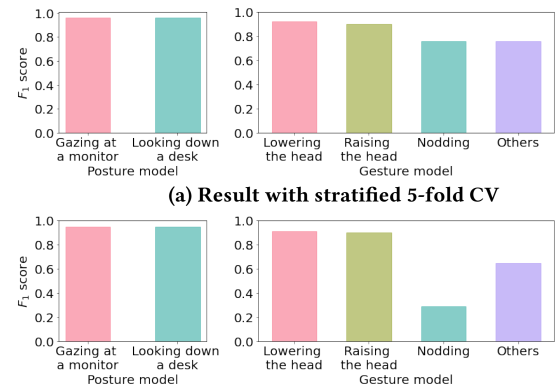
(c) Gazing at a monitor

(d) Lowering the head

(e) Raising the head

(f) Nodding

Figure 4: Motion signal of different postures and gestures



(a) Result with stratified 5-fold CV

(b) Result with leave-one-subject-out CV

Figure 5: Recognition performance

Detection technique: Postures and gestures have distinctive characteristics of motion signal patterns. Figure 4 shows the accelerometer and gyroscope data for some examples. While signals show little change over time while a user is taking a posture (e.g., Figure 4b and 4c), the fluctuation of signals can be easily observed in motion gestures (e.g., Figure 4d, 4e, and 4f). Inspired by such a finding, we devise a two-stage sensing pipeline to detect understanding-relevant behaviors. In the first stage, the pipeline quantifies the degree of movement and identifies whether a given signal segment is from a posture or a gesture. Then, in the second stage, it employs two machine learning models, one for posture detection and the other for gesture detection, and selectively uses them based on the output of the first stage. For each task, posture or gesture detection, a single model is built for all the users.

Movement detection: We take one-second gyroscope samples as input and quantifies the movement by calculating the signal variation. More specifically, we compute the magnitude value of every 3-axis gyroscope sample and calculate the variance of 32 values. The higher variance represents the higher degree of movement. Then, we distinguish between a posture and a gesture using a threshold that we empirically set using our dataset.

Machine learning models: One of the challenges in distinguishing between *gazing at monitor* and *looking down a desk* is that learners mostly remain stationary when they make these behaviors. Accordingly, we can easily expect that traditional activity pipelines for smartphones [16] would not work well because they are mostly designed to utilize the magnitude stream as input to address the arbitrary position of smartphones. On the contrary, the relative direction of the earbuds to a person's head is mostly fixed. Thus, we can leverage the absolute orientation of a device. We empirically found that X-axis and Z-axis show a strong discrimination power to identify understanding-relevant behaviors.

We segment accelerometer and gyroscope data streams into one second-frame. Then, we extract time and frequency-domain features [5] from X and Z streams separately without taking the magnitude and gather the features for the classification. We use PCA to reduce the dimensionality of the features and Support Vector Machine (SVM) as a classifier; we fine-tuned hyper-parameters using our dataset. Two machine learning models have the same architecture, but different target labels. The posture model is to separate looking down a desk and gazing at a monitor, and the gesture model is to separate lowering the head, raising the head, nodding, and others. We do not include the *others* label in the posture model because other postures are hardly observed in our data collection setup. However, we believe our model can be easily extended to cover other postures if needed.

Results and implications: We investigate the recognition performance of our detection technique. We conduct a stratified 5-fold cross-validation and report the F_1 score as a performance metric. The experimental results show that our machine learning models detect understanding-relevant behaviors accurately. Figure 5a show the F_1 score of posture and gesture models, respectively. The posture model shows 0.96 of F_1 score for both postures, looking down a desk and gazing at a monitor. It validates our choice of using axis-specific streams as input, instead of the magnitude. The gesture model also shows reasonable performance. The F_1 scores of two gestures, lowering and raising the head, are 0.92 and 0.90, respectively. However, the F_1 scores of nodding and others are relatively lower, 0.76 for both. This was mainly because the gesture model sometimes confuses the events of nodding and others.

5.2 Variation of Behavioral Patterns

One of the challenges for deploying the automatic estimation of understanding to end-users is the variety of behavioral patterns. First, our participants have a different set of behaviors that show a meaningful relationship with their understanding. For example, while the average duration of the *looking down a desk* posture shows a high correlation for P5, P6, and P7, but has little relationship for P1, P2, and P4.

Second, the signal patterns of behaviors are also different depending on the participant. To investigate how our posture and gesture models work on a new user, we measure their performance with a leave-one-subject-out validation (Figure 5b). The results show that the posture model still achieves high accuracy, i.e., 0.95 of F_1 score. The gesture model also shows the reasonable performance for lowering and raising the head, i.e., 0.91 and 0.90, respectively, which are expected to have little variation across the participants. However, the performance of nodding and others largely decreases, i.e., 0.29 and 0.65 of F_1 score. Observing the collected data, we could see that the participants often did nodding and other behaviors differently, e.g., in terms of direction, count, and strength.

These two findings imply the need for personalized models, i.e., detection model for behavior recognition and regression model for estimating the understanding score. We believe we can develop personalized models using a small amount of the end-user's data with online learning techniques. We leave it as future work.

5.3 Incorporating Additional Sensors

This study currently explores the feasibility of using behavioral features from earables' IMU to estimate students' understanding. It would be possible to incorporate additional sensors for more accurate and robust estimation. Some previous works utilize physiological sensors such as PPG and EDA to detect engagement or attention state of students [6]. While they are different from our target, understanding level, these might be related to each other, considering that engagement in lectures might positively affect the understanding of the lectures. We will further investigate the potential of adopting sensor fusion techniques to use physiological features from PPG and EDA sensors. A further in-depth study will also be necessary to analyze the relationship between the understanding level and the engagement or attention state.

5.4 Application Landscape

We envision that the automatic estimation of learner's understanding will provide significant benefits in online lectures.

Our proposed technique could be used to assist teachers by providing student's understanding status. In many online lectures, especially the pre-recorded video lectures, it is almost infeasible for teachers to have detailed real-time feedback from students, thereby making it difficult to modify and enhance their lectures; even possible, the feedback is often at a high, coarse-grained level. We envision that our solution can monitor student's understanding status, gather this information, and automatically spot parts in online lectures which student perceive difficult to understand.

Our solution could also help students by calling their attention when they do not understand the lecture content well. When our solution detects such moments, it could send a notification message to the students, e.g., "how about watching this part again if you do not understand well?". It might also be possible to provide the students with a summary of difficult parts after the lecture in order to help the student reflect on the lecture.

6 CONCLUSION

This work is our initial step towards a vision to utilize earables as an opportunity to automatically estimate learner's understanding on online lectures. In this explorative study, we observe some behaviors that could be used as clues to estimate the understanding level, and investigate the feasibility of understanding estimation using these behaviors. To realize the automatic estimation of understanding in the wild, we have a range of challenges to address, which will be future avenues of our work.

ACKNOWLEDGMENTS

This research was supported in part by the Ministry of Science and ICT, Korea, under the ITRC support program (IITP-2020-0-01778) supervised by the IITP.

REFERENCES

[1] [n.d.]. *The COVID-19 pandemic has changed education forever. This is how.* Retrieved June 14, 2021 from https://www.weforum.org/agenda/2020/04/coronavirus-education-global-covid19-online-digital-learning/

[2] [n.d.]. *Education during COVID-19; moving towards e-learning.* Retrieved June 14, 2021 from https://www.europeandataportal.eu/en/impact-studies/covid-19/education-during-covid-19-moving-towards-e-learning

[3] Nigel Bosch and Sidney D'Mello. 2019. Automatic detection of mind wandering from video in the lab and in the classroom. *IEEE Transactions on Affective Computing* (2019).

[4] Patricia A Eber and Trent S Parker. 2007. Assessing Student Learning: Applying Bloom's Taxonomy. *Human Service Education* 27, 1 (2007).

[5] Davide Figo, Pedro C Diniz, Diogo R Ferreira, and Joao MP Cardoso. 2010. Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing* 14, 7 (2010), 645–662.

[6] Nan Gao, Wei Shao, Mohammad Saiedur Rahaman, and Flora D Salim. 2020. n-Gage: Predicting in-class Emotional, Behavioural and Cognitive Engagement in the Wild. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–26.

[7] Joseph Grafsgaard, Joseph B Wiggins, Kristy Elizabeth Boyer, Eric N Wiebe, and James Lester. 2013. Automatically recognizing facial expression: Predicting engagement and frustration. In *Educational Data Mining 2013*.

[8] Fahim Kawsar, Chulhong Min, Akhil Mathur, and Alesandro Montanari. 2018. Earables for personal-scale behavior analytics. *IEEE Pervasive Computing* 17, 3 (2018), 83–89.

[9] Dongwoo Kim, Chulhong Min, and Seungwoo Kang. 2020. Towards recognizing perceived level of understanding for online lectures using earables. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 739–740.

[10] David R Krathwohl. 2002. A revision of Bloom's taxonomy: An overview. *Theory into practice* 41, 4 (2002), 212–218.

[11] Ann Lehman. 2005. *JMP for basic univariate and multivariate statistics: a step-by-step guide.* SAS Institute.

[12] Chulhong Min, Akhil Mathur, and Fahim Kawsar. 2018. Exploring audio and kinetic sensing on earable devices. In *Proceedings of the 4th ACM Workshop on Wearable Systems and Applications*. ACM, 5–10.

[13] Selene Mota and Rosalind W Picard. 2003. Automated posture analysis for detecting learner's interest level. In *2003 Conference on Computer Vision and Pattern Recognition Workshop*, Vol. 5. IEEE, 49–49.

[14] 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.

[15] Tarmo Robal, Yue Zhao, Christoph Lofi, and Claudia Hauff. 2018. IntelliEye: Enhancing MOOC Learners' Video Watching Experience through Real-Time Attention Tracking. In *Proceedings of the 29th on Hypertext and Social Media*. ACM, 106–114.

[16] Aras Yurtman. 2019. *Activity recognition invariant to position and orientation of wearable motion sensor units*. Ph.D. Dissertation. bilkent university.