Sense the classroom: AI-supported synchronous online education for a resilient new normal

Krist Shingjergji[a], Deniz Iren[a], Corrie Urlings[a] and Roland Klemke[a,b]

[a] Open University of the Netherlands 6419 AT Heerlen, The Netherlands
[b] Faculty of Cultural Sciences, TH Köln, Cologne, Germany

Abstract
Following the COVID-19 pandemic, as the user-base of online synchronous communication systems skyrocketed, the shortcomings of synchronous online learning systems became more visible. Any attempt to overcome these shortcomings should be considered worthwhile due to the magnitude of potential impact. Improving the quality and addressing the shortcomings of online education is more important than ever. The goal of this multidisciplinary study that lies in the intersection of the fields of Education Science and Computer Science is to address a number of challenges of online education by incorporating AI. This study focuses on developing methods and means to ethically collect and use non-verbal cues of participants of online classrooms to assist teachers, students, and course coordinators by providing real-time and after-the-fact feedback of the students' learning-centered affective states.

1. Introduction and motivation

Online learning provides a means of education to students with physical limitations or inconvenience to participate in physical, face-to-face classroom education. During the COVID-19 pandemic, this limitation became relevant for all students. Approximately, 1.2 billion learners were affected by the closure of schools at the time of the pandemic [1] and educational institutions worldwide made a mandatory transition to online/hybrid learning [2]. As the utilization of online learning reached unprecedented levels, the already-known challenges of online education became painfully visible for both students (e.g., feeling of isolation [3]) and teachers (e.g., lack of face to face interaction with the students [4]).

Students' learning experience and performance are highly related to their psychological, physiological, and emotional states [5]. Teachers can notice when students are distracted, confused, tired, etc., and have the opportunity to adjust their teaching approach accordingly [6], and choose appropriate interventions to keep the learning experience of the class optimal. However, many teachers who gave an online lecture have experienced the severe lack of an understanding of the *learning-centered affective states* of students in the classroom, thus, missing opportunities to improve the overall learning experience. This also directly impacts individual students. In online lectures, students are more prone to distractions and use the Internet for purposes unrelated to the educational activity [7]. The lecturers cannot give timely feedback to guide

the attention since they do not observe the students physically. As a result, the students are left alone to manage their learning experience, stay motivated, and struggle not to fall behind during the educational activity. The underlying reason that leads to these challenges is the communication modality limitations of video conferencing technologies.

In this study, we build on Media Naturalness Theory to examine the limitations of video conferencing as a medium of communication for online, synchronized education. Our objective is to develop artificial intelligence (AI) models to detect a multitude of components of learning-centered affective states (e.g., gestures, micro-expressions, and macro-expression) of the learners, and present the aggregated information to the teacher and the course coordinator in a privacy-protecting manner, and provide the individual information to the students themselves.

The remainder of this paper is structured as the following. Section 2 sheds light on the background and related work. Section 3 lays out the details of the overall research methodology. Section 4 highlights important discussion points such as the theoretical and practical implications, and ethical considerations.

2. Background and related work

In this section we explain the overall methodology that is proposed for this study.

2.1. Communication modalities and Media Naturalness Theory

Human communication occurs in multiple modalities such as voice, speech, facial expressions, and body language [8]. One important type of human communication is the non-verbal communication which is the way of conveying information without the use of words via non-verbal cues i.e., facial expressions and body language [9]. Video conferencing platforms fall short in conveying non-verbal cues among participants. The shortcomings of video conferencing as a communication medium can be analyzed and improved based on the Media Naturalness Theory (MNT). MNT describes the criteria to assess the degree of naturalness of a communication medium that partly relates to the capability of transmitting body language, facial expressions, and natural speech [10]. According to this theory, a reduction in a medium's naturalness may lead to a decrease in learning effectiveness, and a potential increase in ambiguity of the conveyed message [11].

2.2. Learning centered affective states

Many studies that aim to detect the relationship between online learning and emotions by applying emotion recognition techniques, use the basic emotions, namely, happiness, sadness, fear, disgust, anger, and surprise [12]. A plethora of studies that report an accurate mapping among facial expressions and emotions exist in the literature [13], [14], [15]. However, D'Mello in [5] states that the basic emotions are quite infrequent in the context of learning with educational software which raised the need of focusing on the learning-centered affective states, such as *engagement, concentration, boredom, anxiety, confusion, frustration,* and *happiness.* In contrast to emotion recognition, the mapping between facial expressions and learning-centered affective states has been severely understudied [16].

The observable non-verbal cues consist of gestures and body postures (e.g., head-tilt, nod, shake), micro-expressions (e.g., movement of inner eyebrows and lips), other expressions (e.g., smile, frown, confusion), and other activities (e.g., note-taking, active-listening, looking-away) [17]. The state-of-the-art facial-expression recognition (FER) and gesture recognition (GR) models use Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) hybrid networks [18]. These models perform discrete/momentary measurement (i.e., in short intervals), generally on single modality, and they are trained on datasets in which the non-verbal cues are mimicked by actors (i.e., not naturally occurring). As previously mentioned, the detection of constructs other than the six universal emotions[2], such as learning-centered affective states does not have a rich literature. However, the collection of high-quality data for the recognition of learning-centered affective states has been the subject of several studies

---

[2] sadness, happiness, fear, anger, surprise, and disgust

that have certain important limitations, for instance; focusing only on game-based interfaces [19], being explicit to certain ethnic groups [20], and having a limited target affective set such as the level of engagement on a scale [21], and the lack of interest and boredom [22].

In this study, we will bridge this gap by improving the CNN-RNN hybrids architecturally by introducing attention layers, formulating fitting objective functions, fusing data from multiple modalities, and applying transfer learning to train models with the multi-modal data collected from synchronous online education settings.

3. Methodology

In this section we explain the overall methodology that is proposed for this research.

3.1. Research model

Active and engaged learning is an important model in online higher education. Therefore, this study aims at addressing the motivational and emotional side of online education, by providing information that can assist educators to refine the educational activities that they have devised. Our aspiration is to utilize theories of learning, motivation, and emotion in combination to (1) define the relationship between learning-centered affective states of students and observable non-verbal cues, (2) develop specialized multi-modal AI algorithms for the recognition of learning-centered affective states, (3) and design tools to present this information in an actionable way for teachers and students to improve the learning process respecting the privacy of all participants (Figure 1).

Thus, the research questions of our studies are as follows:
1. Which are the specific non-verbal feedback needs (e.g., facial expressions) of teachers and students in online lectures?
2. How can we automatically detect non-verbal cues and translate them to learning-centered affective states of multiple participants in online, synchronous, educational activities?
3. How can we present this information to teachers in real-time so that they can take

actions to positively influence the learning-centered affective states of the students?
4. How can we provide students with this information so that learning-centered affective states are positively influenced?
5. How can we design a system that is ethically sound, that respects privacy concerns and keeps all collected data secure?
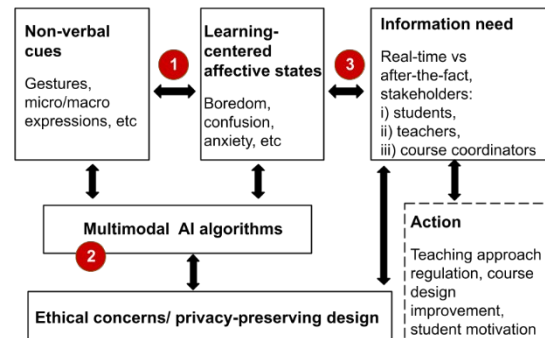

Figure 1: The research model

3.2. Approach

In this study, we will employ Design-Based Research (DBR) and experiments throughout multiple iterations (Figure 2). Teachers and students will be involved in focus groups and co-designing of prototypes [23]. The DBR iterations consist of literature study, requirements elicitation, participatory design, and evaluation of the interventions (i.e., integrated AI models) in pilot studies of online learning. The AI models will be developed through experimentation cycles which comprise data collection, annotation, algorithm development, and model training and evaluation. We will collect data from public-domain video repositories and online lecture sessions recorded by us with the informed consent of participants. The data will be annotated by multiple experts in terms of observed non-verbal cues. Consecutively, we will develop algorithms, train and test FER-GR, and the learning-centered affective states recognition AI models on multiple datasets to ensure generalizability. We will rely on metrics that are commonly used in machine learning, i.e., precision, recall, and F-1 measure to evaluate the accuracy of our models. Data management will be conducted in line with the FAIR data principles [24].
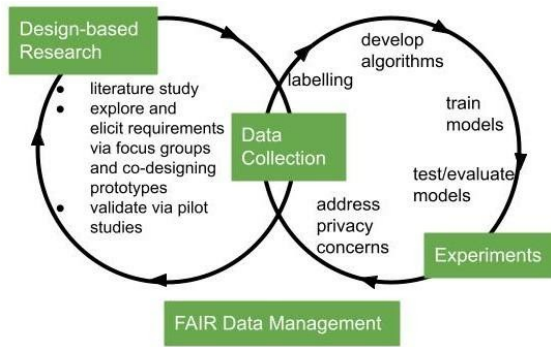
Figure 2: The research approach

We envision the system to provide certain information in real-time and after-the-fact to various stakeholders for different purposes (Table 1). The information flow is targeted at specific educational purposes for each party involved, with short-term and long-term educational benefits.

Table 1
Information flow to the actors of the system

When	Who	What	Why
Real-time	Teacher	Aggregated information regarding the overall learning-centered affective states of students in the classroom.	Teacher may (a) alter the teaching style and/or (b) initiate interventions (e.g., breakout rooms).
Real-time	Student	A semiotic indicator that shows their own learning-centered affective states, as well as suitable nudges (e.g., pop questions).	Keep the student engaged and active, positive influence on the learning process and self-regulation.
After-the-fact	Teacher	A report that shows the detected learning-centered affective states of the overall classroom matched to different parts of the educational activity (e.g., slides, activities, interventions) as a timeline.	Improve design of educational activity and delivery style.
After-the-fact	Course Coordinator	Learning-centered affective states trend of a course throughout multiple online educational activities.	Evidence-based course design and ensuring educational quality.

4. Discussion

In this section, we discuss several important aspects of this study including the theoretical and practical implications, the privacy concerns as well as the limitations.

4.1. Theoretical and practical implications

The outcomes of our experiments will potentially allow us to gain a deeper understanding on how learning-centered affective states are indicated by observable non-verbal cues, and how these states can be related to an effective learning experience. Our results will also contribute to the Media Naturalness Theory by extending it to cover widely used video conferencing platforms and tailor it for education scenarios.

The practical outcome of this study will be an analytical platform that is integrated to video conferencing clients of the students and the teachers. The platform will be able to provide feedback both on *real-time* and *after-the-fact* for all the involved actors in the course: *students*, *teachers*, and *course coordinators* (Table 1). In real-time, the platform will provide the teachers with aggregated information regarding the learning-centered affective states of the students. This information will give the teachers the opportunity to respond in different ways such as changing the teaching style and/or intervene in the course content flow. Students are going to receive information regarding their own learning-centered states, which they can use to self-regulate and be active and engaged. Regarding the real-time feedback, the system will be designed in a way that optimizes its use while taking part in the educational activity. On a longer-term aspect, this aggregated information will be useful for the teachers and the course coordinators as it would play the role of an evidence-based course evaluation which can be used for future improvement of delivery style and course design from the side of the teacher and the course coordinator respectively.

4.2. Privacy

We acknowledge the privacy-sensitive nature of this study. To protect the privacy of students, and to prevent a possible misuse of the technology, e.g., using the obtained information to evaluate students, we design core privacy-preserving measures to shape our research around them. Firstly, all data collection and experimentation will be voluntary and with the informed consent of the participants. The ethical board will be consulted prior to all data collection phases. The training data will be collected anonymously with no possibility to link to individuals. Secondly, the designed system will keep sensitive individual data (e.g., video) on individuals' computers. We will use a virtual webcam that implements AI models and analyzes video data on client computers. This feature will also allow students to keep their camera off (use avatars or nothing at all) while still benefiting from the system. Only the processed and anonymized data (i.e., numerical representations of non-verbal cues) will be transferred, and the teacher will only be provided with information that is aggregated at classroom-level. Finally, this study solely aims at developing a method for improving the quality of online education, and not as a way of individual assessment of the students or the teachers. We are confident that this privacy-preserving design will not allow any misuse of the system.

4.3. Limitations

The source of the data in this study will be the participants' cameras, which results in two important limitations. First, we cannot observe the entire environment of a student, thus, it is not possible to differentiate whether the observed non-verbal cues of an individual student are the result of an event in the classroom or an off-task activity. Second, in the online classrooms, students are in control of their cameras and may refuse to turn them on even when the proposed privacy-preserving methods are in place. In that case, the proposed method is not applicable.

5. References

[1]   Education: From disruption to recovery, 2021   URL: https://en.unesco.org/covid19/educatio nresponse.

[2]   S. Dhawan, Online Learning: A Panacea in the Time of COVID-19 Crisis, Journal of Educational Technology Systems 49 (2020) 5–22. doi: 10.1177/0047239520934018.

[3]   M. Alawamleh, L. M. Al-Twait, G. R. Al-Saht, The effect of online learning on communication between instructors and students during Covid-19 pandemic, Asian Education and Development Studies (2020). doi: 10.1108/AEDS-06-2020-0131.

[4]   S. Gurung, Challenges faced by teachers in online teaching during Covid-19 pandemic, The online journal of distance education and e-Learning 9 (2021).

[5]   S. D'Mello, A selective meta-analysis on the relative incidence of discrete affective states during learning with technology, Journal of Educational Psychology 105 (2013) 1082-1099. doi: 10.1037/a0032674.

[6]   K. Bahreini, R. Nadolski, W. Westera, Towards multimodal emotion recognition in e-learning environments, Interactive Learning Environments, (2016) 590–605. doi: 10.1080/10494820.2014.908927.

[7]   A. Lepp, J. E. Barkley, A. C. Karpinski, S. Singh, College Students' Multitasking Behavior in Online Versus Face-to-Face Courses, SAGE Open 9 (2019), doi: 10.1177/2158244018824505.

[8]   C. Jewitt, J. Bezemer, K. O'Halloran, Introducing multimodality, 1st. ed., Routledge, London, 2016. doi: 10.4324/9781315638027.

[9]   APA Dictionary of Phycology, 2020. URL: https://dictionary.apa.org/nonverbal-communication.

[10]   N. Kock, Media naturalness theory: human evolution and behaviour towards electronic communication technologies, in: S Craig Roberts, Applied evolutionary psychology, Oxford University Press Inc., New York, 2012, pp. 381-398. doi: 10.1093/acprof:oso/9780199586073.001.0001.

[11]   O. Weiser, I. Blau, Y. Eshet-Alkalai, How do medium naturalness, teaching-learning interactions and Students' personality traits affect participation in synchronous E-learning?, Internet and Higher Education 37 (2018) 40–51. doi: 10.1016/j.iheduc.2018.01.001.

[12]   P. Ekman, An Argument for Basic Emotions, Cognition and Emotion 6 (1992) 169–200. doi: 10.1080/02699939208411068.

[13]   R. Reisenzein, M. Studtmann, G. Horstmann, Coherence between emotion and facial expression: Evidence from laboratory experiments, Emotion Review 5 (2013) 16–23. doi: 10.1177/1754073912457228.

[14]   M. Wegrzyn, M. Vogt, B. Kireclioglu, J. Schneider, J. Kissler, Mapping the emotional face. How individual face parts contribute to successful emotion recognition, PLoS ONE 12, (2017) doi: 10.1371/journal.pone.0177239.

[15]   P. Tarnowski, M. Kołodziej, A. Majkowski, R. J. Rak, Emotion recognition using facial expressions, Procedia Computer Science 108 (2017) 1175–1184. doi: 10.1016/j.procs.2017.05.025.

[16]   M. A. A. Dewan, M. Murshed, F. Lin, Engagement detection in online learning: a review, Smart Learning Environments 6 (2019). doi: 10.1186/s40561-018-0080-z.

[17]   D. Umnia Soraya, K. Candra Kirana, S. Wibawanto, H. Wahyu Herwanto, C. Wijaya Kristanto, Non-Verbal Communication Behavior of Learners on Online-based Learning, in: Proceedings of the 2nd International Conference on Vocational Education and Training (ICOVET 2018), 2019, pp. 4-6. doi: 10.2991/icovet-18.2019.2.

[18]   M. Sharma, D. Ahmetovic, L. A. Jeni, K. M. Kitani, Recognizing Visual Signatures of Spontaneous Head Gestures, in: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 2018, pp. 400-408. doi: 10.1109/WACV.2018.00050.

[19]   Nigel Bosch, Sidney D'Mello, Ryan Baker, Jaclyn Ocumpaugh, Valerie Shute, Matthew Ventura, Lubin Wang, Weinan Zhao, Automatic detection of learning-centered affective states in the wild, in: Proceedings of the 20th International Conference on Intelligent User Interfaces (IUI '15), Association for Computing Machinery, New York, 2015, pp. 379–388. doi: 10.1145/2678025.2701397.

[20]   T. S. Ashwin, R. M. R. Guddeti, Affective database for e-learning and classroom environments using Indian students' faces, hand gestures and body postures, Future Generation Computer Systems 108 (2020), 334–348. doi: 10.1016/j.future.2020.02.075.

[21]   J. Whitehill, Z. Serpell, Y. C. Lin, A. Foster, J. R. Movellan, The faces of engagement: Automatic recognition of student engagement from facial expressions, IEEE Transactions on Affective Computing 5, (2014) 86–98. doi: 10.1109/TAFFC.2014.2316163.

[22]   L. B. Krithika, G. G. Lakshmi Priya, Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric, in Procedia Computer Science 85

(2016) 767–776. doi: 10.1016/j.procs.2016.05.264.

[23] K. Holstein, B. M. McLaren, V. Aleven, Co-designing a real-time classroom orchestration tool to support teacher–ai complementarity, Journal of Learning Analytics 6 (2019) 27–52. doi: 10.18608/jla.2019.62.3.

[24] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J. W. Boiten, L. O. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. C. Edmunds, C. T. A. Evelo, R. Finkers, A. N. González-Beltrán, A. J. G. Gray, P. Groth, C. A. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. W. W. Hooft, T. Kuhn, R. G. Kok, J. N. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. Laerte Packer, B. Persson, P. Rocca-Serra, M. Roos, R. C. van Schaik, S. Sansone, E. A. Schultes, T. Sengstag, T. Slater, G. O. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. M. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data 3 (2016). doi: 10.1038/sdata.2016.18.