Towards knowing students' emotional states from their voices while interacting with teachers

Ho Tan Nguyen[1,†], Mohammad Nehal Hasnine[2,*,†] and Hiroshi Ueda[3,†]

[123] *Research Center for Computing and Multimedia Studies, Hosei University, Tokyo 184-8584, Japan*

Abstract

In the rapidly evolving landscape of online education, accurately gauging student engagement and emotional state presents a critical challenge. We developed the Motion and Emotion (MOEMO) Learning Analytics Framework to address this. The MOEMO system relies on camera functionality, harnessing the power of visual analysis to decipher the complex tapestry of students' emotions and affective states. However, reliance on camera functionality constitutes a vulnerability; when cameras are off—whether by choice or necessity—educators are rendered blind to their students' affective states and nonverbal cues, a critical component of effective teaching and learning. Therefore, this research addresses the limitation above by proposing an innovative integration: adding a voice analysis component to the existing MOEMO framework. With this new component, the VoiceMetrics system, we aim to understand students' emotional states—for example, when they ask questions using a microphone and their camera is off. The synergy of auditory and visual data offers a more robust, emotional learning analytics approach to students' emotion recognition.

Keywords

affect, emotions, emotional learning analytics, voice analytics, multimodal learning analytics

1. Introduction

Understanding students' emotions is essential for teachers, as emotions play a crucial role in how and why students learn. In emotional learning analytics, emotion analysis analyzes a student's emotions from the student's remarks or facial expressions. Using this method, teachers understand the situation of students from the results of the emotion analysis and give accurate advice [1]. According to an article published in Times Higher Education, whether in a face-to-face or online environment, learners' and teachers' emotional states can influence one another [2]. By looking at their facial expressions, it is easy for an instructor to determine whether they are enjoying the class, confused about the content, or happy or sad about the lecture. Also, it is relatively easy to determine students' emotional and affective states if their web cameras are on during an online learning session with AI or deep learning-based emotion-aware learning analytics systems. However, it is exceedingly difficult for teachers to know their students' emotions and affective states if they cannot see them via web camera. Therefore, we need more sophisticated emotion-aware learning analytics systems to identify students' affective and emotional states from their remarks, for example, when they ask questions using a microphone or text in the chat panel.

In our previous research, we developed the MOEMO Learning Analytics Framework [3], [4], which can generate students' affective and emotional states by analyzing their facial expressions. However, if the camera is off, the MOEMO system cannot understand students' affective and emotional states. Therefore, we aimed to develop a method for analyzing students' voice and text data to understand

Joint Proceedings of LAK 2025 Workshops, co-located with 15th International Conference on Learning Analytics and Knowledge (LAK 2025), Dublin, Ireland, March 03-07, 2025.

[2*] Corresponding author. † These authors contributed equally.

✉ nguyentanhoit@gmail.com (H. T. Nguyen); nehal.hasnine.79@hosei.ac.jp (M. N. Hasnine); uep@hosei.ac.jp (H. Ueda)
iD 0009-0009-7795-376X (H. T. Nguyen); 0000-0003-4761-4002 (M. N. Hasnine); 0000-0001-6776-6500 (H. Ueda)

their affective and emotional states. We leverage the interactions between students and teachers during the discussion and Q&A sessions, typically after the lecturing time and practice-based classes, where students actively participate via microphone. These scenarios provide rich verbal data for emotion analysis, making the system applicable in online and blended learning environments. This paper introduces the VoiceMetrics system and discusses its usage in online learning.

2. Related works

Several emotion-aware and multimodal learning analytics systems have been developed in recent years. Emotion analysis techniques are applied to identify and guide students with problems, such as those who cannot participate in discussions. Muñoz, S., Sánchez, E., & Iglesias, C. A. (2020) developed an emotion-aware e-learning platform that recognizes students' emotions and attention to improve their academic performance [5]. Their proposed system recognizes students' emotions and engagement. It also lets educators consider this information to adapt their content and methodologies. Tanaka, M., Takao, I., & Keisuke, M. (2022) developed an 'AI teacher' system using coaching technology [1]. They embedded an emotion analysis module and a chatbot in the 'AI teacher' system and tested the system's effectiveness in a problem-based learning context. Ez-Zaouia, M., Tabard, A., & Lavoué, E. (2020) developed Emodash [6], an emotion-aware learning analytics dashboard for retrospective awareness. This system helped online tutors to incorporate more affective components to their reports and assignments. Oliver Fredriksen Werner, A. (2023) designed and developed a smart emotion-aware reflective system for teachers [7]. In this system, emotional metrics such as Engagement, Happiness, Anger, Entertainment, and Stress are displayed to teachers, indicating the class's emotional state at a glance. In addition to the abovementioned systems, Rienties, B., & Alden, B. (2014) discussed earlier years' research on using students' emotions in the learning and teaching process [8].

Several learning theories are tested in learning analytics, where emotional attributes are explored. The theories of motivation proposed by Maslow and Reiss and Emotional Reaction proposed by Plutchik (presented in Figure 1) are often cited in educational research [9]. Control-value learning theory is another popular theory that associates students' achievement emotions, and therefore, its implications are tested for educational practices [10]. Social-emotional learning for developing students' self-awareness, self-control, and interpersonal skills are explored in several studies [11]. Pekrun, R., & Stephens, E. J. (2012) categorized academic emotions into four categories, namely achievement emotions, topic emotions, social emotions, and epistemic emotions [12]. According to Chapter 12 of the Handbook of Learning Analytics, students show achievement emotions when they are asked to do some learning activities such as *homework* or *taking a test*, and their outcomes are in the form of *success* or *failure*. Topic emotions are linked with the learning content, such as *empathy for a protagonist*. Social emotions such as *pride*, *shame*, and *jealousy* occur because education requires interacting with others. Epistemic emotions are more connected with cognitive processing, such as *confusion in the face*.

Our study considers seven basic emotions: 'angry,' 'disgust,' 'fear,' 'happy,' 'sad,' 'surprise,' 'fearful', and 'neutral.' In this paper, we explored approaches to detecting these emotions from students' remarks, including when they ask questions using a microphone, via chat, or during interactive online discussion sessions with the teacher.

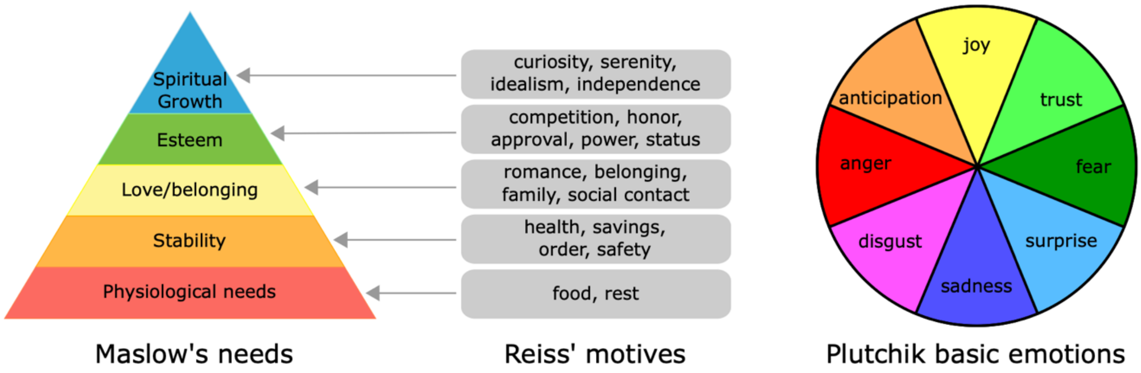


Figure 1: Theories of Motivation (Maslow and Reiss) and Emotional Reaction (Plutchik) [9]

3. MOEMO

The proposed VoiceMetrics system is an added function to the MOEMO system. Hence, we first introduce the MOEMO system. Figure 2 is the dashboard of the MOEMO system [4]. The dashboard reports on five engagement types: strong, high, medium, low, and disengagement. It also reports on the concentration level (focused or distracted). In addition, it reports to the teacher about the students' seven emotional states: 'angry,' 'disgust,' 'fear,' 'happy,' 'sad,' 'surprise,' 'fearful', and 'neutral.' The system analyzes data from students' facial expressions to generate these insights in the dashboard.

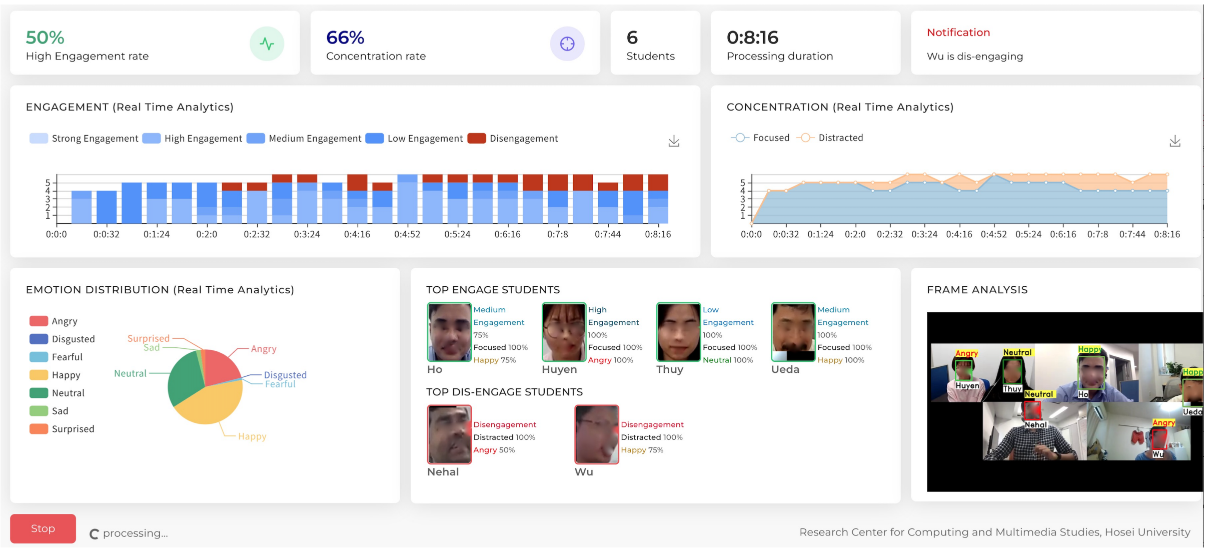
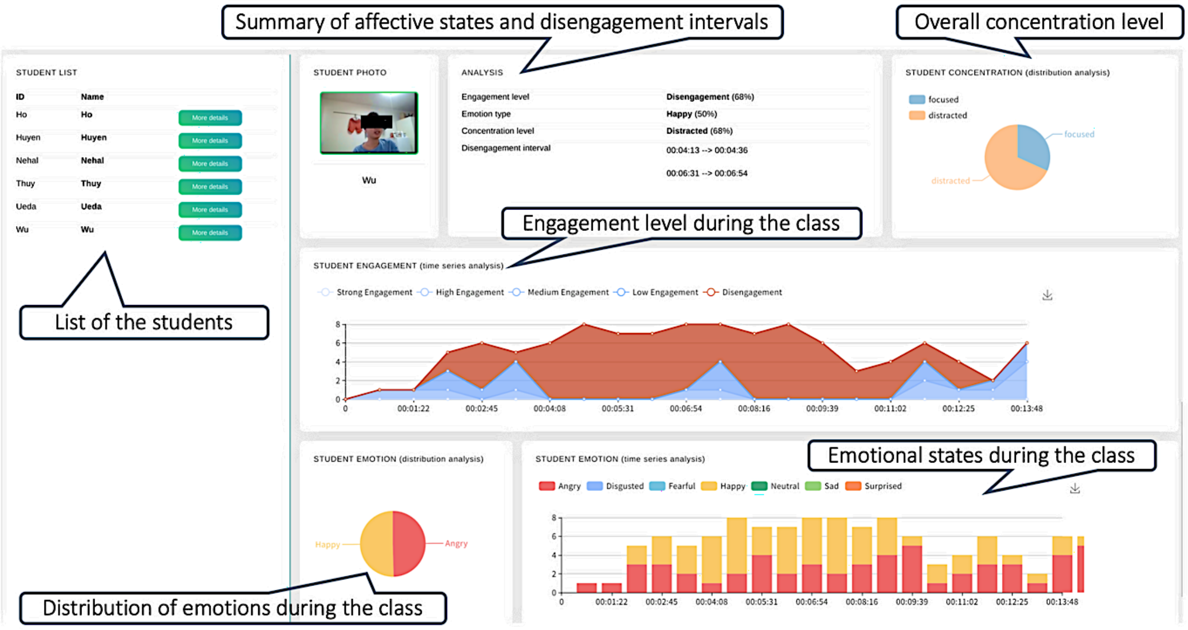

50%
High Engagement rate
66%
Concentration rate
6
Students
0:8:16
Processing duration
Notification
Wu is dis-engaging
ENGAGEMENT (Real Time Analytics)
Strong Engagement
High Engagement
Medium Engagement
Low Engagement
Disengagement
0:0:0
0:0:32
0:1:24
0:2:0
0:2:32
0:3:24
0:4:16
0:4:52
0:5:24
0:6:16
0:7:8
0:7:44
0:8:16
CONCENTRATION (Real Time Analytics)
Focused
Distracted
EMOTION DISTRIBUTION (Real Time Analytics)
Angry
Disgusted
Fearful
Happy
Neutral
Sad
Surprised
TOP ENGAGE STUDENTS
Medium Engagement 75% Focused 100% Happy 75%
Ho
High Engagement 100% Focused 100% Angry 100%
Huyen
Low Engagement 100% Focused 100% Neutral 100%
Thuy
Medium Engagement 100% Focused 100% Happy 100%
Ueda
TOP DIS-ENGAGE STUDENTS
Disengagement Distracted 100% Angry 50%
Nehal
Disengagement Distracted 100% Happy 75%
Wu
FRAME ANALYSIS
Stop
processing...
Research Center for Computing and Multimedia Studies, Hosei University

Figure 2: The Dashboard of MOEMO System [4].

In MOEMO, a closer look at a student's learning and engagement is also possible. Figure 2 shows a learning analysis of a student [13]. This is how a teacher could monitor a classroom. Moreover, a teacher could understand when a student was disengaged and had negative emotions during a lecture.



STUDENT LIST

ID	Name	
Ho	Ho	More details
Huyen	Huyen	More details
Nehal	Nehal	More details
Thuy	Thuy	More details
Ueda	Ueda	More details
Wu	Wu	More details

STUDENT PHOTO

Wu

ANALYSIS

Engagement level	**Disengagement** (68%)
Emotion type	**Happy** (50%)
Concentration level	**Distracted** (68%)
Disengagement interval	00:04:13 --> 00:04:36 00:06:31 --> 00:06:54

Figure 3: Example of a student's learning analysis [13].

4. Methodology

As stated earlier, the current version of the MOEMO system can produce insights into a student's emotional and affective states if they keep their camera on during the online lecture. It cannot generate results for those students who keep their web cameras off. This limitation led us to research the approaches to generate emotional states from their voices. We explored two approaches to detect students' emotions from their voices. The first approach is processing students' voices from scratch. And the second approach is processing voice using Zoom API, as the MOEMO system is integrated with Zoom.

4.1. Voice Processing from Scratch

This approach extracts raw audio data from a recorded lecture video. Then, we apply the silence detection mechanism to preprocess the data. After that, we use Speech-to-Text API on the non-silenced parts. By doing so, we get a new version of the audio transcript with clearer data. Then, we extract the timestamp and transcript from this newer version of the audio transcript. Finally, we match students' audio and visual data to maintain accuracy. Figure 4 shows the approach 1.

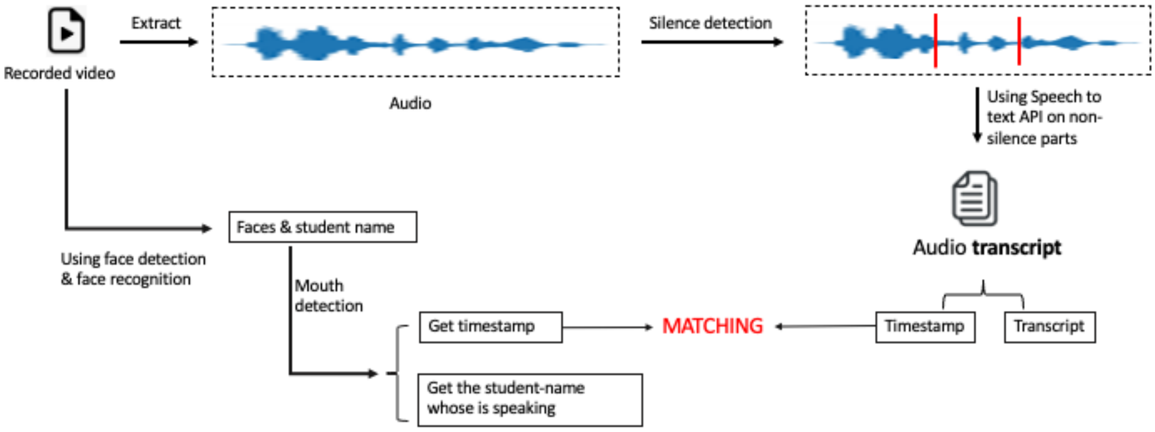


Figure 4: Voice Processing from Scratch

4.2. Voice Processing using Zoom API

This approach (Figure 5) uses Zoom API to process the data. We downloaded the transcript and recorded the video from the Zoom cloud. Then, we separated the video and audio data. We separated the timestamps, names of the students who spoke during the session, and transcripts from the audio data. Then, we send the data to MOEMO's analysis server.

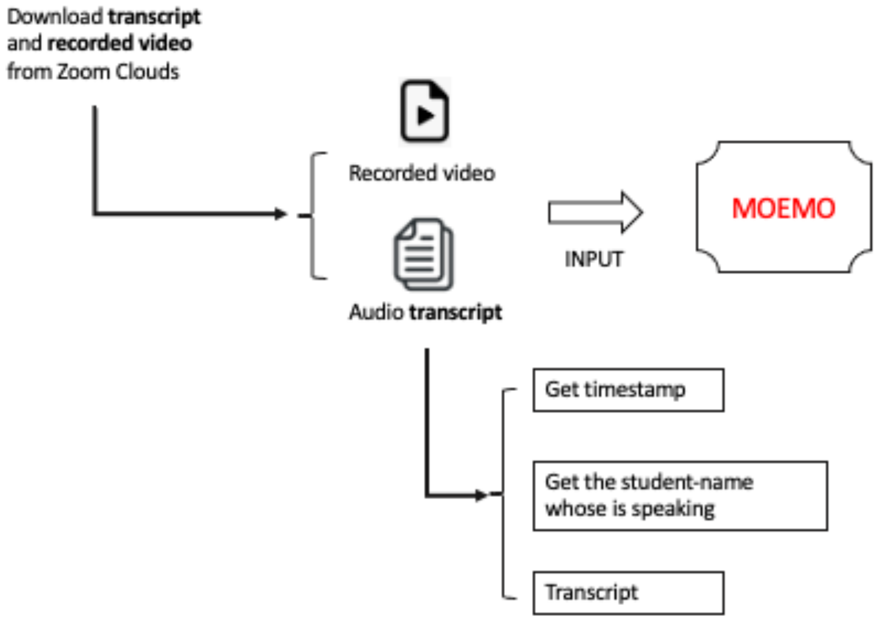

Download transcript and recorded video from Zoom Clouds
Recorded video
Audio transcript
INPUT
MOEMO
Get timestamp
Get the student-name whose is speaking
Transcript

Figure 5: Voice Processing using Zoom API

4.3. Rationale of the Approaches

We can analyze and compare the two approaches to emphasize their strengths and weaknesses. The first approach (Figure 4) is adaptable across various platforms, offering flexibility and customization for the system. It provides complete control, enabling fine-tuning of the entire processing pipeline. Also, it reduces reliance on external services, such as privacy concerns and ongoing costs. However, building and maintaining the whole processing pipeline requires more time and expertise. Also, due to the need for improvement and optimization, the accuracy may be lower than that of the pre-built solutions. This second approach (Figure 5) offers better performance due to pre-trained models and advanced processing capabilities provided by Zoom, saving time and resources due to faster deployment. However, the system will be tied to Zoom, limiting adaptability to other platforms. Also, this potentially involves subscription fees or API usage costs.

4.4. Concerns about Data Privacy in Audio Data

Since we use audio data from students, this can raise privacy concerns, as recordings might accidentally capture sensitive or personal conversations. Addressing these issues and proposing a data privacy policy for this system is important. For example, we should obtain explicit consent from all participants before collecting audio data and ensure data is securely deleted once it is no longer needed.

5. The VoiceMetrics System

5.1. Architecture

The architecture of the VoiceMetrics system integrates advanced technologies and methodologies to analyze student engagement and emotional states in an online learning environment. The system architecture is designed to process and analyze audio and textual data effectively, providing a comprehensive understanding of student interactions. Figure 6 presents the architecture of the VoiceMetrics system.

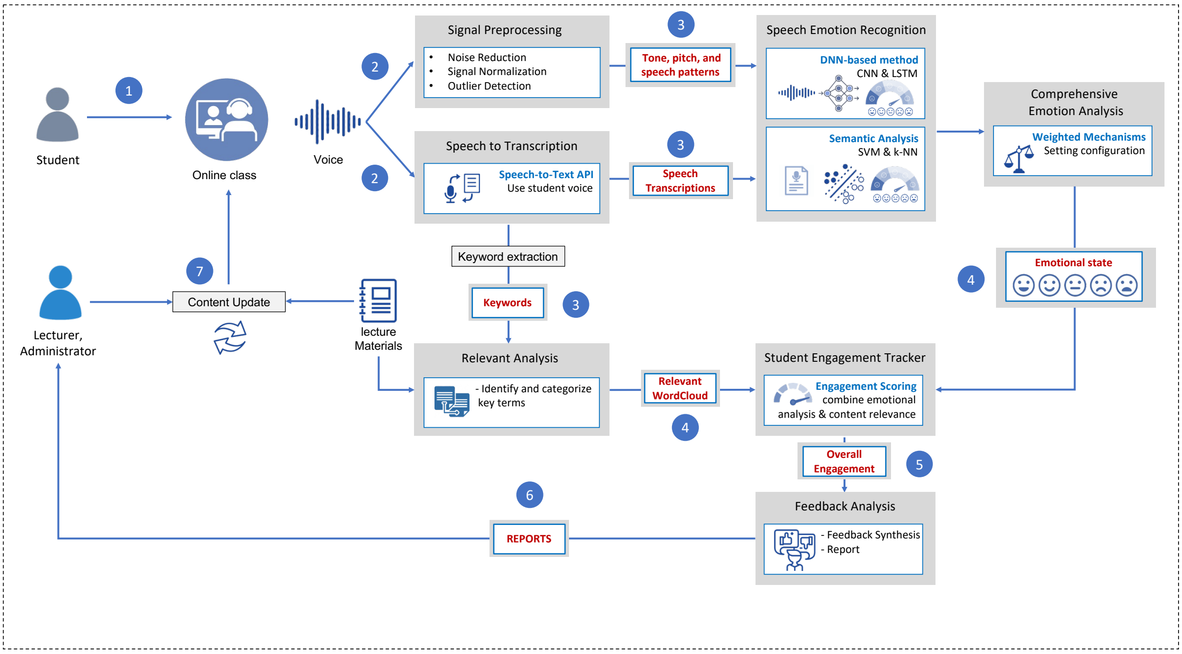


Figure 6: The low-level architecture

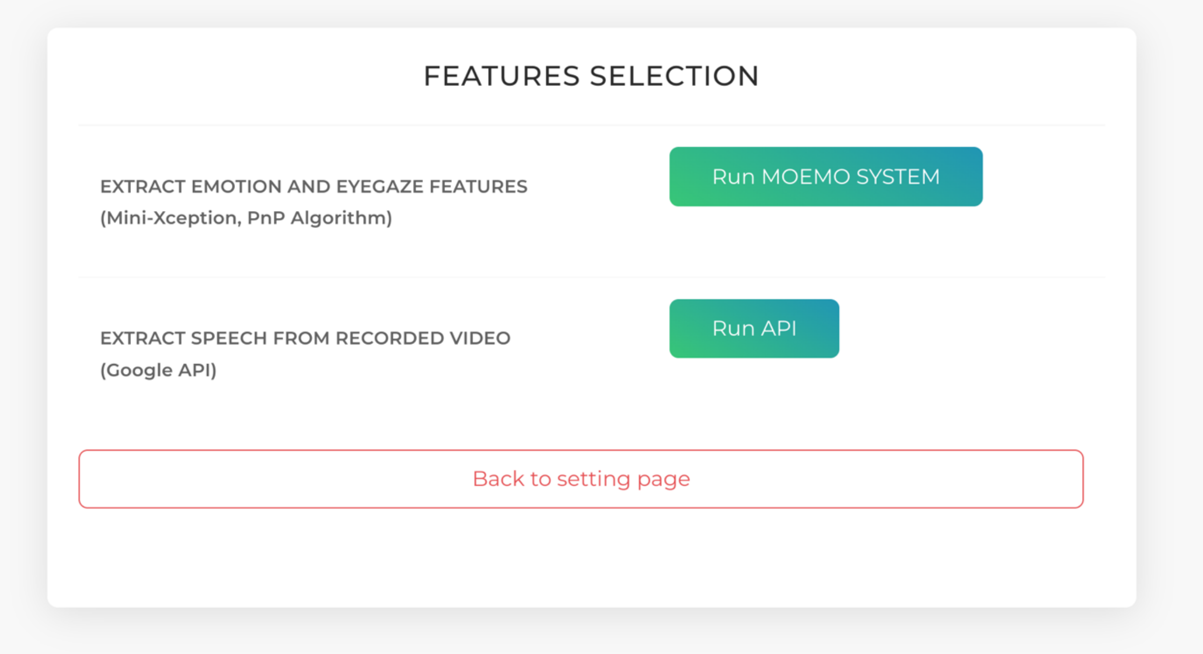
5.2. Functionality and Workflow

The VoiceMetrics system analyzes student interactions during online classes, focusing mainly on voice data to ascertain emotional and engagement levels. The process begins with a thorough preprocessing of the audio signal, where techniques such as noise reduction, outlier detection, and speech normalization are applied to ensure the clarity and reliability of the data. After preprocessing, the system employs deep learning-based methods to extract and analyze features from the cleaned audio signals. This analysis aims to detect various emotional states from the students' voices, providing real-time feedback on their engagement and well-being. At the same time, the VoiceMetrics system converts spoken words into text transcriptions using advanced speech recognition technologies. These transcriptions are then analyzed to assess the sentiment of the conversation, helping to determine the emotional content conveyed by the students. This dual analysis of voice tone and spoken content allows the VoiceMetrics system to offer a nuanced view of student emotions. Voice emotion analysis and speech transcription outputs are integrated to comprehensively view a student's emotional state.

In addition, the system analyzes the relevance of the speech content to the lesson topics by extracting keywords and generating word clouds. This analysis helps assess how closely student interactions are related to the lesson content, indicating their engagement level. The Engagement Tracker module synthesizes data on emotional states and topic relevance to evaluate each student's engagement level. This module is crucial for educators to understand how students interact with the course material. Furthermore, the Feedback Analysis module utilizes the engagement data to identify which parts of the lesson may require enhancements. By comparing student engagement levels with the content delivered, this module pinpoints areas where adjustments could maximize understanding and interest. The Emotion Feature Extraction Submodule leverages deep learning algorithms. This module extracts meaningful features from the preprocessed audio. These features typically include pitch, tone, and speech dynamics, which are crucial for determining emotional states.

5.3. Integration

We developed an interface to integrate the functions of the VoiceMetrics system with the existing MOEMO system so that the two systems can share each other's features. This interface integrates the emotional data with the insights from the word clouds to evaluate overall engagement. This assessment considers the intensity and context of emotional reactions about the relevance of the content being discussed. By correlating emotional states with key educational content at crucial moments, the system can discern whether high emotional responses correspond to critical learning points, gauging engagement depth and quality. Figure 7 shows the user interface of the integration module.



FEATURES SELECTION

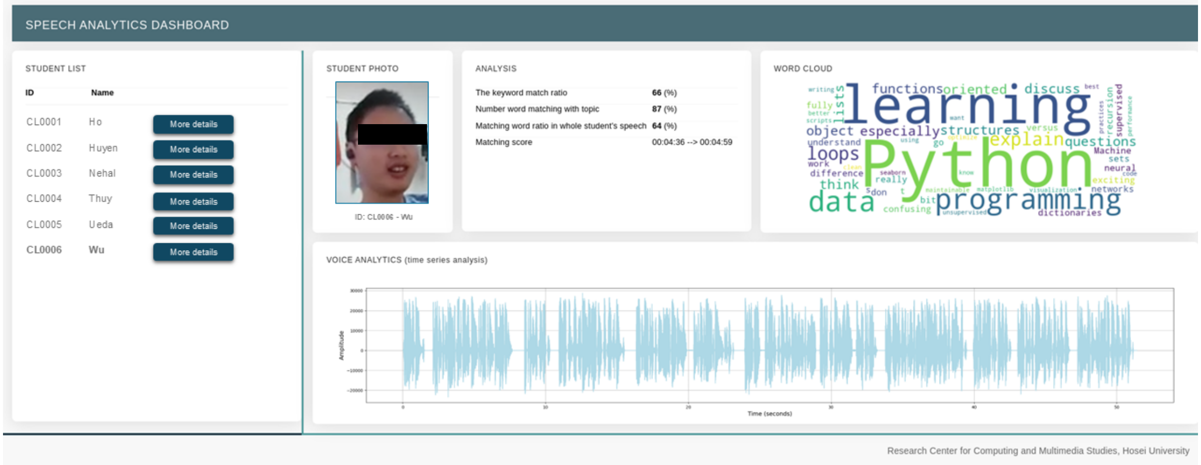
EXTRACT EMOTION AND EYEGAZE FEATURES
(Mini-Xception, PnP Algorithm)

EXTRACT SPEECH FROM RECORDED VIDEO
(Google API)

Figure 7: Integration Interface

5.4. Speech Analytics Dashboard

In addition, we developed a module called Speech Analytics Dashboard for VoiceMetrics system. The Speech Analytics Dashboard provides a comprehensive analysis of students' speech and includes key features, such as: 1) Keyword Summarization generates word clouds to visualize key topics discussed between students and teachers. By looking at the keywords, a teacher can understand whether the students' questions were related to the class content. 2) Lesson-related Metrics can analyze metrics related to the lesson, such as keyword match ratio, topic word match percentage, and more. 3) Flexible Time Frame Analysis supports analysis over the entire learning process or specific time intervals. The dashboard is an excellent tool for evaluating students' participation and their level of understanding during their studies. Figure 8 shows the dashboard presenting a result that we simulated.



SPEECH ANALYTICS DASHBOARD

STUDENT LIST

ID	Name	
CL0001	Ho	More details
CL0002	Huyen	More details
CL0003	Nehal	More details
CL0004	Thuy	More details
CL0005	Ueda	More details
CL0006	**Wu**	More details

STUDENT PHOTO

[redacted]

ID: CL0006 - Wu

ANALYSIS

The keyword match ratio	**66** (%)
Number word matching with topic	**87** (%)
Matching word ratio in whole student's speech	**64** (%)
Matching score	00:04:36 --> 00:04:59

WORD CLOUD

VOICE ANALYTICS (time series analysis)

Research Center for Computing and Multimedia Studies, Hosei University

Figure 8: Speech Analytics Dashboard

6. Discussion

This research demonstrates the feasibility and effectiveness of incorporating voice analysis into the MOEMO framework. It shows that it is possible to capture a broad spectrum of emotional states even in the absence of visual cues. This capability is crucial for educators, providing deeper insights into students' psychological and emotional states and enabling more informed and empathetic interactions. The proposed VoiceMetrics system's ability to analyze and integrate data from multiple sources—audio signals and speech transcriptions—into a coherent assessment of student engagement presents a robust model for future developments in educational technology. Looking ahead, the potential for expanding this technology to incorporate additional modalities and data sources offers exciting prospects for developing even more nuanced and adaptive educational tools. This study contributes to the broader discourse on educational technology by highlighting the importance of emotional recognition in enhancing student engagement and learning outcomes. It provides a foundation for future research to explore the complex interplay between emotion, engagement, and educational content, which is essential for the continued evolution of remote learning platforms.

The current study has several limitations. First, we acknowledge the importance of evaluating both the accuracy of speech recognition and the effectiveness of the dashboard as key areas for future research. To address this issue, we intend to measure speech recognition accuracy in the future. In the future, we also plan to develop an educator dashboard that provides educators with actionable insights and detailed reports to enhance teaching and learning experiences. This dashboard will allow educators to quickly grasp the dynamics of student interactions and adapt their teaching strategies effectively. Furthermore, we plan to investigate more student engagement levels, emotional states, topic relevance, and areas recommended for content enhancement.

7. References

[1] M. Tanaka, I. Takao, and M. Keisuke, “Understanding and utilizing students’ attitudes toward participation in discussions by using emotion analysis,” in Towards a new future in engineering education, new scenarios that european alliances of tech universities open up, Universitat Politècnica de Catalunya, 2022, pp. 1696–1703. Accessed: Nov. 25, 2024. [Online]. Available: https://upcommons.upc.edu/handle/2117/385816

[2] “Emotions and learning: what role do emotions play in how and why students learn?,” THE Campus Learn, Share, Connect. Accessed: Nov. 25, 2024. [Online]. Available: https://www.timeshighereducation.com/campus/emotions-and-learning-what-role-do-emotions-play-how-and-why-students-learn

[3] M. N. Hasnine, H. T. Bui, T. T. T. Tran, H. T. Nguyen, G. Akçapınar, and H. Ueda, “Students’ emotion extraction and visualization for engagement detection in online learning,” Procedia Computer Science, vol. 192, pp. 3423–3431, 2021. https://doi.org/10.1016/j.procs.2021.09.115

[4] M. N. Hasnine, H. T. Nguyen, T. T. T. Tran, H. T. Bui, G. Akçapınar, and H. Ueda, “A real-time learning analytics dashboard for automatic detection of online learners’ affective states,” Sensors, vol. 23, no. 9, p. 4243, 2023. https://doi.org/10.3390/s23094243

[5] S. Muñoz, E. Sánchez, and C. A. Iglesias, “An Emotion-Aware Learning Analytics System Based on Semantic Task Automation,” Electronics, vol. 9, no. 8, Art. no. 8, Aug. 2020, doi: 10.3390/electronics9081194.

[6] M. Ez-Zaouia, A. Tabard, and E. Lavoue, “Emodash: A dashboard supporting retrospective awareness of emotions in online learning,” International Journal of Human-Computer Studies, vol. 139, 2020. https://doi.org/10.1016/j.ijhcs.2020.102411

[7] A. Oliver Fredriksen Werner, “Design of a smart emotion-aware reflection system for teachers,” Master thesis, NTNU, 2023. Accessed: Nov. 26, 2024. [Online]. Available: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3097876

[8] B. Rienties and B. Alden, “Emotions used in Learning Analytics: a state-of-the-art review,” Measuring and understanding learner emotions: Evidence and prospects. Accessed: Nov. 26, 2024. [Online]. Available: https://oro.open.ac.uk/72634/

[9] H. Rashkin, A. Bosselut, M. Sap, K. Knight, and Y. Choi, “Modeling Naive Psychology of Characters in Simple Commonsense Stories,” May 16, 2018, arXiv: arXiv:1805.06533. doi: 10.48550/arXiv.1805.06533.

[10] R. Pekrun, A. C. Frenzel, T. Goetz, and R. P. Perry, “Chapter 2 - The Control-Value Theory of Achievement Emotions: An Integrative Approach to Emotions in Education,” in Emotion in Education, P. A. Schutz and R. Pekrun, Eds., in Educational Psychology. , Burlington: Academic Press, 2007, pp. 13–36. doi: 10.1016/B978-012372545-5/50003-4.

[11] M. J. Elias, Academic and social-emotional learning, vol. 11. International Academy of Education Brussels, Belgium, 2003. Accessed: Nov. 26, 2024. [Online]. Available: https://www.orientation94.org/uploaded/MakalatPdf/Manchurat/prac11e.pdf

[12] R. Pekrun and E. J. Stephens, “Academic emotions.,” 2012, Accessed: Nov. 26, 2024. [Online]. Available: https://psycnet.apa.org/record/2011-11778-001

[13] M. N. Hasnine, H. T. Nguyen, G. Akçapinar, R. Morita, and H. Ueda, “Classroom Monitoring using Emotional Data.,” in LAK Workshops, 2024, pp. 83–88. Accessed: Nov. 25, 2024. [Online]. Available: https://ceur-ws.org/Vol-3667/DC-LAK24-paper-10.pdf