

# **Multimodal Confidence Detection in Cross-Cultural E-Learning**

## **Environments**

### **Research Overview**

This study explores a multimodal method which is used to estimate the confidence of learners in e-learning environments by combining behavioral and physiological signals for confidence estimation. Based on the cross-cultural situation, this study analyzed how eye movement and rPPG signals can be combined to better support adaptive learning systems.

### **Research objective**

Confidence is an important indicator of learners' internal cognitive states, closely related to engagement and learning performance. Therefore, accurately identifying learners' confidence levels is essential for providing adaptive support and improving learning experiences in e-learning environments. However, traditional e-learning environments usually lack effective mechanisms to capture and respond to learners' confidence states. Recently, most studies rely on self-report or single data sources, which are limited in accuracy and versatility.

Recent studies in educational technology suggest that behavioral and physiological signals can provide richer information about learners' states. For example, eye movement data can reflect visual attention and cognitive processing, rPPG signals can reflect arousal and affective states. Nevertheless, the effective fusion of these two modalities remains challenging, particularly in real-world scenarios where rPPG signals are often unstable and noisy.

In addition, cultural differences may influence learners' behavioral and physiological responses during learning. However, limited research has explored how these methods work for learners from different cultural backgrounds. As education becomes more international, it

is important to understand these differences.

Moreover, estimating learners' confidence can support more effective learning analytics and adaptive interventions. For example, identifying low-confidence states may help detect learning difficulties early, while high-confidence states may indicate successful understanding. Reliable confidence estimation is therefore important for improving both learning experiences and adaptive learning system design.

To address these challenges, this study aims to develop a multimodal method to estimate confidence levels by combining eye movement and rPPG signals. The proposed approach is further evaluated in a cross-cultural learning context.

## **Research methods**

To evaluate the proposed multimodal confidence estimation method, this study conducts an experiment in an e-learning setting. In each session, participants watch an educational video, answer questions, and report their confidence levels, which are used as ground-truth labels for confidence estimation. The participants include Japanese students and international students. Each participant completes multiple sessions, providing repeated measurements of learner behavior and confidence responses.

The learning materials consist of short educational lecture clips from different subject areas, such as science, economics, and humanities.

Facial videos were recorded using the webcam, and both eye movement and rPPG signals were extracted using video-based methods. The eye movement features reflect learner's visual patterns, while rPPG signals provide additional physiological signals.

We propose a multimodal framework that combines eye movement and rPPG signals for confidence estimation. Eye movement data are the main source of information because of their stability, while rPPG signals provide additional physiological signals. The two modalities are

processed separately and then combined in the model.

Because rPPG signals are sensitive to noise and video quality variations, we introduce a mechanism which can dynamically adjust the contribution of rPPG features based on signal quality, allowing the model to rely more on physiological information when the signal quality is high and reduce its contribution when the signal quality is low. This design enables the model to reduce the influence of unreliable rPPG signals during confidence estimation.

## **Results and discussion**

Preliminary observations suggest that combining eye movement and rPPG signals may support multimodal confidence estimation in online learning environments. Eye movement data provide relatively stable behavioral information, while rPPG signals may offer additional physiological information when signal quality is good.

This study proposes a multimodal confidence estimation approach that dynamically adjusts the contribution of physiological signals based on signal quality. The proposed approach is further examined in a cross-cultural learning context to explore differences in confidence-related patterns among learners from different backgrounds.