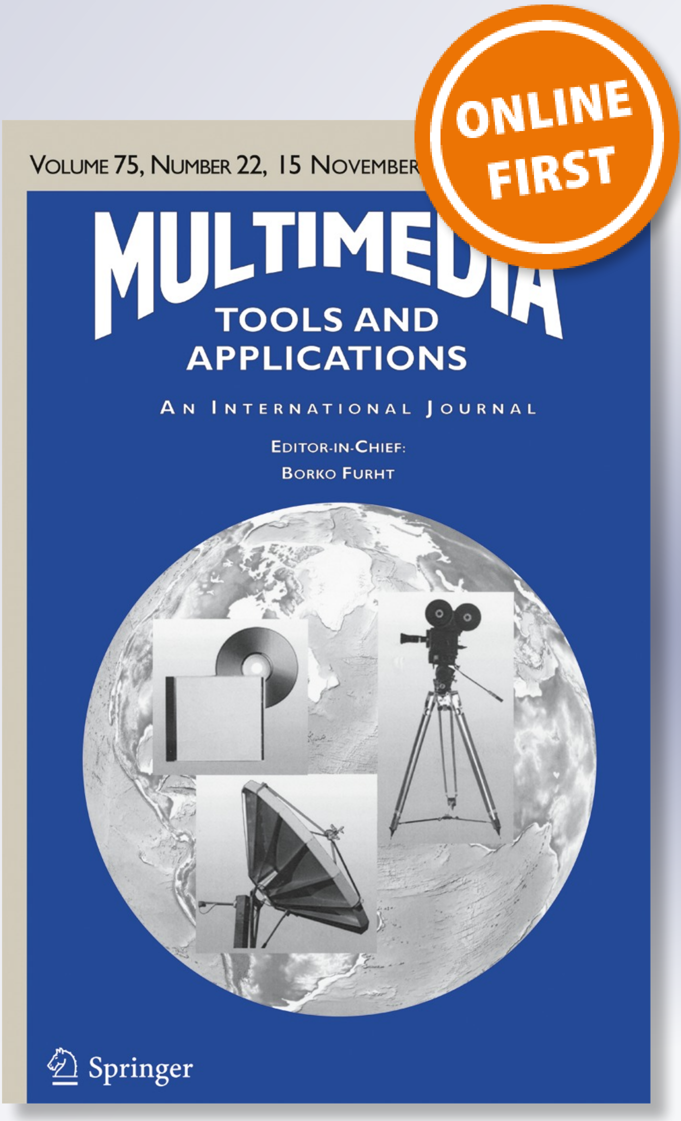
Augmenting teacher-student interaction in digital learning through affective computing

Jason Chi-Shun Hung, Kun-Hsiang Chiang, Yi-Hung Huang & Kuan-Cheng Lin

Multimedia Tools and Applications
An International Journal

ISSN 1380-7501

Multimed Tools Appl
DOI 10.1007/s11042-016-4101-z



Springer

Multimed Tools Appl
DOI 10.1007/s11042-016-4101-z

Augmenting teacher-student interaction in digital learning through affective computing

Jason Chi-Shun Hung[1] **· Kun-Hsiang Chiang**[2] **·**
Yi-Hung Huang[3] **· Kuan-Cheng Lin**[2]

Received: 28 July 2016 / Revised: 24 October 2016 / Accepted: 27 October 2016

Abstract Interactions between teachers and students can be effectively enhanced if teachers can capture the spontaneous nonverbal behaviors (e.g., facial expressions and body language) of their students in real time, thereby effectively improving teaching strategies and the learning effectiveness of students. In this study, we implemented an expression–response analysis system (ERAS) to analyze facial expressions. The ERAS employs a web camera to capture the facial images of students. Their facial expressions are analyzed to assess their attitude toward progressively more difficult course content, and to determine the relationship between their social interactions and learning effectiveness. The ERAS identified 10 facial feature points that form 11 facial action units (AUs). Subsequently, the AUs were classified as positive, neutral, and negative social interactions by applying a rule-based expert system, and cognitive load theory was applied to verify the classifications. The experimental results showed that student with high coding abilities could adapt to the multimedia digital learning content, as evidenced by the comparatively higher expression of neutral and positive social interactions, whereas students with low coding abilities reported a higher frequency of negative social interactions resulting from the increase in cognitive load. Simultaneously, the real time detection of social interactions can provide a basis for diagnosing student learning difficulties and assist teachers in adjusting their teaching strategies.

Keywords Affective computing · Cognitive load · Facial expression · Digital learning · Teacher-student interaction

✉ Kuan-Cheng Lin
kclin@nchu.edu.tw

[1] Department of Information Technology, Overseas Chinese University, Taichung, Taiwan

[2] Department of Management Information Systems, National Chung Hsing University, Taichung, Taiwan

[3] Department of Mathematics Education, National Taichung University of Education, Taichung, Taiwan

Published online: 11 November 2016

1 Introduction

In daily communication, people typically analyze the inner affective states of other people by observing their spontaneous or unconscious nonverbal behaviors, such as their facial expressions or body language. This type of communication and interaction is called social interaction. For example, people can determine whether a person agrees or disagrees with a concept based on his or her facial expression [2, 7]. Social interactions occur during face-to-face instruction in school classrooms, and teachers observe their students' nonverbal behaviors to determine whether they understand the course content.

To be effective in the classroom, teachers often ask students whether they have questions. Their intention is to understand how effective their students are at learning, and subsequently determine to proceed with new content or reinforce already-taught content; however, students frequently do not indicate they have questions, which can be frustrating for teachers. In digital teaching environments, teachers can determine whether their students understand the lesson content by applying affective computing [13], social signal processing [17], and other information technologies.

Teaching is a type of information transmission. Teachers transmit content to their students via information channels; students process this information, and store it in their recognition memory. If the information stimuli accords with a previous positive experience, the students experience positive emotions; otherwise, they experience negative emotions. Nonverbal emotional cues are difficult to perceive, although facial expressions are detectable. Picard [13] elaborated on affective computing, which can be employed to detect people's emotions. It can be further applied in computer-assisted learning, information retrieval, and computer–human interactions. Improvements in computer computational efficiency have resulted in numerous studies on facial expression recognition. In these studies, web cameras were employed to record human faces, and various computerized algorithms were applied to classify their facial expressions and further evaluate their emotional reactions. Digital learning can be augmented by affective computing-based human–machine interfaces to obtain effective and objective feedback during teaching activities, which can provide a reference for adjusting learning strategies and enhancing teaching skills. Facial expression recognition systems can analyze learning status, learning difficulties, mental states, emotions, and attention among students based on facial expression features [8–10, 12, 14, 19, 20].

Students learn from multimedia textbooks differently, because of differences in their coding abilities when assessing information stimuli. Coding is the first process in the cognition of information. Additional effort is spent when coding tasks are more complex, which also affects how effective a student is at learning. Learning is a mental activity, and expressed behaviors can indicate whether a student has successfully learned a concept; for example, a student's eyes and facial expression might indicate a sense of pleasure or satisfaction.

Previous studies on affective computing have employed simple web cameras to capture images. Because of limitations in image resolution, quality, and evenness of light, certain expressions which were meant to be representative of real facial expressions were inevitably exaggerated by researchers, to enable detection and recognition. In this study, we captured and analyzed real facial expressions of students in a multimedia learning environment. To elicit noticeable differences in the students' facial expressions during the learning process, we selected a series of increasingly difficult multimedia English-learning textbooks as an information source; the students involved in this study were Taiwanese, and were unfamiliar with the content. The reactions that were identified from the coding results and cognitive load were exhibited through the students' facial expressions, which were detected using a facial expression recognition system. In this paper, the following questions are discussed:

1) How does the system detect and recognize facial expressions and reactions in real learning environments?
2) What are the reactions of students with different English listening and visual skills when learning from multimedia textbooks of varying difficulty?
3) What is the relationship between facial expressions and reactions in learning?

2 Background and related work

2.1 Cognitive theory of multimedia learning

Multimedia learning refers to student learning from both words and pictures; thus, multimedia learning is also termed *dual-code learning* or *dual-channel learning*. Words can be categorized as either written (i.e., visual stimulus) or spoken (i.e., auditory stimulus), whereas pictures can be categorized as static (e.g., illustrations, coordinate graphs, diagrams, photos, and maps) or dynamic (e.g., animation and video); both of these constitute human information processing (Fig. 1). Multimedia information are sensed and stored as aural and visual memories in students' minds. Aural and visual memories are processed using word and pictorial models, respectively. If a student wants to learn a concept, they must integrate both of these models with their prior knowledge (i.e., previous experiences) in the long-term memory of the brain by establishing this comprehensible and meaningful knowledge in their working memory, and then storing this knowledge as a new schema in their long-term memory [11].

Decoding is the process of restoring the output symbols of a signal source. This process changes based on personal experience, and it is essential in forming facial expressions. Facial expressions result when an addressee interprets information based on his or her previous experience, and this interpretation is converted into an internal feeling. In theory, the working memory controls the senses, and elicits a specific reaction after a concept is integrated into long-term memory. These reactions include facial expressions, sounds, physical actions, and physiological responses. In this study, we focused on facial expressions. Multimedia English textbooks were used as stimuli to elicit various responses to indicate the differences between the visual and aural abilities of two students, in the context of learning a foreign language. The recorded facial expressions were given to the teachers to provide a reference for adjusting their teaching strategies, and for evaluating the effectiveness of the students' learning of, and reactions to, the content of the textbooks.

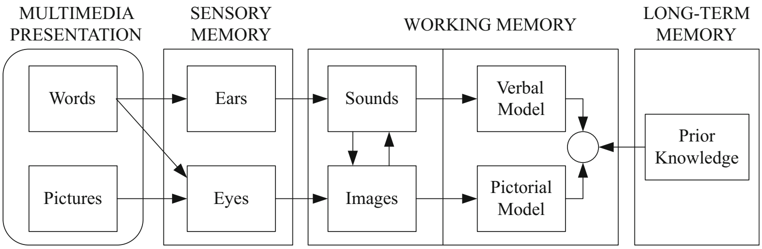


Fig. 1 Cognitive theory of multimedia learning (ref. [2])

2.2 Cognitive load theory

Sweller, van Merriënboer, and Paas [16] conceptualized cognitive load as mental load, which occurs when a specific task is imposed upon a learner's cognitive system. Learning effectiveness diminishes when the load exceeds a specific threshold. Gerjets and Scheiter [6] extended this theory, and proposed the cognitive load theory framework for teaching design, textbook content, and textbook complexity, as shown in Fig. 2.

That study also indicated that cognitive load is excessive if the information received by a learner exceeds his or her cognitive capacity. Consequently, the person becomes unable to recognize the meaning of new information, and the information process is disrupted. Conversely, if the total cognitive load is within a person's cognitive capacity, they can continue processing the information effectively. In information processing theory, cognitive load refers to learning load in the working memory. The working memory has limited storage capacity, and overload can result in excessive cognitive load (i.e., high cognitive load). When this happens, learners lose their ability to concentrate, which may have a negative emotional effect. Hence, their learning effectiveness is reduced [16].

Brunken et al. [3] classified two methods for measuring the cognitive load (i.e., mental load) spent by learners: 1) Subjective measurement: This method is suitable for assessing learners who can recall their cognition during learning, and can clearly indicate the consumption cost of their mental efforts. For example, if a questionnaire survey was conducted, learners would be capable of accurately indicating their metal efforts while learning, as well as the textbook difficulty, and the level of their ability. 2) Objective measurement: This method assumes that changes in cognitive load can affect physiological changes (e.g., heartbeat, brain activity, eye movement, and behavior).

In the proposed expression–response analysis system (ERAS), the objective method was applied. Additionally, the subjective method was employed to verify the feasibility of the proposed ERAS by comparing the measurements obtained using both methods.

2.3 Facial expression recognition and facial action coding system

Affective computing is a type of artificial intelligence. It requires considerable computational resources, and applies deduction and learning to augment perception, the most critical component of which is affective perception. Affective perception signals induce a range of

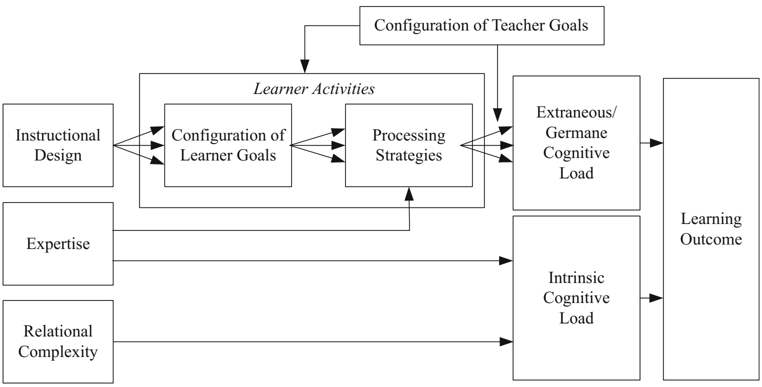


Fig. 2 Theoretical framework of cognitive load (ref. [6])

objective physiological changes (e.g., changes in facial expression, sound, pulse, breathing, brain wave activity, and movement of the eye or body). Emotion is the cognitive appraisal of sensory stimuli, and is a reaction based on prior emotional experiences. Reactions can manifest as facial expressions (e.g., happiness, anger, sadness, fear, surprise, and disgust), or through the visceral nervous system as physiological changes (e.g., changes in pulse rate, breathing, blood pressure, pupil dilation, body temperature, brain wave activity, or muscular as well as skin electrical conductivity). Facial expressions are a natural manifestation of inner emotional reactions. Emotional expressions typically involve movement of the mouth, eyes, and eyebrows. Thus, the proposed ERAS must be capable of recognizing and analyzing these three areas.

Ekman taxonomized various combinations of facial expressions and proposed the facial action coding system (FACS) [4, 5]. The FACS includes numerous individual or combined facial action units (AUs) that are relevant to this study because they facilitate the recognition of specific facial expressions. Ekman and Friesen defined 44 facial AUs of the upper and lower face, as well as the following six basic categories of emotional facial expressions: 1) happiness, 2) anger, 3) sadness, 4) fear, 5) surprise, and 6) disgust. Because of the complex relationship among the various types of human facial expressions, the AUs can be combined. Thus, the description of facial expressions and degree of variation may differ across studies. Although the FACS defined 44 specific facial AUs, researchers can select AUs that are related to their operational goals, and the optimal recognition effect can be achieved.

3 The design of expression response analysis system

The proposed ERAS was implemented based on the following three stages: 1) face detection, 2) feature extraction and positioning, and 3) classification and recognition of facial expressions.

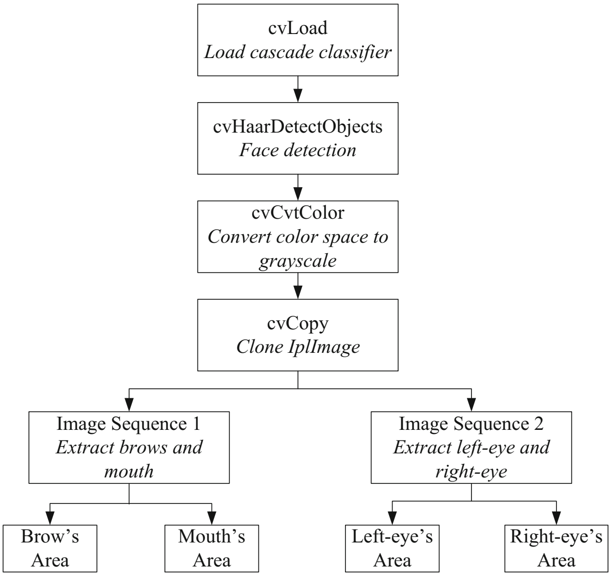
3.1 Facial detection

OpenCV AdaBoost, a machine-learning algorithm, and Haar-Like were employed for face detection [1, 18]; OpenCV was employed to train the face detection cascade classifier to detect the face, eyes (i.e., left eye, right eye, and both eyes), nose, and mouth and identify feature positions. Static face images were extracted, and areas of the eyes, nose, and mouth were examined. The total mean detection time was approximately 0.6 seconds, and the detection frequency was extremely high. Because the eyes and eyebrows are close to each other, the image sequence was duplicated, and four areas were extracted (eyebrows, mouth, left eye, and right eye) to avoid mutual interference in these two extraction areas. Figure 3 shows the flow chart for the face detection and feature extraction process.

3.2 Feature extraction and position

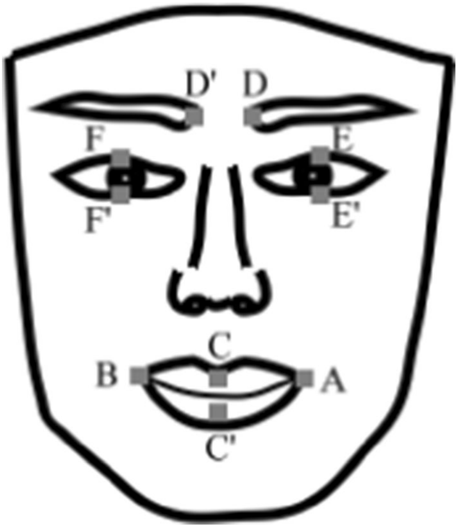
Changes in facial expression were observed in three key areas: the eyebrows, eyes, and mouth. Feature points were applied to perform image processing and displacement measurements for these areas. Accordingly, ten feature points were selected as a basis of measurement (Fig. 4). Points A and B represented the left and right corners of the mouth, respectively. When the corners of mouth move upward, Points A and B rise. The upper (Point C) and lower (Point C') lips were linked and changes in the length of CC' represent the mouth opening or closing. The

Fig. 3 Flow chart for face detection and feature extraction



eyelids were also linked, and changes in the length of EE' or FF' respectively represented the left or right eye opening or closing. In addition, the duration that the eyes were closed was examined to determine whether the eyes were blinking or closed. The inner corners of the eyebrows are critical indicators in identifying facial expressions. Changes in the eyebrows were observed based on the upward and downward movement of Points D and D', as well as the length of Line DD'.

Fig. 4 Location of face feature points



3.3 Feature distance vector

Feature vectors are critical for classifying facial expressions. Several feature vectors were implemented to correspond to AUs, as follows: 1) AU12 represented a person smiling. The corners of the mouth move upward, which causes *vtAB* to expand and enlarge, (*vtA* and *vtB* also enlarge); 2) AU20 represented a swallowing action. Although *vtAB* is enlarged, the corners of the mouth do not move upward; thus, *vtA* and *vtB* are not enlarged; 3) AU12 + 24 also represented a person smiling. The upper and lower lips become thin, and *vtCC'* becomes shorter; 4) AU12 + 25,26,27 represented a person laughing. The upper and lower lips are opened, and *vtCC'* increases; and 5) AU15 + 25,26,27 represented a person yawning or feeling sad. The opening of the upper and lower lips causes *vtCC'* to increase. However, the corners of mouth move downward. Figure 5 shows the original data for the *vtAB* feature vector.

Real time image detection can be affected by movements such as breathing. Consequently, the performance of the subsequent classification and identification could be impaired by changes in the feature vectors. To address this, the participants maintained a neutral facial expression for 2–3 seconds to establish the neutral value *neu* as a benchmark for calculating the threshold value *cnv* (a ± 10 % upper and lower threshold was applied), as shown in Equation 1.

Figure 6 shows the changes in vector *vt* after calculating *cnv*. Figures 7 and 8 show the feature vector changes for the mouth and eyelids, respectively.

3.4 Expression recognition and classification

This study examined the reactions of students, represented by their natural facial expressions, after learning from various textbooks. A rule-based expert system was employed to identify the feature vectors of the discussed feature areas. The FACS AUs applied in the recognition process are detailed as follows.

3.4.1 AUs for the brow (AU1, AU4)

The brow AUs were used to identify the upward and downward changes of Points D and D', as well as changes in the length of DD'. However, because the length of DD' might remain constant, the identification rule was divided into two parts for the feature classification process.

3.4.2 AUs for the eyes (AU5, AU7, blinking, closed)

Eye movement is rapid, and the maintenance time (i.e., the time between calculations) for analyzing this action requires specific clarification. If the length of EE' or FF' is within the threshold value (e.g., when the eye is fully opened), the maintenance time is relatively short; conversely, a comparatively short distance (e.g., when the eye is closed or almost closed) requires a longer maintenance time. Furthermore, because these considerations are exclusive, it

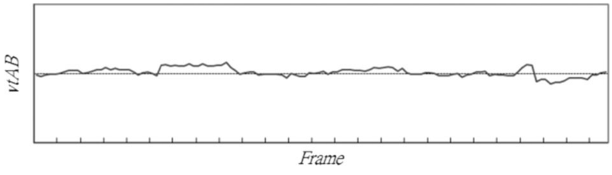
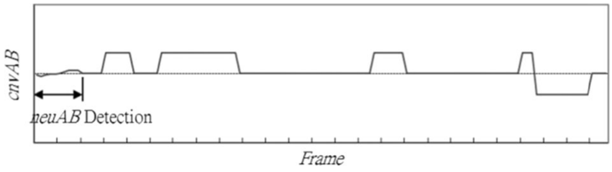


Fig. 5 Diagram of vt image sequence (before conversion)

Fig. 6 Diagram of cnv image sequence (after conversion)



is impossible for AU5 (i.e., the eyes are opened) and AU7 (i.e., the eyelids are tightening, blinking, or closing) to occur in the same frame. The frame rate of the web camera was set at 10 frames per second, and the maintenance time was set at 0.5 seconds (approximately 5 frames). When the length of EE' or FF' was ≤ 0.01 mm and the maintenance time was ≤ 0.5 seconds, the eye was considered blinking; conversely, the eye was considered closed if the maintenance time > 0.5 seconds.

3.4.3 AUs for the mouth (AU12, AU15, AU20)

The action of the corners of the mouth is crucial in producing mouth expressions, the categorization of which is based on the upward or downward movement of Points A and B, as well as the length of AB. When people feel pleasure, the corners of mouth typically move upward (e.g., when smiling); conversely, when people feel displeasure, there is either no facial expression, or the corners of mouth move downward. In addition, when people feel bored or uninterested, the corners of their mouth may stretch, indicating escape.

3.4.4 AUs for the lips (AU24, AU25, AU26, AU27)

Opening and closing the lips enhances the reaction range of mouth expressions, as determined by the length of CC'. For example, if CC' < *neuCC'*, the lips are tense; conversely, if CC' > *neuCC'*, the mouth is open, where AU25, AU26, and AU27 indicate a small part in the lips, dropped jaw, and stretched mouth, respectively.

3.4.5 AU combinations

Ekman and Friesen defined 44 AUs for the upper face and lower face, as well as six basic emotions (happiness, anger, sadness, fear, surprise, and disgust). This paper summarizes 11

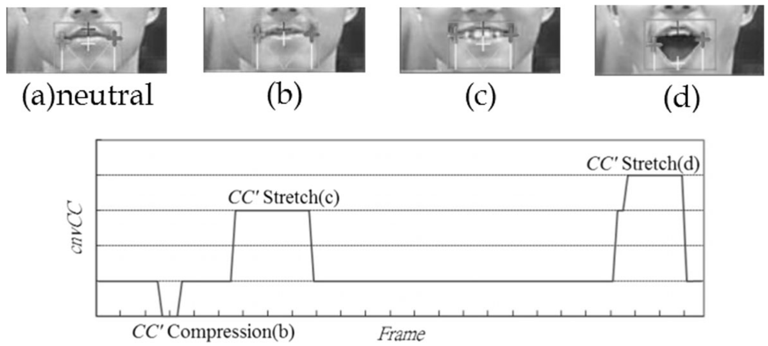


Fig. 7 Diagram of change in mouth feature vector

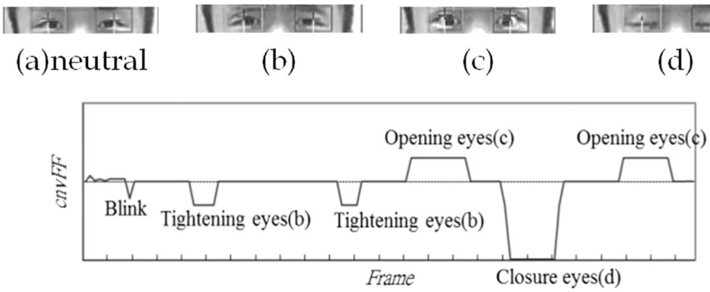


Fig. 8 Diagram of change in feature vector of upper and lower eyelids

AUs based on the facial expressions of the students after they received information stimuli. The AU combinations were defined based on six basic emotions (Table 1).

4 The implementation of expression response analysis system

The ERAS contains the following three modules: 1) image extraction; 2) facial expression recognition; and 3) emotional reaction analysis.

4.1 Image extraction module

The image extraction module is a simple front-end interface that allows the user to adjust the image location, and includes warning functions. This module was installed on students' computers. Because the interface is simple to operate, it imposes a minimal psychological effect on students, and it ensures that the image and position of the student's face are consistent. The facial expressions of the students are saved in AVI format, which is convenient for future processing and application. Figure 9a shows that when the light is green, the position

Table 1 AU combinations of basic emotions

Location	Emotions	Positive		Negative			
		Happiness	Surprise	Fear	Anger	Sadness	Disgust
Upper Face	AU1	●	●	●		●	
	AU4			●	●	●	●
	AU 5		●	●	●		
	AU 7	●				●	●
Lower face	AU 12	●					
	AU 15		●	●	●	●	●
	AU 20			●			●
	AU 24	●		●	●	●	●
	AU 25	●	●	●	●	●	●
	AU 26	●	●	●	●	●	
	AU 27	●	●	●	●	●	

of the student's face is suitable, and the distance to the lens is normal (i.e., approximately 60–70 cm); when the light is yellow, the position of the student's face has deviated, although it is acceptable; however, when the light is red, the student's face cannot be detected, or the system cannot perform processing and recognition. Thus, students must avoid the red light. During program execution, the correction window (Fig. 9b) is displayed in the central upper edge of the screen (below the web camera). Its function is similar to a mirror, allowing the students to practice performing facial expressions in addition to adjusting the position of their face. After clicking "Record," the correction window is hidden to prevent interference during the test.

4.2 Facial expression recognition module

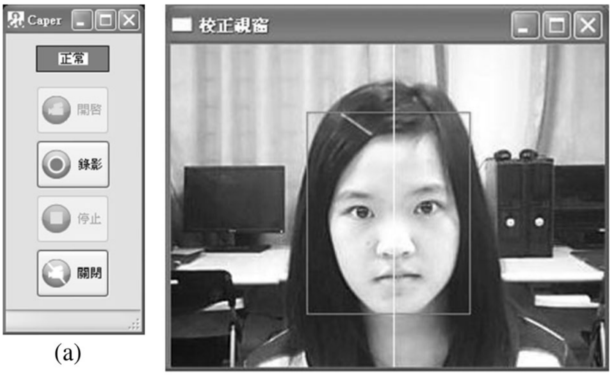
The facial expression recognition module is a back-end program. It primarily processes the facial expression images, including the image processing, marking the feature points, calculating the feature vectors, and conversion. Finally, all of the feature vectors are categorized as AUs, and the text files are inputted. Section 3.4 details the processing method. This module can be used for detecting dynamic images (Fig. 10).

4.3 Facial expression analysis module

The facial expression analysis module is a back-end program. Because of the close relationship between the facial expression and time sequence, facial expression analysis was conducted after all of the AUs were recognized. AUs were analyzed in each frame sequentially from Frame 1, and each frame was categorized as a facial expression by applying the facial expression analysis rules based on the AU combinations shown in Table 1. The facial expression analysis module (Fig. 11) used the AU recognition results from the facial expression recognition module to analyze the facial expressions.

5 ERAS teaching experiment

Grade 3 students were recruited from the data processing department at a vocational college in Taichung, Taiwan. The students had unimpaired sight and hearing and were proficient in using computers. From this group, one student (S1) with a low English score (low coding ability),



(a) (b)

Fig. 9 Operation picture of image extraction module

Fig. 10 Facial expression recognition module—dynamic image detection

and another student (S2) with a high English score (high coding ability) were selected to participate. Based on the difference between the two students' visual and aural abilities, we analyzed and compared their emotional reactions and facial expressions throughout the learning activities, and examined the results to evaluate the students' learning effectiveness and difficulties while learning from the textbooks.

Fig. 11 Operation picture of facial expression analysis

5.1 Teaching material and object

To elicit noticeable differences in the visual and aural abilities of the students, we selected three textbooks with progressive degrees of difficulty: "Chinese/Chinese" (Textbook 1), "Scenario/ English" (Textbook 2), and "English/English" (Textbook 3; Table 2). The three textbooks provided visual and auditory sensory stimuli, which formed the basis for discussing the social reactions (i.e., facial expressions) to assess the difference in their visual and hearing listening skills.

To ensure that the students understood the learning objectives, the teachers introduced the learning units, and assisted them in establishing the learning goal for each unit before distributing the textbooks. Following the experiments, a subjective measurement method was employed to assess the imposed cognitive load. A postlearning test was also employed to assist in evaluating how effective the students were at learning the lesson content. Several questions were designed to determine whether the students clearly understood the content, which formed the basis for appraising the learning effectiveness. Students with low and high coding abilities were distinguished to provide a reference for comparing the facial expressions based on the experimental results. Table 3 shows information on the students' previous academic performance.

In the experiments, we captured the facial expressions of S1 and S2 while they learned from the three textbooks, and observed the time sequence of facial expressions in reaction to the textbook content. During the experiment, we examined the students' facial expressions under normal learning conditions. The students were not requested the students to intentionally constrain or perform certain facial expressions; thus, subtle changes in facial expression and lighting conditions could affect the detected values. The learning state of the students was observed based on the frequency and interval between facial expressions. Based on information processing and cognitive load theories, if the students adapted to the textbook content, they exhibited positive reactions or maintain neutral facial expressions; otherwise, they exhibited negative facial expressions. To estimate the cognitive load, the subjective

Table 2 Textbook introduction

Textbook 1	Textbook 2	Textbook 3
Chinese	Scenario	English
Chinese	English	English
Easy to learn	A little difficult	Extremely difficult
One-to-one multi-media textbook teaching; introduce computer VB Language-Arithmetic Operation unit.	One-to-one multi-media textbook teaching; play English scenario dialogue textbook without subtitle.	One-to-one multi-media textbook teaching; introduce computer data structure-Stacking unit.
Discuss facial expression reactions in learning without any restriction on visual and auditory sensory stimuli.	Discuss facial expression reactions in learning with slight restriction on visual and auditory sensory stimuli.	Discuss facial expression reactions in learning with restriction on visual and auditory sensory stimuli.
Easy	Easy	Medium
Easy	Medium	Difficult
81 sec	59 sec	96 sec

Table 3 Background data of tested students

Tested student No.	Average score of computer	Information ability	Average score of English	Language seeing and hearing abilities
S1 (Male)	88	Strong	60	Weak
S2 (Female)	94	Strong	78	Strong

measurement adopted a 3-point scale, with different scoring systems for Question 1 (*Yes* = 3, *Average* = 2, and *No* = 1) and Questions 2–4 (*Yes* = 1, *Average* = 2, and *No* = 1).

5.2 Facial expression and social reaction analysis

5.2.1 Experiment 1: facial expression for textbook 1

The textbook and audio content were presented in Chinese. The information input for the visual and auditory sensory organs was unrestricted; only their language visual and auditory information coding abilities differed, which exerted an effect on their cognitive load, thereby eliciting the observed facial expressions. The experimental results (Table 4) showed that compared with S2, S1 exhibited fewer neutral (n = 766) and more negative (n = 36) facial expressions; conversely, S2 exhibited a higher frequency of positive (n = 22) facial expressions. This indicated that S2 experienced less difficulty than S1 in adapting to the textbook.

The experimental results showed that 95.5 % and 97.3 % of facial expressions were neutral for S1 and S2, respectively. Both of the students concentrated on the learning activities. In a normal learning environment, ther emotional state was stable. Thus, a high frequency of neutral facial expressions was considered normal. The time sequence of the facial expressions revealed that S1 maintained a negative facial expression between 38 and 65 seconds (Fig. 12a), which coincided with increased blinking (Fig. 12b). The ERAS results categorized the negative facial expressions as "disgust." According to the subjective measurement results, S1 reported that he thought the teaching speed was too fast, and the increased cognitive load coincided with negative facial expressions. In the second half of the learning activity, the blinking decreased, indicating a reduction in cognitive load. During the period from 20–30 seconds, S2 appeared interested in the robotic animations of the operators, and her facial expression was positive (Fig. 13a). The frequency of blinking was approximately 8.7 times per minute (Fig. 13b), which was similar to that of a person who is attentively reading information on a screen. This indicated that the cognitive load of S2 was within her load capacity, and that she adapted to and concentrate on the textbook content.

The subjective measurement results revealed that both of the students fully understood the content of Textbook 1, and both of them answered all of the questions correctly. The subjective

Table 4 Detection results of facial expression reactions to textbook 1

	Image length (sec)	Cognitive load	Positive		Neutral	Negative				Blink
			Happiness	Surprise	None	Anger	Fear	Sadness	Disgust	
S1	82	7	0	0	766	0	0	0	36	16
S2	83	4	22	0	794	0	0	0	0	12

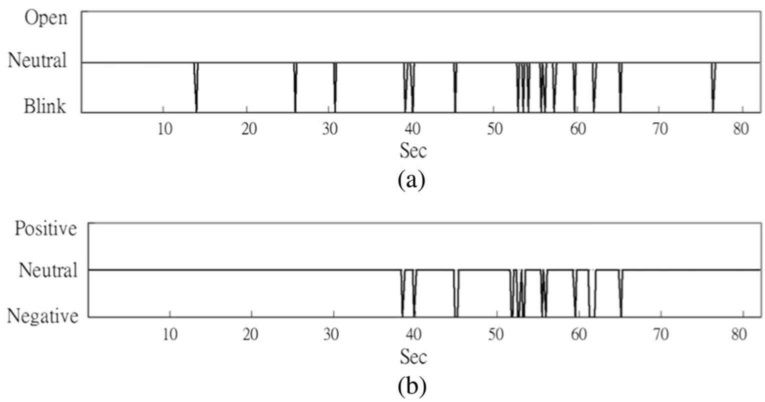


Fig. 12 Analysis chart of S1 facial expression to textbook 1

measurement results were consistent with the postlearning results (Table 5). Although both of the students exhibited many neutral facial expressions, their learning effectiveness was high when they concentrated on the learning activities.

5.2.2 Experiment 2: facial expressions for textbook 2

The content of Textbook 2 was considered of an intermediate level of difficulty. The visual content was scenario- and action-based (presented in Chinese), and the listening section was presented in English. This experiment was designed to create a situation where the students might experience difficulty understanding the auditory information. If the students could correctly code the auditory information, the visual aspect assisted them in increasing their learning effectiveness. However, if they incorrectly coded the auditory information, they may have misunderstood the textbook based on only visual information. Thus, the scenario could only be coded correctly by combining visual and auditory information.

The experimental results demonstrated that S1 exhibited 178 negative facial expressions toward the textbook, and S2 exhibited 40 positive facial expressions (Table 6) in addition to 542 neutral facial expressions. This shows S2 experienced less difficulty than S1 in adapting to

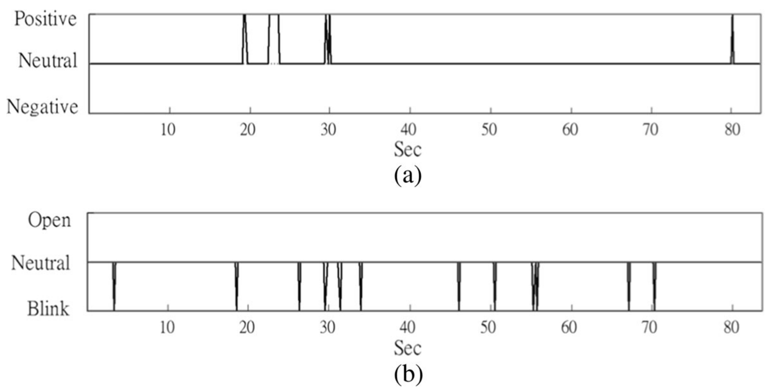


Fig. 13 Analysis chart of S2 facial expression to textbook 1

Table 5 Learning effectiveness of textbook 1 and answers in learning sheet

Item	S1	S2
Do you think teaching speed of the textbook is too quick?	Yes (3)	No (1)
Are you interested in contents of the textbook 1?	Fair (2)	Yes (1)
Can you understand all the contents of the textbook 1?	Yes (1)	Yes (1)
Can you understand teaching contents of the textbook 1?	Yes (1)	Yes (1)
Operators of the robot are / and Mod?	No(O)	No(O)
2 * (1 – 3) ^ 2 is equal to 8?	Yes(O)	Yes(O)
2 Mod 5 – 30 \ 3 * 2 is equal to -18?	No(O)	No(O)
The rule of operator precedence is exponent, negative, quotient, multiplication, division, addition and sub?	No(O)	No(O)
Are the operators with the same priority are calculated from left to right?	Yes(O)	Yes(O)
Does integer division means taking integer part of quotient?	Yes(O)	Yes(O)

the textbook. In addition, increasing the textbook difficulty resulted in increased negative facial expressions by S1.

Compared with the experimental results of Textbook 1, the percentage of negative facial expression exhibited by S1 increased by 31 %. Because the information for Textbook 2 was transmitted in spoken English, the auditory channel was restricted, thereby limiting the dual-channel information to a single channel, which reduced the students' learning effectiveness, and increased their cognitive load. Consequently, a higher frequency of negative facial expressions was observed. The duration of facial expressions indicate that S1 exhibited negative facial expressions from 17 to 35 seconds, and again from 39 to 50 seconds (Fig. 14a). After these two intervals, S2's facial expressions were positive (Fig. 15a). The first part of the textbook describes a hero who had forgotten his wallet, and the final part describes a scenario where the hero locked his car key in his car. The hero was involved in two events that elicited positive facial expressions from S2. The learning sheet results demonstrated that S1 answered all of the listening test questions incorrectly, although half of his answers were correct in the visual test (Table 7), indicating that he used his visual sensory channel to compensate for the auditory sensory channel deficiency. Consequently, the learning effectiveness was reduced, and the information became less interesting because it was difficult to understand. Thus, S1 did not exhibit the same positive facial expression as S2. This indicates that the students' facial expressions are relative to their cognitive load. S1 tended to exhibit negative facial expression because of the cognitive overload.

On average, S1 blinked 22.0 times per minute while learning the Textbook 2 content (Fig. 14b), compared with 11.7 times per minute while learning from Textbook 1. Thus, the blinking became

Table 6 Detection of facial expression reactions to textbook 2

	Image length (sec)	Cognitive load	Positive		Neutral	Negative				Blink
			Happiness	Surprise	None	Anger	Fear	Sadness	Disgust	
S1	60	8	0	0	396	0	0	0	178	22
S2	60	4	40	0	542	0	0	0	0	10

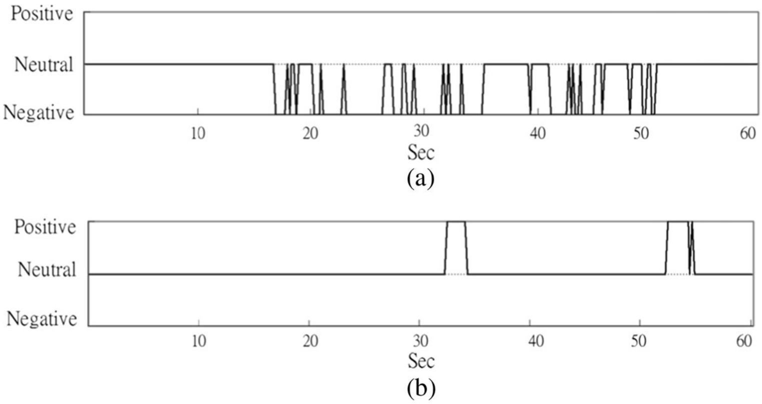


Fig. 14 Analysis chart of S1 facial expression to textbook 2

more frequent. S2 blinked 10.0 and 8.7 times per minute while learning from Textbook 2 (Fig. 15b) and Textbook 1. Thus, the difference was minor for S2. Moreover, S2 continued to exhibit positive facial expressions. These results revealed that S2 adapted to the Textbook 2 content. Additionally, she answered all of the questions correctly for Textbook 2 (Table 7).

The subjective measurement results of the learning cognition test showed that S1 fully understood the visual information, although he only partially understood the listening information. Furthermore, despite scoring 100 % on the visual perception test, none of the listening test answers were correct. This indicated that the subjective measurements overestimated the listening aspect. The subjective measurement results of the students' cognitive load indicated that S2 fully understood both the visual and listening information, which was consistent with the postlearning results (Table 7).

5.2.3 Experiment 3: facial expressions for textbook 3

In the third and final experiment, an advanced level textbook was employed to impose an increased cognitive load on the students to test their facial expressions. Specifically, both the visual and listening content were in English; thus, both the visual and auditory senses were restricted, and the students were required to code the information by applying their language

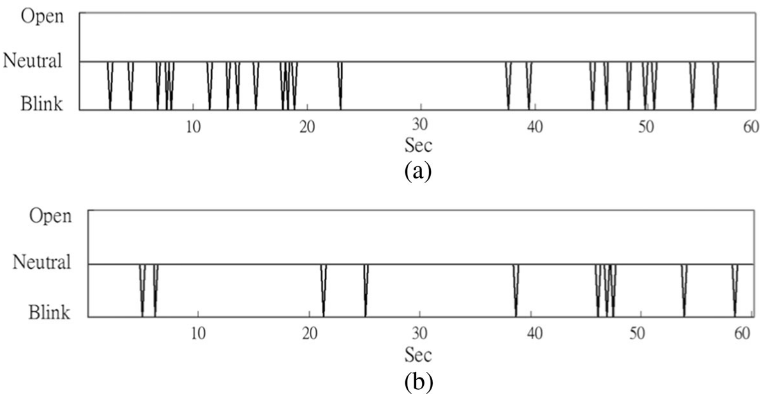


Fig. 15 Analysis chart of S2 facial expression to textbook 2

Table 7 Learning effectiveness of textbook 2 and answers in learning sheet

Item	S1	S2
Do you consider the hero and heroine talked too quickly?	Yes (3)	No (1)
Are you interested in contents in the textbook 2?	Fair (2)	Yes (1)
Can you fully understand scenario content of the textbook 2?	Yes (1)	Yes (1)
Can you fully understand talk content in the textbook 2?	Fair (2)	Yes (1)
Does the heroine sit on the left side of the hero?	Yes(O)	Yes(O)
Is the dining table rounded?	Yes(O)	Yes(O)
Did the hero forget the car key when he returned to the house?	No(O)	No(O)
Did the hero return the house to make an important call?	Yes(X)	No(O)
Has the hero found the wallet?	Yes(X)	No(O)
Did the heroine guess the car key is locked in the car?	No(X)	Yes(O)

coding ability. Textbook 3 was unrelated to any standard course offered at vocational colleges; specifically, the content was about data structure (push and pop in the stacking unit). For the students, the content of Textbook 3 was unusual and difficult to understand; consequently, the information-coding process was arduous.

The experimental results showed that S1 exhibited 308 negative facial expressions (Table 8 and Fig. 16) while learning from Textbook 3; negative expressions thus increased to 32.4 % of all facial expressions. Furthermore, the percentage of correct answers was only 16.7 % (Table 8). Learning effectiveness was non-existent, and the frequency of negative facial expressions corresponded with the increase cognitive load. S2 exhibited 26 negative facial expressions (Table 8 and Fig. 17), increasing to 2.7 % of all facial expressions, indicating that Textbook 3 caused S2 to reach a critical cognitive load. Furthermore, the percentage of correct answers decreased to 83.3 % (Table 9). Teachers should pay particular attention to this result; specifically, teachers should avoid using unusual approaches to presenting the textbook content. Teachers who list terms and charts on the blackboard but fail to convey the language n meaningful terms cause cognitive overload of students; thus, this approach is an ineffective approach to teaching.

S1 and S2 blinked 26.6 and 18.8 times per minute, respectively; moreover, Textbook 3 resulted in the highest rate of blinking among the three textbooks in this study. This showed that textbook difficulty was positively correlated with blinking frequency. In addition, difficult textbooks impose high cognitive load, disrupt concentration, and cause students to blink frequently (Figs. 16b and 17b). The subjective measurement results showed that S1 partly understood the visual information, whereas the listening information was not understood. Only one answer was correct in the visual perception test, and all of the answers were wrong in the

Table 8 Detection result of facial expression to textbook 3

	Image length (sec)	Cognitive load	Positive		Neutral	Negative				Blink
			Happiness	Surprise	None	Anger	Fear	Sadness	Disgust	
S1	97	11	0	0	642	188	0	0	120	43
S2	99	9	0	0	952	0	0	18	8	31

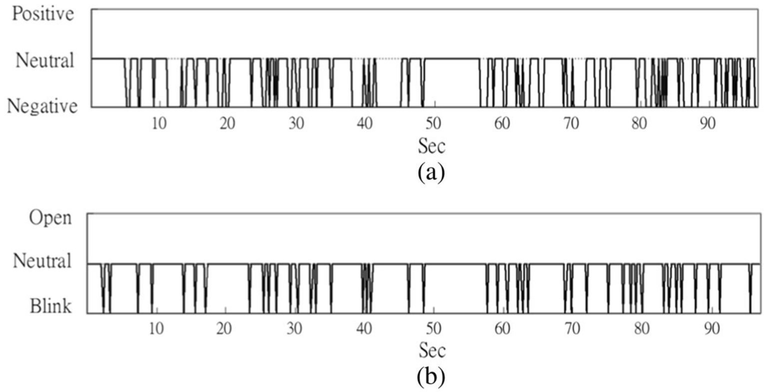


Fig. 16 Analysis chart for S1 facial expression to textbook 3

auditory perception test. The subjective measurement result was consistent with the postlearning test results (Table 9). The subjective measurement results showed that S2 partially understood the visual and understand listening information. However, all of the visual perception test answers were correct, and only one answer was wrong on the auditory perception test. The subjective measurement underestimated the outcome of the visual information. The listening aspect was primary cause of cognitive load reaching a critical level. The subjective measurement for listening aspect was consistent with the postlearning test result.

5.3 Experimental results and analysis

Information processing theory and limited capacity of cognitive load theory purport that high cognitive load results in information processing failure. Students experience worry and fear, and negative emotions such as frustration, which manifests as negative facial expressions. Conversely, people can effectively process information provided that their cognitive load is not exceeded. In a normal teaching environment, the mental state of students is stable, and facial expressions are natural. Thus, neutral facial expressions are rational and expected. Furthermore, learning effectiveness is achieved when students can concentrate on their learning tasks. The results of this study indicate that high anxiety from learning a foreign language is negatively correlated with the students' reading performance. Conversely, the performance

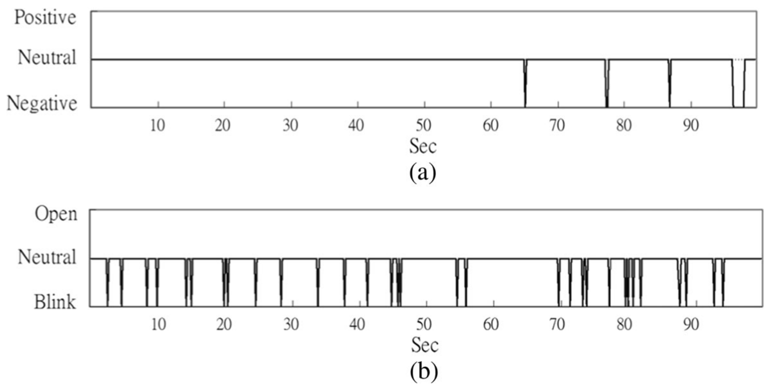


Fig. 17 Analysis chart for S2 facial expression to textbook 3

Table 9 Learning effectiveness of textbook 3 and answers in learning sheet

Item	S1	S2
Do you think teaching speed of the textbook 3 is too quick?	Yes (3)	Yes (3)
Are you interested in the textbook 3 contents?	No (3)	Fair (2)
Do you fully understand text contents of textbook 3?	Fair (2)	Fair (2)
Do you understand teaching content of the textbook 3?	No (3)	Fair (2)
In empty stack, size () can be used for detection?	Yes(X)	No(O)
In operation example, is number 5 firstly placed in the stack?	Yes(O)	Yes(O)
In operation example, is number 6 firstly placed in the stack?	No(X)	No(O)
Is adding data to the stack called pop?	No(X)	No(O)
Is stack top mark called push?	Yes(X)	Yes(O)
In empty stack, is the returned value from detection equal to 1?	Yes(X)	Yes(X)

of students with low anxiety is markedly higher than that of students with high anxiety. Students with low anxiety exhibited superior English reading performance in comparison with highly anxious students. Consequently, anxiety can affect the performance of students when tested. During learning, cognitive load is high if the students exert an excessive level of mental effort. Mental pressure and learning anxiety are the primary cause of study emotions. The experimental results showed that facial expressions while learning are positively correlated with cognitive load. Thus, students with positive expressions exhibit superior learning effectiveness in comparison with those with negative facial expressions. When designing teaching activities and course content, teachers should consider the cognitive load of the students to maintain effective learning practices.

In this study, we employed subjective measurements to assess students' mental efforts. The two students considered Textbooks 1 as having the simplest content, and Textbook 3 as having the most difficult content. The results were consistent with the experimental design. The ERAS observed the learning state of the students by detecting physiological reactions such as facial expressions and eye movement. The experimental results showed that learning difficulty corresponded with cognitive load. Negative emotions instinctively manifested as negative facial expressions. Furthermore, distraction increased blinking frequency. S2 responded with positive facial expressions toward the operator robots in Textbook 1 and the hero's behavior in Textbook 2. In addition, S2 presented neutral facial expressions, which could be interpreted as normal facial expressions, toward learning. The students may have shown positive facial expressions toward content that they are interested in; however, the duration of these were short. The positive facial expressions ceased when the textbook presented new information, the students were guided into another coding phase, or their level of interest decreased. When the total cognitive load of the students exceeded their load capacity, their facial expressions became negative as a result of the pressure and anxiety. The duration of negative expressions tended to be longer than that of positive facial expressions, and it endured until the students completed an information coding task, the cognitive load capacity receded to a tolerable level, or when the negative emotions disappeared. Thus, teachers should allow sufficient time for students to rest, such as sharing an anecdote, telling a joke, dismissing the class, and allowing more time to code information, thereby reducing their cognitive load. Table 10 shows the summay of questionnaire for teaching experiment.

Table 10 Summary of experiment results

Questions	S1	S2
Are you nervous in the experiment?	No	No
Did you keep natural facial expressions as usual in the experiment?	Yes	Yes
Did you concentrate your attentions on learning in the experiment?	Yes	Yes
How do you prioritize the three textbooks according to levels of difficulty?	Textbook 1 Textbook 2 Textbook 3	Textbook 1 Textbook 2 Textbook 3
Which textbook you can adapt to best among the three textbooks?	Textbook 1	Textbook 1
Which textbook you cannot adapt to most among the three textbooks?	Textbook 3	Textbook 3
How do you feel when you can understand the textbook contents in the experiment?	Relaxed and pleased	Relaxed and pleased
How do you feel when you cannot understand the textbook contents in the experiment?	Disgust and reject Seriously gaze	Sad and helpless Fear pressure

In this paper, a dual-channel-interaction cognitive theory of multimedia learning was applied in experimental learning. Facial expressions were elicited by increasing the students' cognitive load, namely, by employing various language-related visual and auditory abilities to deliver course content. ERAS effectively detected and recognized students' facial expressions, and assisted in assessing the learning effectiveness of students, and the difficulty of textbooks provided by the teachers.

6 Discussion and recommendations

The experimental results showed that the analysis of learning effectiveness through facial expressions has the following characteristics.

Facial expressions are a projection of inner emotions. When a textbook arouses a state of interest, students' learning emotions are projected in their facial expressions. Generally, information can be interpreted based on a person's previous experience, which is then converted into inner perception. Positive experiences or emotions (e.g., joy, surprise, happiness, pleasure, expected, affirmation, trust, humor, and curiosity) can elicit positive facial expressions. Negative emotions (e.g., anger, fear, sadness, disgust, boredom, dejection, rejection, pressure, anxiety, pessimism, nervousness, and doubt) can manifest as negative facial expressions. S2 was interested in the operator robots in Textbook 1, and she smiled; however, S1 did not react. Conversely, S1 was displeased with the increased teaching speed in the final part of that lesson, and the ERAS detected that S1 exhibited a negative facial expression—disgust; however, S2 had no facial expression (Fig. 18). In Textbook 2, similar situations were observed; S2 was interested in the hero's behavior, which elicited a positive reaction. S1 could not understand the scenario, and attempted to guess the scenario content; consequently, her degree of interest toward the hero's behavior decreased, whereas S2 felt relaxed because she understood the dialogue. The antics played a role (Fig. 19).

Students exhibit positive facial expressions when they are interested in novel textbook content. Conversely, negative information in textbooks elicited negative facial expressions. When people are in a natural state and are unaffected by emotions, their facial expressions

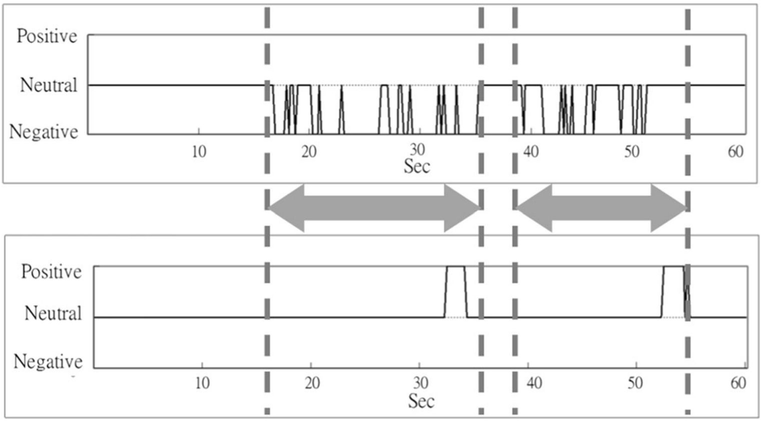


Fig. 18 Analysis chart of facial expressions to textbook 1

remain unchanged. Unconscious habitual actions are meaningless, and these expressions were classified as neutral facial expressions in this study. The above three facial expressions have the same value in evaluation learning effectiveness; positive expressions do not imply the optimal learning effectiveness, just as negative expressions do not imply the worst learning effectiveness. Teachers should present facial expressions of the students toward textbook content, and verify whether they are consistent with those in actual learning situations. If the teacher presents a positive expression and students exhibit a negative expression (or vice versa), this is unsatisfactory. For Textbook 1, 95.5 % and 97.3 % of the reactions elicited from S1 and S2 were neutral facial expressions, and 100 % of the answers were correct (Table 11), indicating that it is normal for students to have neutral facial expressions while learning. The cognitive load was within the cognitive load capacity, their working memory was not overloaded, and no negative emotions occurred; thus, their facial expressions remained unchanged. The learning was effective when the students concentrated on the learning tasks and were engaged in mental activities.

Attention was paid to the time in observing facial expressions. The students exhibited positive facial expressions when they were interested in the textbook content; however, this was sustained for only a short period, after which their positive expressions returned to neutral expressions (Fig. 19) when new information was presented in their textbooks, when the

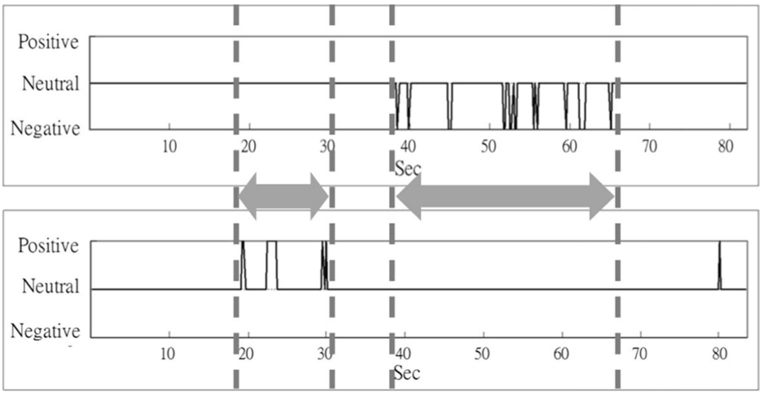


Fig. 19 Analysis chart of facial expressions to textbook 2

Table 11 Summary for percent of social interaction and correct rate

Textbook(student) \social interaction	Positive	Neutral	Negative	Correct rate
Textbook 1(S1)	0 %	95.5 %	4.5 %	100 %
Textbook 2(S1)	0 %	69 %	31 %	50 %
Textbook 3(S1)	0 %	67.6 %	32.4 %	16.7 %
Textbook 1(S2)	2.7 %	97.3 %	0 %	100 %
Textbook 2(S2)	6.9 %	93.1 %	0 %	100 %
Textbook 3(S2)	0 %	97.3 %	2.7 %	83.3 %

teacher guided introduced a new coding phase, or when their interest subsided. If negative emotions caused by cognitive overload induce negative facial expressions, neutral expressions can be restored after the negative facial expressions or cognitive load are decreased to the within the load range; thus, the duration is longer (in Fig. 19). The time when the facial expression appears is also critical; appropriate facial expression shall occur at the appropriate time.

Only facial expressions that occur for a certain period of time are meaningful. Previous research [15] indicated that the sufficient time for accurately recognizing microexpressions is approximately 200 ms. Isolated or accidental facial expressions, or those that last for less than 200 ms can be regarded as noise. Figure 20 shows a diagram of the facial expressions of S1 and S2 in response to Textbook 3 (Fig. 20); S1 maintained a negative facial expression throughout the learning activity, which indicates that S1 performance experienced difficulty when learning from the textbook. Moreover, S2 did not exhibit any negative expressions until the final experiment, indicating that S2 also experienced difficulty with this textbook. Therefore, in similar situations, the teacher should rearrange the lesson content or provide additional teaching activities. Future studies should focus on determining the clustering of particular facial expressions. Quantitative data that is more objective can be obtained from larger samples.

In group or remote teaching environments, teachers cannot observe and record the learning state of each student; thus, the textbook differentiation cannot be evaluated. The system can assist teachers in evaluating students' adaptability to textbooks that contain difficult content. If the students adapt well, the learning difficulty is low; therefore, students can progress to the

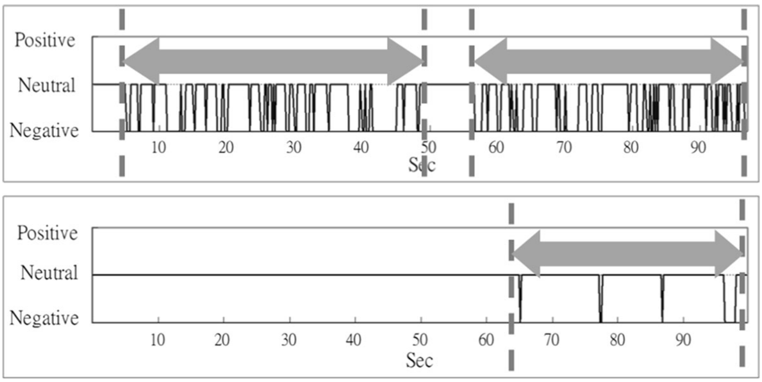


Fig. 20 Analysis chart of facial expressions to textbook 3

next unit or examine other textbooks to gain an in-depth understanding of relevant content. Conversely, if students cannot adapt well, the learning difficulty is high; accordingly, the teacher can provide additional instruction to assist the students in relearning the textbook, or a basic textbook can be employed as supplementary material.

The experimental results showed that facial expressions can change based on the textbook difficulty. When the textbook difficulty increased, the students' negative facial expressions became increasingly obvious as the cognitive load increased. The proposed ERAS analyzed the facial features of students and evaluated the level of difficulty experienced by the students while learning from the textbooks. The negative expressions of S1 become more apparent as the difficulty of the textbook increased. The percentage of the negative facial expressions increased from 4.5 % (Textbook 1) to 31 % (Textbook 2), and then 32.4 % (Textbook 3). Conversely, S2 produced no negative facial expressions toward Textbooks 1 or 2, although 2.7 % of her facial expressions were negative while learning from Textbook 3, indicating that Textbook 3 was the most difficult. In similar situations, teachers can introduce the content before the class, request the students to prepare the content, and progress slowly through the textbook. Teachers can also capitalize on the repeatability of the multimedia textbooks, and request students to review the lesson material after class. In addition, this can be used as reference for learning appraisal.

7 Conclusion

Teaching involves teachers who apply teaching methods to transfer information stimuli based on textbook content to the minds of their students via information sensory channels. These information stimuli are organized and integrated into the brain. Understanding this information is called coding ability, which is related to the cognition of information stimuli. The brain can react appropriately only when the information is coded successfully. Increasingly difficult textbooks were employed in the experiments. Information stimuli coding ability affects the inner emotions, and is indirectly manifested through facial expressions. We recognized these facial expressions and classified them as neutral, positive, and negative reactions.

The FACS, OpenCV, and image processing method were employed for face detection, feature extraction, and feature position. In this study, a facial expression analysis system was implemented. The proposed ERAS extracted 10 facial feature points, and 9 feature vectors and neutral values were identified as measurement benchmarks for detecting when a student blinked based on 11 facial AUs. The ERAS classified the AUs into six basic facial expressions (happiness, surprise, fear, anger, sadness, and disgust). Subsequently, these AUs were categorized into three AU combinations (positive, neutral, and negative) of social interaction.

In the experiments, the students were not required to perform specific facial expressions for the detection process. Neutral facial expressions are normal in a normal learning state. If learning effectiveness requires assessment, other physiological signals such as eye movement should be considered to enhance the reliability of the system identification. The experimental results showed that by using an identical textbook, S2 adapted to the textbook content with less difficulty, and also exhibited more neutral expressions and positive expressions; conversely, S1 exhibited more negative facial expressions. To obtain more detailed results, we increased the difficulty of the textbook and observed that the textbook difficulty and ratio of positive facial expressions have an inverse relationship; in other words, the more difficult a textbook is, the more frequently negative expressions occur. Thus, students can manage cognitive load before they show many negative expressions. Moreover, teachers can improve their awareness of whether their students understand the textbook content, as well as their adaptability to new information.

References

1. Bolme DS, Strout M, Beveridge JR (2007) FacePerf: benchmarks for face recognition. Proc Int Symp Workload Character: 114–119
2. Bousmalis K, Mehu M, Pantic M (2013) Towards the automatic detection of spontaneous agreement and disagreement based on non-verbal behaviour: a survey of related cues, databases, and tools. Image Vis Comput 31(2):203–221
3. Brunken R, Plass J, Leutner D (2003) Direct measurement of cognitive load in multimedia learning. Educ Psychol 38(1):53–61
4. Ekman P (2003) Emotions revealed: understanding faces and feelings. Weidenfeld & Nicolson
5. Ekman P, Friesen WV (1978) The facial action coding system: a technique for the measurement of facial movement. Consulting Psychologists Press, San Francisco
6. Gerjets P, Scheiter K (2003) Goal configurations and processing strategies as moderators between instructional design and cognitive load: evidence from hypertext-based instruction. Educ Psychol 38(1):33–41
7. Ghanem K, Caplier A An efficient method to classify positive and negative expressions. Int J Tomography Stat 17(11)72–78
8. Lin K-C, Lin R-W, Chen S-J, You C-R, Chai J-L The classroom response system based on affective computing. Third IEEE Int Conf Ubi-Med Comput, Zhejiang Normal University, Jinhua, China
9. Lin K-C, Chen S-J (2012) Affective computing in distance learning: based on facial emotion recognition. 2nd Int Conf Future Comput Educ, Shanghai, China
10. Loh MP, Wong YP, Wong CO (2006) Facial expression recognition for E-learning systems using gabor wavelet & neural network. Proc Int Conf Adv Learn Technol
11. Mayer RE (2005) The Cambridge handbook of multimedia learning. Cambridge University Press, NY
12. Nosu K, Kurokawa T (2006) Facial tracking for an emotion-diagnosis robot to support E-learning. Proc Int Conf Mach Learn Cybernet: 3811–3816
13. Picard RW (1995) Affective computing. Technical Report TR-321, MIT, Media Laboratory
14. Shen L, Wang M, Shen R (2009) Affective e-learning: using “emotional” data to improve learning in pervasive learning environment. Educ Technol Soc 12(2):176–189
15. Shen X, Wu Q, Fu X (2012) Effects of duration of expressions on the rercognition of microexpressions. J Zhejiang Univ-Sci B: Biomed Biotechnol 13(3):221–230
16. Sweller J, van Merrierboer J, Paas F (1998) Cognitive architecture and instructional design. Educ Psychol Rev 10(3):251–296
17. Vinciarelli A, Pantic M, Heylen D, Pelachaud C, Poggi I, D’ericco F, Schroeder M (2012) Bridging the gap between social animal and unsocial machine: a survey of social signal processing. IEEE Trans Affect Comput 3(1):69–87
18. Whitehill J, Omlin CW (2006) Haar features for FACS AU recognition. Proc Int Conf Autom Face Gesture Recognit: 5–9
19. Wu Y, Wang TT, Chu XN (2009) Affective modeling and recognition of learning emotion: application to E-learning. J Software 4(8):859–866
20. Zhang Y, Ji Q (2005) Active and dynamic information fusion for facial expression understanding from image sequences. IEEE Trans Pattern Anal Mach Intell 27(5):699–714

Jason Chi-Shun Hung received a BS and MS degree in computer science and information engineering from Tamkang University in 1996 and 1998 respectvely. He also received his PhD in computer science and information engineering from the Tamkang University in 2001. He is currently an associate professor at the Department of information technology, Overseas Chinese University, Taiwan. His current research interests include multimedia computing and networking, digital learning and data mining.

Kun-Hsiang Chiang received his MS in Department of Management Information Systems from the National Chung Hsing University, in 2012. His primary interests lie in the areas of affective computing and digital learning.

Yi-Hung Huang received the BS, MS, and Ph.D. degrees from the Department of Applied Mathematics, National Chung-Hsing University, Taiwan, in 1986, 1990, and 1999, respectively. He is currently an associate professor at the Department of Mathematics Education, National Taichung University of Education, Taiwan. His research interests include cognitive load theory, Instructional design, machine learning and mobile communications.

Kuan-Cheng Lin received a BS in chemistry from National Taiwan University in 1988 and a PhD in applied mathematics from the National Chung-Hsing University in 2000. He is currently a professor with the department of management information systems at National Chung-Hsing University, Taichung, Taiwan. His current research interests include affective computing, intelligent tutoring system and data mining.