Augmenting Social Interaction in Digital Learning through Affective Computing

Kuan-Cheng Lin, *Member, IEEE*, Kun-Hsiang Chiang, and Jason C. Hung

Abstract—Social interaction between teachers and students can be effectively enhanced if teachers can extract spontaneous non-verbal behavior of students in real time, such as facial expressions and body language. In view of this, teaching strategies of teachers and learning effectiveness of students can be effectively improved. This study implemented a facial expression analysis system. The system can capture facial images of students via webcam and analyze facial expressions of students towards learning contents with different difficulties and a relationship between social interaction and learning effectiveness. The system selected ten face feature points for identification of 11 facial action units. Then these action units are classified into positive, neutral and negative social interactions by using a rule-based expert system. After verification using cognitive load theory, the teaching experiment results showed in multimedia digital learning the student with high coding ability can adapt to the learning content, and displayed more neutral and positive social interactions, while the student with low coding ability displayed more negative social interaction due to increase of cognitive load. At the same time, real-time detection of social interaction can act as basis for diagnosis of student learning difficulty and help teachers adjust teaching strategies.

Index Terms—Affective computing, cognitive load, face expression, digital learning

1 INTRODUCTION

IN daily communication, we often analyze inner affective state of other persons by observing their spontaneous or unconscious non-verbal behaviors such as facial expression or body language. This kind of communication and interaction is called social interaction. For example, we can judge whether some person agrees or disagrees with a thing through his/her facial expression when they hearing the thing [1,11]. At the same time, social interaction may occur in face-to-face teaching in school classrooms, and teachers may observe students non-verbal behaviors to judge whether students understand curriculum contents.

Teachers often like to ask “any question?” for feedback in class in order to maintain teaching effectiveness. They intend to know students learning effectiveness in unit teaching, and then decide to reinforce current contents or teach new contents. However, students often have no response or look around silently. Teachers often feel frustrated. In digital teaching, teachers can know whether students have understood the teaching contents through affective computing [2], social signal processing [3] and information technologies.

Teaching is a kind of information transmission. Teachers transmit teaching contents to student sensory organs via information channels, forming information stimulus. The information stimulus can stimulate student minds to process the information, and produce recognition memory. If the information stimulus accords with positive experience, the students may generate positive emotions, otherwise they will generate negative emotion. Emotions are difficult to perceive but facial expression can be detected. Picard [2] elaborated affective computing which has the ability to detect people's emotions. It can be further applied in CAL, information retrieval and computer-human interaction. With improvement of computer computational efficiency, there are many mature researches on facial expressions recognition. These researches captured human faces through web cams, classified the facial expressions through various computerized algorithms, and further evaluated emotional reactions. Digital learning assisted with affective computing-based human-machine interface provides effective and objective feedback for teaching activities and can serve as reference for adjustment of learning strategies and teaching skills. Facial expression recognition system can perform learning status analysis, learning difficulty diagnosis, learning mental state detection, emotional recognition, and attention analysis according to facial expression features [4-10].

Students have different degree of perception to multi-media textbook contents because every one has different coding abilities for information stimulus. Coding ability is the start of human cognizing information. More efforts are spent when coding tasks are more and complicated. The learning effectiveness is also affected. Learning is a mental activity. It can be determined whether students produce learning through overt behaviors. One of the clues is to observe whether eyes and faces of students show pleased and satisfied ex-

- *Kuan-Cheng Lin is with the Department of Management Information Systems, National Chung Hsing University, 250 Kuo Kuang Road, 402 Taichung, Taiwan, R.O.C., E-mail: kclin@ nchu.edu.tw.*
- *Kun-Hsiang Chiang is with the Department of Management Information Systems, National Chung Hsing University, 250 Kuo Kuang Road, 402 Taichung, Taiwan, R.O.C., E-mail: msger99@hotmail.com.*
- *Jason C. Hung is with the Department of Information Management, Oversea Chinese University, 402 Taichung, Taiwan, R.O.C., E-mail: jhungc.hung@gmail.com.*

pressions.

The past affective computing used simple webcams as image capture devices. Due to limited resolution, image quality and evenness of light, in order to capture apparent facial expressions, some expressions were inevitably exaggerated to fulfill detection and recognition. In this paper, we captured real facial expressions of students in multimedia learning environment. In order to make facial expressions in learning process have more significant difference, multimedia English textbooks to which Taiwanese students are not unaccustomed were selected as information source, and teaching contents with different difficulties were used to give information stimulus to students. The social reactions resulted from coding results and cognitive load were displayed on the facial expression, which were detected through facial expression recognition system. The following issues are mainly discussed.

1) How does the system detect and recognize facial expression and social reaction in natural learning state?

2) What are social reactions of students with different English listening and visual abilities when learning multimedia textbooks with different difficulties?

3) Discuss a relationship between facial expressions and social reactions in learning.

2 Background and Related Work

2.1 Cognitive Theory of Multimedia Learning

Multimedia learning means students learn from both words and pictures, so multimedia learning is also called dual-code learning or dual-channel learning. Words contain written words input for eyes and sound words input for ears, and belong to word type. Pictures contain static pictures (illustrations, coordinate graphs, diagrams, photos and maps) and dynamic pictures (animation and video), and belong to picture type; both are information processing system for humans (as shown in Figure 1). Multimedia information forms sound and image memories in minds through input by sensory organs. The sound memory is processed through word model, and the image memory is processed through picture model. If the students want to produce learning, they must integrate the results from processing of word model and picture model with the previous knowledge (old experience) in long-term memory area, establish comprehensible and meaningful knowledge in the working memory area, and store the knowledge as new schema in long-term memory area [12].

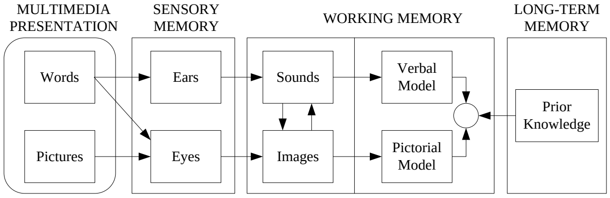


Fig.1 Cognitive Theory of Multimedia Learning

Decoding is the restoration process of output symbols of the signal source. This process changes with individual experience, and is crucial to producing facial expression. The facial expressions can occur when the addressee can interpret the information based on his experience and convert it into his inner feeling. In this paper, multimedia textbooks are used to give visual and auditory stimuli to learners. In theory, working memory can control sensory organs to make reaction after integration. These reactions include facial expression, sound, body actions, physiological responses and etc. This paper focuses on facial expressions. The English textbooks which Taiwanese students are not familiar with were used to produce difference in seeing and hearing abilities of foreign language. Facial expression of students can be observed in learning textbooks with different stimulus. The observed facial expressions were sent back to teachers, providing reference for adjusting and evaluating student learning effect and reactions to the textbooks.

2.2 Cognitive Load Theory

Sweller, van Merrierboer, and Paas [13] considered cognitive load means mental load which occurs when specific task is added to learner's cognitive system. The leaning effectiveness will be affected when the load exceeds the range which an individual can bear. Gerjets and Scheiter [14] also extended this theory, and suggested framework of cognitive load theory for teaching design, textbook contents and textbook complexity.

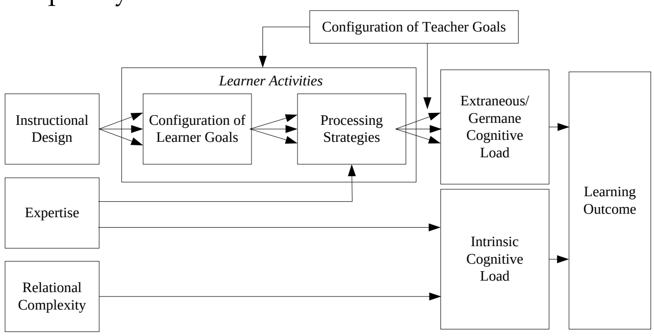


Fig.2 Theoretical framework of cognitive load

The research also indicated cognitive load is excessive if information received by a learner exceeds his cognitive ability. As a result, the individual cannot recognize information meanings, and information process failed. If the total cognitive load by the information is within the individual load capacity, the individual can continue to process the information effectively. In information processing theory, cognitive load means learning load in working memory area. The working memory has limited storage capacity. Overloaded working may cause excessive cognitive load (high cognitive load). In this case, the learners cannot concentrate their attention, and this may have negative impact on learning emotions. Hence, learning effectiveness is lower [13].

Brunken et al. [15] suggested two methods of measuring the cognitive load (mental load) spent by individuals in learning, as follows:

a. Subjective measurement: This method is based on the subjects who have an ability of recalling cognition in learning and can clearly indicate consumption of mental efforts. If questionnaire survey is conducted, the subjects are required to indicate metal efforts in learning, textbook difficulty and individual ability level.
b. Objective measurement: This method assumes cognitive load change can be reflected in physiological changes, including heartbeat, brain activities, eye movement, neuroimaging and behaviors.

The proposed ERAS is an application of the objective measurement. However, in order to demonstrate the system feasibility, the cognitive load is verified by comparing with the cognitive load by subjective measurement.

2.3 Facial Expression Recognition and Facial Action Coding System

Affective computing belongs to artificial intelligence. It uses massive computing, deduction and learning to enhance perception. The first important one is affective perception. Affective perception signals can be from objective physiological changes, such as facial expression, sound, pulse, breathe, brain waves, eyeball movement, body movement and etc. Emotion is a brain's appraisal for sensory stimuli, and a reaction resulted from emotional experiences. The reactions shown on faces are facial expressions, such as happiness, anger, sadness, fear, surprise and disgust; reactions shown on visceral nerve may change pluse, breathe, blood pressure, pupils, body temperature, brain waves, muscle and skin electric conductivity. Facial expression is a natural way to show inner emotional reactions. The emotional expressions are mainly shown on mouths, eyes and eyebrows. Thus, the system must recognize and analyse the three parts.

Ekman suggested facial expression combinations of Facial Action Coding System (FACS) [16-17]. It has one or combines several relevant facial action units, which can enhance facial expressions recognition. Ekman and Friesen defined 44 facial action units (AU) from upper face and lower face, and six basic emotions as per the facial expressions: Happiness, Anger, Sadness, Fear, Surprise, and Disgust et al. Due to complicated relationship between human facial expressions, the action units may be combined with other action units. Thus, description of facial expressions and degree may be different. Although FACS defined 44 facial action units, the operator can select action units related to the operation goal in practice according to needs, and optimal recognition effect can still be achieved.

3 The Design of Expression Response Analysis System(ERAS)

Implementation of ERAS contains three steps: facial detection, feature extraction and location, and expression classification and recognition.

3.1 Facial Detection

We used OpenCV AdaBoost (machine learning algorithm), and Haar-Like for facial detection [18,19]. OpenCV trained facial detection cascade classifier, including face, eyes (including left eye, right eye and eyes), nose and mouth to finish detection of face or specific locations. One static facial image contains eyes, nose and mouth. The total average detection time is around 0.6 sec, and detection frequency is extremely high. The distance between eyes and brows are too close. In order to avoid mutual interference in the two extraction areas, the image sequence is duplicated to extract four areas: brows, mouth, left eye and right eye. The flow chart for facial detection and feature extraction is shown in fig.3.

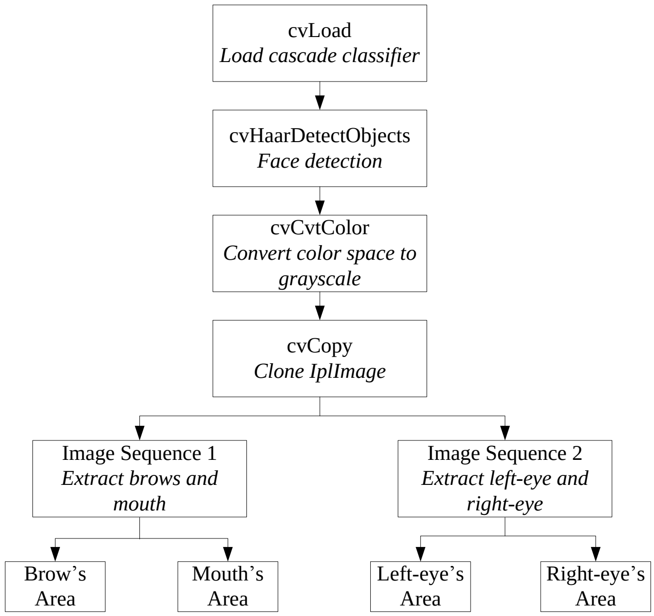


Fig. 3 Flow chart for face detection and feature extraction

3.2 Feature Extraction and Location

We observe changes in facial expressions mainly through three parts, brows, eyes and mouth. The feature points performed image processing and displacement measurement for the three areas. Thus, 10 feature points were selected as measurement basis (as shown in Figure 4). Point A and Point B represent left and right corners of mouth respectively. When corners of mouth turn upwards, point A and point B raise. The upper and lower lips are linked, CC′ distance can represent mouth opening and closing. Eyelids are linked. EE′ and FF′ distance can represent eye opening and closing. Even, eye blink or closing can be determined through closing time. Inside corners of eyebrows are also important in facial expression. This can be observed through up and down of Point D and Point D′ and distance between the two points.

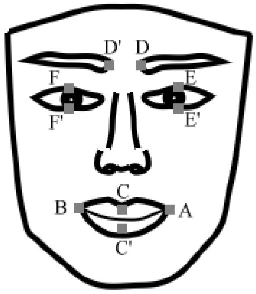


Fig. 4 Location of face feature points

3.3 Feature Distance Vector

Feature vector is important basis for classification of facial expressions. The data changes have meanings. AB feature point is taken as an example. When smile is shown on face, corners of mouth turn upwards, vtAB is expanded and enlarged, and vtA and vtB are also enlarged. This is AU12. Swallow may also enlarge vtAB. However, corners of mouth do not turn upwards. Thus, vtA and vtB are not enlarged. This is AU20. With smile, the upper and lower lips become thin, and vtCC′ are shortened. This is AU12+24. When laugh is shown in face, the upper and lower lips are opened, and vtCC' is enlarged. This is AU12+25,26,27 according to vtCC' stretched length. When yawning or feeling sad, opening upper and lower lips enlarges vtCC′. However, corners of mouth are downwards. Thus, this is AU15+25,26,27. The original data of vtAB feature vector are shown in Fig. 5.

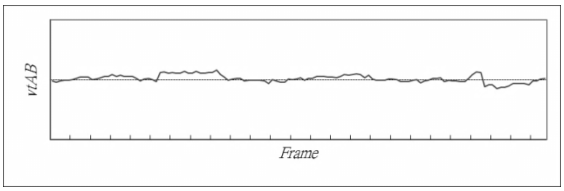


Fig. 5 Diagram of vt image sequence (before conversion)

The real-time image detection and human breathe may cause fluctuation and jiggle. The feature vector may change slightly. This is not good for subsequent classification and identification. Thus, the experimental subjects must keep no facial expression for 2-3 seconds. This can facilitate the system to detect neutral (neu). The neutral value is used as benchmark, -10% is lower threshold value, and +10% is upper threshold value. They are converted through Equation 1.

$$cnv = \begin{cases} neu*0.7 \text{ , } if \;\; vt \leq neu*0.9 \\ neu*1.3 \text{ , } if \;\; vt \geq neu*1.1 \\ neu \quad\quad \text{ , } other \end{cases}$$

After threshold value conversion (cnv), vt vector is shown in Fig. 6. Fig. 7 and 8 represent feature vector change of mouth and upper and lower eyelids.

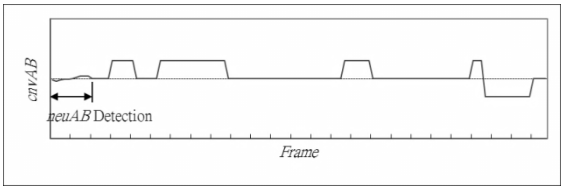


Fig. 6 Diagram of cnv image sequence (after conversion)

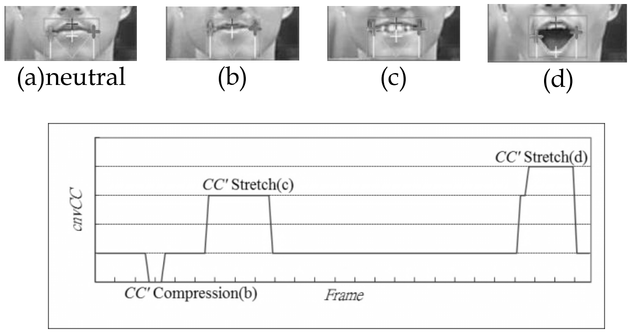


Fig. 7 Diagram of change in mouth feature vector

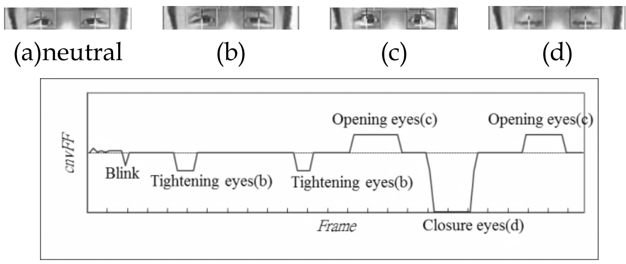


Fig. 8 Diagram of change in feature vector of upper and lower eyelids

3.4 Expression Recognition and Classification

This paper aims at social reactions represented by natural facial expressions of students after they received textbooks with different stimuli. Rule-Based Expert System was used for identification of feature vectors of each location, and FACS AU for recognition.

Brow's AUs (AU1, AU4)

Brow's AUS are used to identify up and down of points D and D′ and DD′ flexing distance. DD' distance may not change when we consider up and down of points D and D'. Thus, the identification rule is divided into two parts for classification.

Eye's AUs (AU5, AU7, Blink, Close)

Eye action is very rapid, and action maintenance time must be considered. If eyelid closure distance (EE′ and FF′) is longer, the maintenance time is shorter; if the eyelid closure distance is shorter, the maintenance time is longer. Both are exclusive. In the same frame, when AU5 (opening eyes) is determined, AU7 (eyelids tightening, blink and closure) is impossible to occur. The webcam shooting speed is 10FPS, and the maintenance time is 0.5 second (5 frames). When the eyelid closure distance is ≤0.01 and the maintenance time of action is less than 0.5, the facial expression is blink. If the maintenance time exceeds 0.5, the facial expression is blink.

Mouth's AUs (AU12, AU15, AU20)

Action of corners of mouth is crucial to producing mouth expression. The basis is up and down of points A and B, and AB distance. When people feel pleased, corners of mouth may turn upwards, showing a smile; when people feel unpleased, there is no facial expression, or corners of mouth are downwards; when people feel bored or uninterested, corners of mouth may stretch, showing escape.

Lips's AUs (AU24, AU25, AU26, AU27)

Opening and closing of upper and lower lips can lift reaction range of mouth expression, and the main basis is CC′ distance. We can take neuCC′ as an example. If less than neuCC′, the upper and lower lips are tense, and CC′ distance is shortened; if exceeding neuCC′, the mouth is opened, and CC′ distance is enlarged. There is small open-wide AU25, medium open-wide AU26 and big open-wide AU27.

AU Combinations

Ekman and Friesen defined 44 action units from upper face and lower face, and defined six basic emotions: Happiness, Anger, Sadness, Fear, Surprise and Disgust. This paper summarizes 11 action units according to the facial expressions of the students after they received information stimuli. The AU combinations are defined according to the basic six emotions, as shown in Table 1.

Table 1 AU combinations of basic emotions

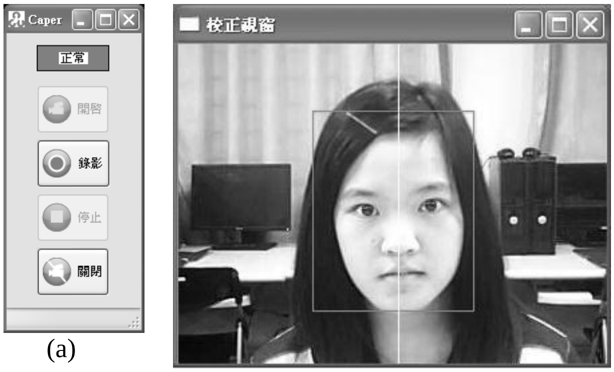
Location	Emotions	Positive		Negative			
		Happiness	Surprise	Fear	Anger	Sadness	Disgust
Upper Face	AU1	●	●	●		●	
	AU4			●	●	●	●
	AU 5		●	●	●		
	AU 7	●				●	●
Lower face	AU 12	●					
	AU 15		●	●	●	●	●
	AU 20			●			●
	AU 24	●		●	●	●	●
	AU 25	●	●	●	●	●	●
	AU 26	●	●	●	●	●	
	AU 27	●	●	●	●	●	

4 The Implementation of Expression Response Analysis System(ERAS)

ERAS contains three modules: image extraction module, facial expression recognition module, and emotional reaction analysis module.

4.1 Image extraction module

Image extraction module is simple in interface operation, and has image location adjustment and warning functions. It runs on the computers of the tested students. Due to small interface and simple functions, it can reduce psychological impact on tested students, and make images and locations of tested students consistent. Its main outcome is AVI file of the facial expressions of tested students in learning. These files are convenient for future processing and application. When the green light is on (as shown in Figure 9a), face location of one tested student is suitable, and distance to the lens is normal (60~70cm); when the yellow light is on, face location of the tested student deviates, and is still accepted; when the red light is on, face of the tested student cannot be detected, or the system cannot perform processing and recognition. The tested students shall avoid red light. During program execution, the correction window (as shown in Figure 9b) is in central upper edge of the screen, and below the webcam. Its function is like a mirror, and tested students can exercise facial expressions besides adjustment of face location. After pressing down "Record", the correction window is hidden to prevent interference with testing.



(a) (b)

Fig. 9 Operation picture of image extraction module

4.2 Facial expression recognition module

The facial expression recognition module runs at back end. It mainly process facial expression images of the tested students, including image processing, feature point marking, calculation of feature vector and conversion. At last, all the feature vectors are recognized to action units, and then text files are input. The detail processing method is illustrated in section 3.4. This module can be used for detection of dynamic images, as shown in Figure 10.

Fig. 10 Facial expression recognition module—dynamic image detection

4.3 Facial expression analysis module

The facial expression analysis module is also backend program. Due to close relations between facial expression and time sequence, facial expression analysis is conducted after all AU are recognized. If hold time of eye closure is ⩽0.5 seconds (around 5 frames), it is eye blink. If the time exceeds 0.5 sec, it is eye closure. The facial expression analysis module takes frame as analysis unit. It analyses AU from one frame. Thus, one frame produces one facial expression. The facial expression analysis rule is performed one frame by one frame as per AU combination table (Table 1). The facial expression analysis module (as shown in Figure 11) uses AU recognition results from the facial expression recognition module to analyse facial expressions.

Fig. 11 Operation picture of facial expression analysis

5 ERAS FOR TEACHING EXPERIMENT

The study investigated classes of grade three in data processing department of a vocational college in Taichung, Taiwan. Among the students who have normal visual and auditory sensory organs and are good at computers, one student with lower English score (low coding ability) and one student with higher English score (high coding ability) were selected. From difference between the two students in seeing and listening abilities, we analyzed and compared emotional reactions and facial expressions of the two students in learning, and applied the results to evaluate learning effectiveness and textbook difficulties.

5.1 Taching Material and Object

In order to make the seeing and hearing abilities of the students have significant difference, the textbooks with three stimulus degrees “Chinese/Chinese”, “Scenario/English”, and English/English were used in the experiment (as shown in Table 2). The three textbooks have visual and auditory sensory stimulus functions. Thus, we can discuss social reactions represented by facial expressions of students who had difference in seeing and hearing abilities after receiving stimuli of the three textbooks.

Please note that math equations might need to be reformatted from the original submission for page layout reasons. This includes the possibility that some in-line equations will be made display equations to create better flow in a paragraph. If display equations do not fit in the two-column format, they will also be reformatted. Authors are strongly encouraged to ensure that equations fit in the given column width.

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

Before teaching textbooks, teachers shall introduce learning units of the textbooks, and assist students in establishment of learning goal in each unit. This can help students know the learning points of the textbooks. After experiment, subjective measurement is used to assess cognitive load of mental efforts. The post-learning test may also evaluate learning effectiveness for teaching content, and asks several questions to test whether students can clearly understand visual and hearing contents. This is the basis for appraisal of learning effectiveness. The student with low English test score is regarded to have “low coding ability (S1)”, and the other student with high English test score is regarded to have “high coding ability (S2)”. The purpose of distinguishing the low and high coding abilities is to conduct difference comparison of facial expressions from experiment results. Background data of the tested students are shown in Table 3.

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

In the experiment, we obtained facial expression of S1 and S2 when learning the three textbooks with different stimuli and observed time sequence of facial expression reactions caused by the textbook contents. In the experiment, we can mainly take facial expressions in normal learning, and did not require students intentionally constrain or show certain facial expressions for system detection. Thus, tiny change in facial expressions and lighting source can change detection value. Learning state of the students can be observed through frequency and time interval of facial expressions. Based on information processing theory and cognitive load theory, if the students adapt to the textbook contents, they will have positive reactions or maintain neutral facial expression reactions; if they cannot adapt to the textbook contents, they will have negative facial expression reactions. For cognitive load estimate, the subjective measurement adopts scale. The first question is assigned with 3 score for Yes, 2 score for Average, and 1 score for No; the questions 2-4 are assigned with 1 score for Yes, 2 score for Average and 1 score.

5.2 Facial expressions and social reaction analysis

Experiment 1: Facial expression reactions towards textbook 1

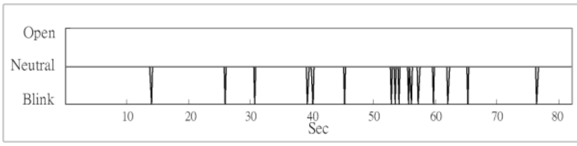
The textbook contents are presented in Chinese, and the auditory part is Chinese. For the students, information input in visual and auditory sensory organs was not restricted. Only their language seeing and hearing abilities were different in information coding, and this affected cognitive load. Thus, facial expressions are shown. According to the experimental results, besides more neutral facial expressions, S1 had 36 negative facial expressions and S2 had 22 positive facial expressions (as shown in Table 4). This indicates the student with high coding ability adapted to the textbook better than the other with low coding ability.

Table 4
Detection results of facial expression reactions to textbook 1

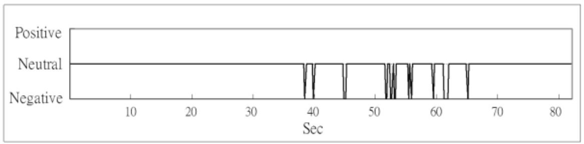
	Image length (sec)	Cognitive load	Positive		Neutral	Negative				Blink
			Happiness	Surprise	None	Anger	Fear	Sadness	Disgust	
S1	82	7	0	0	766	0	0	0	36	16
S2	83	4	22	0	794	0	0	0	0	12

In the experiment, S1 had 95.5% of neutral facial expressions, and S2 had 97.3% of neutral facial expressions. Both concentrated their attentions in learning. They were not required to intentionally constrain or show certain facial expressions. In the normal learning environment, emotional state was stable. Thus, many neutral facial expressions are normal. From time sequence of facial expressions, the negative facial expressions of S1 were concentrated between 38 and 65 sec (as shown in Figure 12a), and the blink frequency was also increased (Figure 12b). The system detection results showed disgust. From the subjective measurement result of S1, S1 considered the teaching speed was too quick. Cognitive load increased, and negative facial expressions showed. In the last half part, the blink frequency was decreased. It seemed that cognitive load may be reduced. During 20~30 sec, S2 were interested in robot animated images of operators and showed positive facial expression reactions (in Figure 13a). Blink frequency is around 8.7 times per minute (Figure 13b). This frequency is similar with that of a person who attentively reads before screen. This indicates cognitive load of S2 was still within his load capacity, and S2 adapted to and concentrate his attention on the textbook contents.

The subjective measurement results of the two students were "fully understand". In the textbook 1, percent of correct answers in the textbook 1 is 100%. Subjective measurement results are consistent with post-learning results (in Table 5). Although the students showed many neutral facial expressions, they can achieve learning effectiveness as long as they concentrated attention.

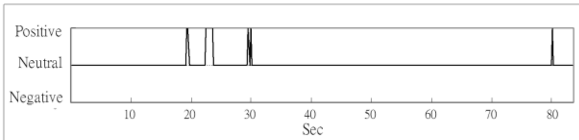


(a)

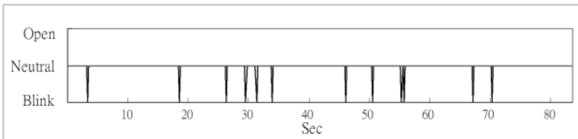


(b)

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



(a)



(b)

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)

Experiment 2: Facial expression reaction to textbook 2
The textbook contents are “somewhat difficult". The text, visual part, is presented by scenario and action, and the listening part is spoken by English. The students may have difficulty in auditory sense. If the students have correct coding for auditory information, the visual part can help them increase learning effectiveness; if they have wrong coding for auditory information, they may misunderstand the textbook only with visual information. Thus, coding of the scenario content can be correct when visual information is combined with auditory information. This is the situation that the experiment intends to create. From the experimental results, S1 showed 178 times of negative facial expression reaction towards the textbook, and S2 still showed 40 times of positive facial expression reaction (as shown in Table 6) besides neutral facial expressions. This shows the student with high coding ability adapted to the textbook better than the other student with low coding ability. Meanwhile, increasing textbook difficulty resulted in high probability of negative facial expressions shown by the student with low coding ability.

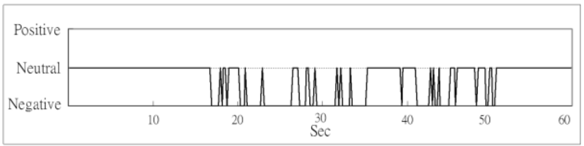
Table 6
Detection results of facial expression reactions to textbook 2

	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

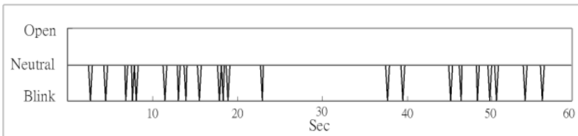
S1 percentage of negative facial expression reactions was increased by 31%. The textbook 2 is spoken by English for information transmission. This restricted hearing channel. Dual-channel information input was reduced to single channel. This also affected learning effect, and increased cognitive load. Thus, more negative facial expression reactions were shown. From the duration of facial expressions, S1 showed negative facial expressions during 17~35 sec and 39~50 sec (as shown in Figure 14a). At the end of the two time intervals, S2 showed positive facial expressions (as shown in Figure 15a). The first part of the textbook describes the hero forgot his wallet, and the last part describes the car key was locked in the car. The hero had two antics which triggered positive facial expressions of S2. In S1 learning sheet, all the answers in the auditory perception test were wrong, but visual perception test were right. The correct rate is 50% (Table 7). It can be seen that S1 used visual sensory channel to complement deficiency of auditory sensory channel. This lowered learning effectiveness, and information interestingness because of difficulty. Thus, S1 did not show the same positive facial expression as S2. This indicates facial expressions of the students are actually related to cognitive load. The student with low coding ability tended to show negative facial expression because of cognitive overload.

The average blink frequency of S1 was 22 times per minute in learning textbook 2 (Figure 14(b)), and 11.7 times per minute in learning textbook 1. The blink became more frequent. The average blink frequency of S2 was 10 times per minute in learning textbook 2 (Figure 15b), and 8.7 times per minute in learning textbook 1. The difference is small. S2 still showed positive facial expression. It can be seen S2 well adapted to the textbook 2. The correct rate of S2 was 100% (Table 7).

The subjective measurement result for S1 learning cognition is “fully understand visual part and partly understand listening part”. All the answers in the visual perception test were right, but in the listening test were wrong. This indicates subjective measurement overestimated listening part. The subjective measurement result of S2 cognitive load is "fully understand the visual part and the listening part". This indicates the measurement result is consistent with post-learning result (Table 7).

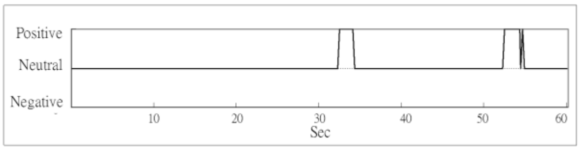


(a)

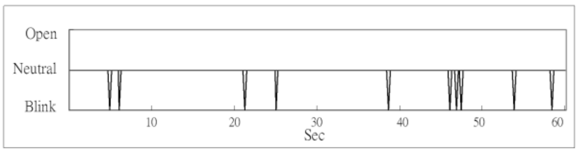


(b)

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



(a)



(b)

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	Yes(O)	Yes(O)

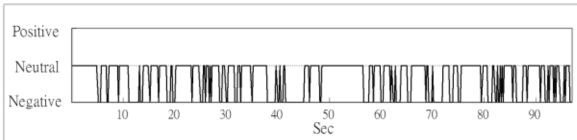
the hero?		
Is the dining table rounded?	Yes(O)	Yes(O)
Did the hero forget the car key when he returned to the house?	No(O)	No(O)
Did the hero return the house to make an important call?	Yes(X)	No(O)
Has the hero found the wallet?	Yes(X)	No(O)
Did the heroine guess the car key is locked in the car?	No(X)	Yes(O)

Experiment 3: facial expression towards textbook 3
In order to further test facial expressions of the student with different language visual and listening abilities under cognitive load caused by textbook difficulty, this textbook contents are extremely difficult. The text, namely visual part, is in English and the listening part is spoken in English. The visual and auditory senses of the students were restricted, and they had to code the information with their language coding ability. The cognitive load was increased by using more difficult textbook contents, so as to observe the facial expressions. The textbook contents are not standard course for vocational colleges, and are about the data structure – push and pop in the stacking unit. For the students, the textbook 3 was strange and not easy to understand. Visual and auditory sensory channels were restricted. Information coding process was arduous.
The experimental results showed S1 with low coding ability had 308 times of negative facial expressions (Table 8 and Figure 16) in learning the textbook, with increase rate of 32.4%. The correct rate only reached 16.7% (Table 8). The visual and auditory sensory channels were restricted. The learning had no effectiveness. Indeed, increase of cognitive load will induce more negative facial expressions. The S2 with high coding ability also had 26 times of negative facial expressions (Table 8 and Figure 17), with increase rate of 2.7%. This reveled the textbook 3 caused S2 cognitive load to reach critical value. The correct rate was decreased to 83.3% (Table 9). Teachers shall pay attention to this result. Teachers shall avoid using a strange way to present the textbook contents. Teachers listed terms and charts on the blackboard but failed to convert the language, resulting in cognitive overload of students. This is not useful for learning effectiveness.

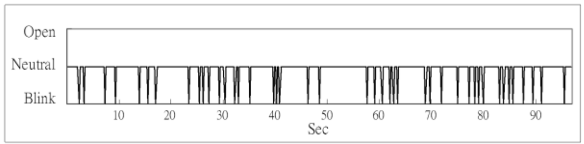
Table 8
Detection result of facial expression to textbook 3

	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

The blink frequency of S1 and S2 was 26.6 times and 18.8 times per minute respectively. In learning the textbook 3, both blink frequency was the highest. This revealed the textbook difficulty is positively related to the blink frequency. High textbook difficulty causes high cognitive load and distractions, accompanied by higher blink frequency (Figure 16(b) and Figure 17(b)). The subjective measurement result of S1 is "partly understand visual part and cannot understand listening parts". Only one answer was correct in the visual perception test, and all were wrong in the auditory perception test. The subjective measurement result is consistent with that of post-learning test (Table 9). The subjective measurement result of S2 cognitive load is "partly understand visual part, and partly understand listening part". All the answers by were correct in visual perception test, and only one answer was wrong in the auditory perception test. The subjective measurement underestimated the visual part. The listening part was main cause for cognitive load reaching the critical value. The subjective measurement for listening part is consistent with that of the post-learning test.

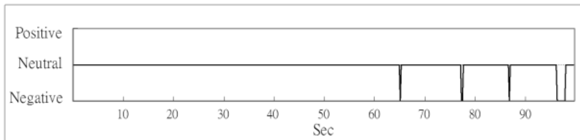


(a)

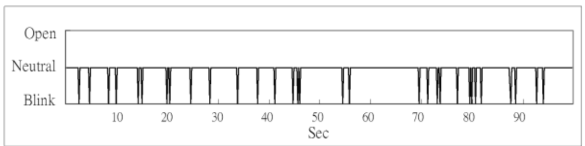


(b)

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



(a)



(b)

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	Yes(O)	Yes(O)

placed in the stack?		
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)

5.3 Experimental Results and Analysis

From information processing theory and limited capacity of cognitive load theory, if the information received by individuals exceed their processing capacity, cognitive load is high, and information processing fails. Students have worries and fears, even negative emotion frustration. The emotions shown on face are negative facial expressions. If the information processing does not exceed individual load capacity, the individuals can still effectively deal with information. In normal teaching environment, mental state of students is stable, and facial expressions are natural. Thus, many neutral facial expressions are rational. Further, learning has effectiveness as long as students concentrate their attentions. The research indicated foreign language anxiety is negatively related to performance of foreign language reading. Performance of the students with low anxiety is significantly higher than performance of the students with high anxiety. The students with low anxiety have better English reading performance than those with high anxiety. The anxiety emotion may affect performance of tested students. During learning, cognitive load is high if the students consume too many mental efforts. Mental pressure and learning anxiety are main cause of producing study emotions. The experimental results demonstrated facial expressions in learning is related to cognitive load. Thus, students with positive expressions have better learning effectiveness than those with negative expressions. In design teaching activities and contents, teachers shall consider cognitive load of the students, so as to maintain learning effectiveness.

We employed often used subjective measurement to assess mental efforts of the students. The two students considered the textbook 1 is the simplest, the textbook 2 is more difficult and the textbook 3 is the most difficult. The result is consistent with the experiment design. ERAS can observe the learning state of the students by detecting physiological reactions facial expression and eyes movement. The experimental results showed the higher learning difficulty is, the higher cognitive load is. The negative emotion instinctively appeared on face to form negative facial expressions. Also, distraction increases blink frequency. S2 showed positive facial expressions towards operator robots in the textbook 1 and hero antic in the textbook 2. In other time, S2 showed neutral facial expressions which can be interpreted as normal facial expressions towards learning. The students may show positive facial expressions towards their interested contents, but the duration is short. When the textbook transfers new information, and guide the students enter another coding phase, or degree of interest was reduced, the positive facial expression terminated. When total cognitive load of the students exceeds their load capacity, the facial expressions are negative due to pressure and anxiety. The duration of the negative expressions is longer, and last until students complete one information coding, cognitive load is reduced to the load capacity or the negative emotions disappear. Thus, teachers shall give sufficient rest time to students, such as telling story and joke, and dismissing class, and make them have more time to code information, and thus reducing cognitive load.

In this paper, dual-channel interaction of CTML was applied in learning experiment. The students were induced to show facial expressions by increasing their cognitive loads with different language visual and auditory abilities. ERAS can effectively detect and recognize facial expressions of the students, and help assess learning effectiveness of students and difficulty of the textbooks provided by the teachers.

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

6 Disscussion and Suggestion

The experimental result revealed that analysis of learning effectiveness through facial expressions has the following characteristics.

Facial expression is outward reflection of inner emotions. When the textbook arouses interesting from the students, the learning emotions can be reflected by facial expression. Generally, information can be interpreted by the individual experience and converted into inner perception. Positive experience or emotions (joyful, surprise, happiness, pleased, expected, affirm, trust, humor and curious) can produce positive facial expressions. Negative emotions (anger, fear, sadness, disgust, dull, dejected, reject, pressure, anxiety, pessimism, nervous and doubt) can produce negative facial expressions. S2 was interested in lovely operator robots in the textbook 1, and smiled. At the same time, S1 had no reaction; in the last half part, S1 was not pleased because of a bit faster teaching speed. Through the system detection, S1 showed the negative facial expression—disgust. At the same time S2 had no facial expression (Figure 18). In the textbook 2, it can be seen the situations were similar. S2 was interested in hero antics, and induced to show positive reaction. S1 cannot understand the scenario and attempted to guess the scenario contents, so interest degree of antics was lowered. S2 understood the dialogue contents and felt relaxed. The antics played a role (in Figure 19).

Students show positive facial expressions when they are interested in novel textbook contents. If they received negative information from negative textbooks, they showed negative facial expressions. When people have natural state and are not affected by emotions, their facial expressions will have no change. Even unconscious habitual actions are meaningless which are classified into neutral facial expressions in this paper. The above three facial expressions have the same value in evaluation of learning effectiveness. Positive expressions do not mean the best learning effectiveness, and negative expressions do not mean the worse learning effectiveness. Teachers shall preset facial expressions of the students towards the textbook contents, and check whether they are consistent with those in actual learning. If the teachers preset positive expression and the students show negative expression, this is not good, vice versa. The neutral reactions of S1 and S2 were 95.5% and 97.3% in the textbook 1. The correct rate was 100%. Especially, S2 neutral reactions exceeded 90%. The correct rate was also high (Table 11). The indicated the students have neutral facial expressions in leaning. The cognitive load was still within the load capacity, and the working memory was not overloaded. No negative emotions occurred. Thus, facial expressions had no change. The learning had effectiveness as long as the students concentrated their attentions on learning and had mental activities.

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%

Attention is paid to time in observing facial expressions. The students show positive facial expressions when they are interested in the textbook contents. However, the duration was short. The positive expressions are changed to neutral expressions (Figure 19) when the textbooks present new information, guide them to another coding phase, or interest degree is lowered. If negative emotions caused by cognitive overload evoke negative facial expressions, the neutral expressions can be restored after negative facial expressions or cognitive load was decreased to the load range. Thus, the duration is longer (in Figure 19). The time when the facial expression appears is also important. Correct facial expression shall occur at the correct time.

Only densely occurred facial expressions are meaningful. The research indicated [20] the sufficient time to accurately recognize micro expression is 1/5sec (200ms). Separate and accidental facial expressions or those expressions with less than 1/5 can be regarded as noise. From the diagram for facial expressions of S1 and S2 towards the textbook 3 (Figure 20), S1 showed negative facial expressions throughout the learning. This revealed S1 performance was not good when learning the textbook. The teacher can rearrange learning. S2 did not show negative expression until the final teaching content. This revealed S2 felt the final content was difficult. The teacher can arrange additional teaching activities for the final content. Determining density of facial expressions is orientation of future research. More objective quantitative data can be obtained through mass and broad experimental samples.

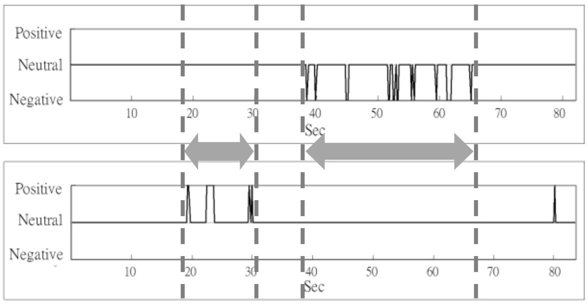


Fig. 18 Analysis chart of facial expressions to textbook 1

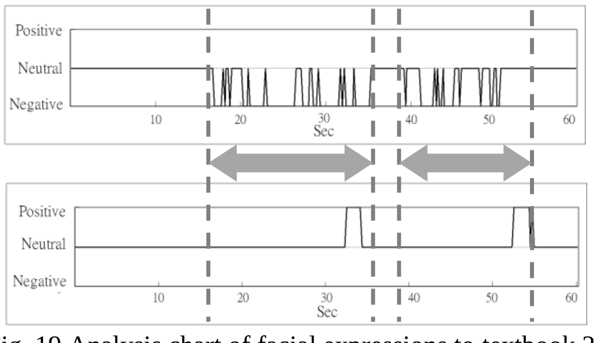


Fig. 19 Analysis chart of facial expressions to textbook 2

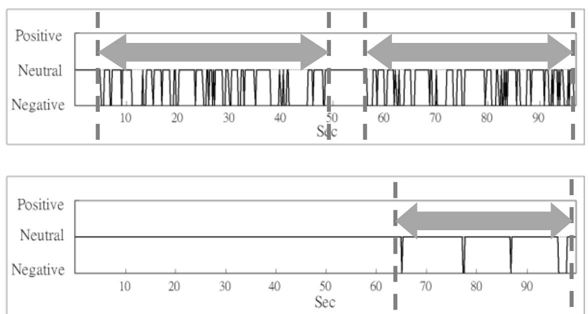


Fig. 20 Analysis chart of facial expressions to textbook 3

In group or remote teaching environment, teachers cannot observe and record learning state of each student. Thus, the textbook differentiation cannot be evaluated. The system can assist teachers in evaluation of adaptability of students to textbook difficulty. If the students adapt to the textbook contents well, learning difficulty of the textbook is lower. The students can learn the next unit or other deeper and broad textbooks; if the students cannot adapt to the textbook well, learning difficulty of the textbook is higher. The teacher can make the students relearn the textbook or select basic additional textbook.

The experimental results showed facial expressions may change with difficulty of textbooks. If the textbook difficulty is increased, negative facial expression are more obvious under high cognitive load. Thus, this system uses this feature in evaluation of learning difficulty of textbooks. The negative expressions of S1 were more apparent as the textbook difficulty increased. The percent of the negative expressions were increased from 4.5% in the textbook to 31% in the textbook 2 and 32.4% in the textbook 3. S2 showed no negative facial expressions towards the textbook 1 and textbook 2 but showed 2.7% of negative expression towards the textbook 3. It can be determined the textbook 3 is the most difficult. The teacher can introduce teaching contents before class, request students prepare the contents, and teach the textbook slowly. The teachers can also make use of repeatability of the multi-medium textbook and request the students to review it after class. Even, this can be used as reference for learning appraisal.

7 Conclusions

Teaching means teachers use teaching method to transfer the information stimulus by the textbook contents to the student brains via information sensory channels. These information stimuli are organized and integrated in a brain. Dealing with the information is called coding ability. The coding ability is related to brain reaction to information stimuli. The brain can make a correct reaction only when the information is coded successfully. We used more difficult textbooks in the experiment. Information stimuli coding ability actually affects inner emotions, and is indirectly reflected on facial expressions of the students. We recognize these facial expressions and classify them into neutral, positive and negative social reactions.

Based on the facial action coding system, Open CV and image processing method were used for face detection, feature extraction and location. A facial expression analysis system was implemented. This system extracted 10 face feature points, and used 9 feature vectors and neutral values as measurement benchmarks for identification and blink detection of 11 facial action units. It classified the action units into six basic facial expressions, happiness, surprise, fear, anger, sadness and disgust. Then these action units are divided by AU combinations into positive, neutral and negative social interactions.

In the experiment, the students were not required to show certain facial expressions for the system detection. In normal learning state, showing many neutral expressions is normal. If learning effectiveness needs to be assessed, other physiological signals such as eye movement should be considered to enhance reliability of the system identification. The experimental results showed with the same textbook the student with high coding ability was adapted to the textbook contents better, and showed more neutral expressions and positive expressions; the student with low coding ability showed more negative facial expressions. In order to demonstrate the result, we intentionally increase textbook difficulty and found the textbook difficulty and the facial expressions have an inverse relationship. The higher the textbook difficulty is, the more easily the negative expressions occur. Thus, students can bear the cognitive load before they show many negative expressions. Teachers can know student understanding of the textbook contents and adaptability to the information.

References

[1] K. Bousmalis, M. Mehu, M. Pantic, "Towards the automatic detection of spontaneous agreement and disagreement based on non-verbal behaviour: A Survey of related cues, databases, and tools," *Image and Vision Computing*. 31(2): pp. 203 - 221, 2013.

[2] R. W. Picard, "Affective Computing," Technical Report TR-321, MIT, Media Laboratory, 1995.

[3] A. Vinciarelli, M. Pantic, D. Heylen, C. Pelachaud, I. Poggi, F. D'ericco, M. Schroeder, "Bridging the gap between social animal and unsocial machine: A survey of Social Signal Processing," *IEEE Transactions on Affective Computing*. 3: pp. 69 - 87, Issue 1. April 2012.

[4] K. Nosu and T. Kurokawa, "Facial Tracking for an Emotion-Diagnosis Robot to Support E-Learning," *Proceedings of the International Conference on Machine Learning and Cybernetics, 2006,* pp.3811-3816.

[5] M. P. Loh, Y. P. Wong and C. O. Wong, "Facial Expression Recognition for E-learning Systems using Gabor Wavelet & Neural Network," *Proceedings of the International Conference on Advanced Learning Technologies,* 2006.

[6] Y. Wu, T. T. Wang and X. N. Chu, "Affective Modeling and Recognition of Learning Emotion: Application to E-learning," *Journal of Software,* Vol 4, No 8, 2009, pp.859-866.

[7] Y. Zhang and Q. Ji, "Active and Dynamic Information Fusion for Facial Expression Understanding from Image Sequences," *IEEE Transactions on Pattern Analysis and Machine Intellience,* vol.27, no.5, 2005, pp.699-714.

[8] L. Shen, M. Wang and R. Shen, "Affective e-Learning: Using "Emotional" Data to Improve Learning in Pervasive Learning Environment," *Educational Technology & Society*，Vol 12, No. 2, 176–189, 2009.

[9] Kuan-Cheng Lin and Szu-Ju Chen, "Affective Computing in Distance Learning: Based on Facial Emotion Recognition," *2012 2nd International Conference on Future Computers in Education, Shanghai, China.*

[10] Kuan-Cheng Lin, Rei-Wen Lin, Szu-Ju Chen, Ciou-Ru You, Jui-Lin Chai, "The Classroom Response System Based on Affective Computing. *The Third IEEE International Conference on Ubi-media Computing, Zhejiang Normal University, Jinhua, China.*

[11] Khadoudja Ghanem and Alice Caplier, "An Efficient Method to Classify Positive and Negative Expressions," *International Journal of Tomography & Statistics*, vol. 17， no. 11， pp.72-78.

[12] R. E. Mayer, "The Cambridge handbook of multimedia learning," *NY: Cambridge University Press, 2005*.

[13] J. Sweller, J. van Merrierboer and F. Paas, Cognitive architecture and instructional design." *Educational Psychology Review,* Vol 10, No 3, 1998, pp.251-296.

[14] P. Gerjets and K. Scheiter, "Goal Configurations and Processing Strategies as Moderators Between Instructional Design and Cognitive Load: Evidence From Hypertext-Based Instruction," *Educational Psychologist,* Vol 38, No 1, 2003, pp.33-41.

[15] R. Brunken, J. Plass and D. Leutner, "Direct Measurement of Cognitive Load in Multimedia Learning." Educational Psychologist, Vol 38, No 1, 2003, pp.53-61.

[16] P. Ekman, "Emotions Revealed: Understanding Faces and Feelings," Weidenfeld & Nicolson, 2003.

[17] P. Ekman and W. V. Friesen, "The Facial Action Coding System: A Technique for The Measurement of Facial Movement", Consulting Psychologists Press, San Francisco, 1978.

[18] D. S. Bolme, M. Strout and J. R. Beveridge, "FacePerf: Benchmarks for Face Recognition," Proceedings of the International Symposium on Workload Characterization, 2007, pp.114-119.

[19] J. Whitehill and C. W. Omlin, "Haar Features for FACS AU Recognition," Proceedings of the International Conference on Automation Face and Gesture Recognition, 2006, pp.5-9.

[20] X. Shen, Q. Wu and X. Fu, "Effects of duration of expressions on the rercognition of microexpressions." Journal of Zhejiang University-Science B: Biomedicine & Biotechnology, Vol 13, No 3, 2012, pp.221-230.

Kuan-Cheng Lin is working as a associate professor in the Department of Management Information System, National Chung-Hsing University.

Kun-Hshiang Chiang is a student in the Department of Management Information System, National Chung-Hsing University.

Jason C. Hung biography appears here.