Prototype of Educational Affective Arousal Evaluation System Based on Facial and Speech Emotion Recognition

Jingjing Liu and Xiaofeng Wu

Abstract—**Educational theories and empirical researches indicate that teaching objectives are not limited to the field of knowledge and cognition, and emotion has also proved to be an important part of learning outcomes and future achievements. An affective arousal evaluation system focusing on the self-system goal of education is proposed in this paper, which introduces artificial intelligence (AI) techniques to evaluate affective arousal level as a part of teaching quality in the form of quantitative indicators. The system embeds facial expression recognition and speech emotion recognition methods, which extract and analyze the video streaming collected by the ubiquitous web cameras from both teacher side and student side during the online teaching process. The experiment verifies the correlation and Granger causality between teacher-student affective sequences. The implementation of the system realizes the output of three quantitative indicators: "Affective Frequency Index", "Affective Correlation Index" and "Affective Arousal Level".**

Index Terms—**Affective arousal, expression recognition, speech emotion recognition, teaching evaluation.**

I. INTRODUCTION

According to Marzano's new taxonomy, educational objectives include four categories: knowledge, cognition, metacognition, and self-system. Marzano proved that the effect size between self-system objective and knowledge acquisition is the biggest among all. One of the sub-level self-system objectives is to check emotional responses [1]. For students themselves, they should clearly realize their willingness and motivation towards learning through checking emotional responses. On the other hand, teacher can also adjust their teaching strategies by understanding students' emotional responses. Tracing back to Bloom's taxonomy, educational objectives are divided into three domains: cognitive, psychomotor, and affective domain [2]. Despite its flaws in cognitive processes, affective domain has been of great significance since then. Affective arousal refers to the process of people's emotional initiation under certain conditions, accompanied by physiological and psychological changes, and affecting the processing of cognitive content. Based on the pedagogical significance and detectability of emotions, this paper attempts to design an AI-based system so as to evaluate the affective arousal level of teaching, and to setup a specific quantitative evaluation index, which can be used as teaching evaluation or teaching aid. The designed affective arousal educational system can realize process evaluation in class at the same time.

The completion of pedagogy research has laid a theoretical foundation for the affective arousal evaluation system, while the rapid development of network and AI techniques have made the realization of the system possible.

In the era of explosive growth of information and communication technology (ICT), online education has penetrated into educational systems at all levels, transforming traditional offline teaching scenarios. Online education refers to learner-centered non-face-to-face education using multimedia networks and, which, like traditional offline classrooms, can also elicit powerful emotions [3]. Online education often uses computers, tablets and mobile media to carry out education and teaching in the form of audio-video interaction. This new form of teaching provides a huge amount of reliable data sources for us to apply AI techniques to evaluate the teaching process.

In order to focus on the affective arousal state of teachers and students in case of online classrooms, expression recognition algorithm and speech emotion recognition algorithm are introduced into our system. It is verified that the face++ expression recognition algorithm used in our system has a 7-category emotion recognition accuracy of 73.95% and a triple classification (i.e. positive, neutral, and negative) accuracy of 84.13% testing on the Extended Cohn-Kanade (CK+) facial expression database [4]. The Mel Frequency Cepstral Coefficient (MFCC) based speech recognition algorithm testing on the CASIA (Institute of Automation, Chinese Academy of Sciences) Chinese emotional speech corpus reached an accuracy of 80.4%. Both the performance of expression recognition and speech emotion recognition can meet our system design requirements.

Under the premise that pedagogical research on emotion is gradually completed and the artificial intelligence techniques can be integrated with online education platform on the technical level, we first proposed and designed the prototype of educational affective arousal evaluation system and implement the system with actual video streaming from online classes.

The rest of the paper is organized as follows. Section II describes some of the works related to our topics of interest. The prototype of our proposed system with details is explained in Section III. In section IV, preliminary experiments are carried out on the proposed prototype. At last, conclusion along with future works is clarified in Section V.

II. RELATED WORKS

A. Pedagogical Research in Affective Domain

Although theoretical study has shown the importance of emotion in education, in the past academic research and

teaching practice, it has been undoubtedly the cognitive field that has received great attention. This phenomenon can be partly ascribed to the fact that existing teaching evaluation indicators are highly dependent on academic performance, and the relationship between cognitive field and academic achievement is very intuitive and highly relevant. So it is relatively easy to do educational research or implementation in cognitive domain rather than affective domain.

The good news is that in the past decade or so, more and more research has focused on the discrete emotions of students in the classroom. Academics and educators have begun to pay attention to the critical role of the affective domain in students' academic achievement and future career choices. There are numerous researches and empirical evidence on the impact of students' emotions. Researches indicate that emotions predict important learning strategies and outcomes (including lifelong learning) from educational-psychological perspective [5], [6]. Emotions in the classroom are also proved to have an impact on future career choices [7].

On the other hand, theoretical and empirical studies have shown that teacher's emotions are the decisive factor in students' mood in class, which indicates that teachers should pay attention to the educational objective of affective arousal.

Ginott first pointed out the influence of teachers' emotions on classroom atmosphere and student emotions. He found that he as a teacher is the decisive factor in the classroom, and his personal approach as well as daily mood create the climate in class [8]. An empirical research shows that teacher's enjoyment and student's enjoyment within classrooms are positively linked and that teacher's enthusiasm mediates the relationship between teacher's and student's enjoyment [9]. Another research was carried on 15 lessons in four different subject domains. Research studied the relationship between teachers' emotions, their instructional behavior, and students' emotions in class by carrying experience sampling analysis. It turned out that teachers' and students' emotions in class are closely related and that teachers' emotions explain incremental variance in students' emotions above and beyond their instructional behavior [10]. Researches on characteristics of teaching also comes to the consistent conclusion that positive teaching characteristics including enthusiasm and fostering attention are positively related to students' enjoyment [11].

However, the emergence of online education has brought new questions to the emotional interaction between teachers and students: Can students get emotional experience online? Psychological research has shown that all courses, including online courses, have the potential to elicit powerful emotions that inspire long-term creation of knowledge [3]. That is to say, the role of emotion in pedagogy also applies to online education.

It can be indicated from these studies that the impaction of teacher emotions on students' emotion in class has been verified in various course subjects and teaching platforms (online or offline), which provides an important theoretical basis for the design of our educational affective arousal system.

B. AI Techniques: Facial and Speech Emotion Recognition

At present, research on facial expression recognition continues to iterate, and existing expression recognition algorithm has already been able to adapt to some application scenarios. The methods of facial expression recognition range from traditional method of SVM classification using facial feature points extracted from facial images [12], to deep learning method using DNN method or CNN method base on feature graph [13], [14], all have good recognition outcome. At the same time, companies such as Microsoft or Face++ provides callable expression recognition API tools that make expression recognition easy to apply to vary application scenes.

Speaking of speech emotion recognition, there are many ways to do speech emotion recognition whereas the recognition rate is relatively uneven. Due to the particularity of speech recognition tasks at the individual level, algorithm migration is a difficult task. In the field of speech emotion recognition, the end-to-end deep learning method for continuous emotion recognition from speech has achieved satisfying results [15]. At the same time, the methods based on the classification of acoustic features such as MFCC still have superiority. Likitha, Gupta and Raju successfully distinguish three emotions (i.e. sad, happy and anger) at an accuracy rate of 80% by simple using standard deviation operation of the extracted MFCC features [16].

C. Status of AI Application in Online Education

Online education relies on video and audio interaction during teaching process, so AI technology can easily get access to those data and directly participate in the monitoring and evaluation of educational processes. Online education has a broad market space around the world with large number of participants. Many of them try to improve the quality of teaching by combining AI technology. Still, almost none of them has a clear pedagogical objective for the participation of AI techniques.

Taking online education institution "Haifeng" (an institution origin in Shanghai, with 8 years' education and training history) as an example, the Cape of Good Hope system launched by Haifeng in April 2018 is said to be the first artificial intelligence application launched in the K12 field in China. This AI system analyzes students' emotions through expression recognition, and analyzes students' attention based on eyeball focus at the same time. The system would send immediate feedback of student's current status to the teacher, focusing on whether the student is attentive. Unfortunately, the role of emotion recognition in online education is always focused on the tracking of students' concentrate level, which is lack of pedagogical significance. The connection between students' concentration and educational achievements has not been admitted by the academic community. Making students focused in class is also not an educational objective. In addition, the existing expression recognition applications are aimed at monitoring student's learning process, while ignoring the evaluation of teacher's teaching performance in the other way.

Viewing three aspects of related works, each aspect is rather independent. Pedagogical research in affective domain proves the existence and impaction of emotional interaction in class. Yet there is no application in accordance with the pedagogical research. AI techniques have promising

potential embedded in educational field. But obviously they are not fully utilized so far. The work of the existing online education system is superficial, without digging into the core of educational objectives or making the best of AI techniques. It is based on this splitting situation that we try to systematically organize the work of our predecessors and design our educational affective evaluation system.

III. EDUCATIONAL AFFECTIVE AROUSAL EVALUATION SYSTEM

A. System Overview

The educational affective arousal evaluation system aims at monitoring the emotional interaction between students and teachers. More importantly, the system would give out automatic teaching quality assessment in affective domain per class regularly, instead of the generally applied method of using students' ratings on assessing teaching quality per semester. The system offers real-time feedback of teaching quality assessment without taking students' time and effort for additional rating.

The educational affective arousal evaluation system is divided into 3 layers: an input layer, a data processing layer, and an output layer, as is shown in Fig. 1.

1) Input layer

Our educational affective arousal evaluation system is based on online education platform. Under online teaching scenario, students and teachers communicate with each other through computers, tablets or mobile phones, usually under the platform of a third-party institution. These media well support the need for video interaction, so that video and audio of teachers and students can be easily collected as input to our system.

Fig. 1. Frame of affective arousal evaluation system.

2) Data processing layer

At the data processing layer, our task is to apply artificial intelligence technology to process video and audio data of teachers and students, mainly using expression recognition technology and speech emotion recognition technology. For the video data of teachers and students, the expression recognition module is applied to identify and score them. So two time sequences with scores, namely sequence 2 and sequence 3 in Fig. 1, can be obtained. For the audio data of

teachers, speech emotion recognition module is applied to get the scored sequence, which is sequence 1 in Fig. 1. in addition, the recognition and evaluation of the head posture changes are performed on the student's video data, as is sequence 4 shown in Fig. 1. The paired data of the teacher and the student are respectively added to obtain a paired affective arousal sequences of both the teacher and the student.

3) Outputl layer

At the output layer, frequency analysis and correlation analysis are conducted on the affective arousal sequences of teachers and students, so as to obtain the output of three quantitative indicators: "Affective Frequency Index", "Affective Correlation Index" and "Affective Arousal Level".

B. Expression Recognition Module

The expression recognition module uses the Face++ Detect API. Its expression recognition return has seven categories: anger, disgust, fear, happy, neutral, sad, and surprise. The return accurate to three significant digits after the decimal point, with a range of [1,100].

Due to the particularity of the application scenario of education, we do not need to classify the classification results into specific seven categories. Instead, the emotions are reclassified into three categories, namely positive emotions, neutral emotion and negative emotions (see Table I).

TABLE I: THE RECLASSIFICATION OF EXPRESSIONS

Positive	Neutral	Negative
happy, surprise	neutral	anger, disgust, fear, sad

These three groups of emotions play different roles in students' academic achievement and further life. Positively emotions lead people to think, feel, and act in ways that promote both resource building and involvement with approach goals, thus promote career success [17], [18]. As what John Masefield said, "The days that make us happy make us wise." Numerous studies have demonstrated the positive impact of positive emotions on all aspects of human achievement. On the other hand, negative emotions also have something to do. They help solve the problem of enlightenment [19].
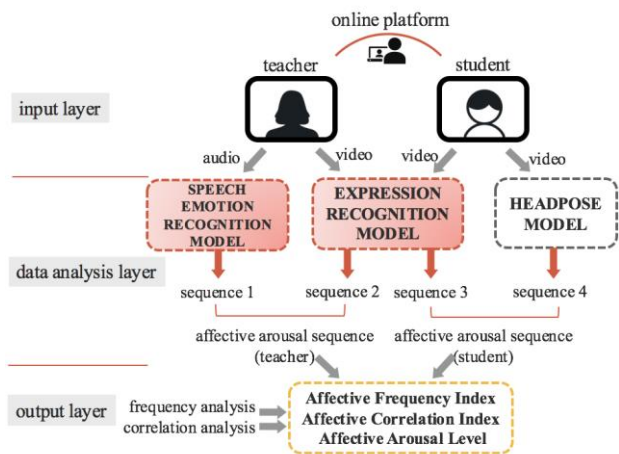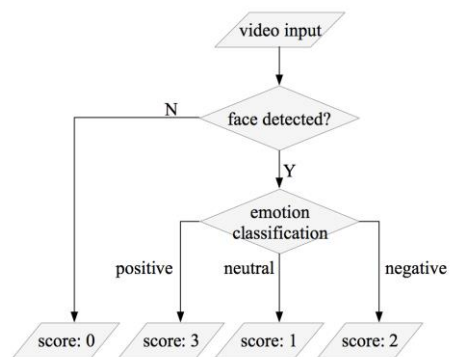

Fig. 2. Scoring method for expression recognition.

In order to present the emotion arousal status during a period, we assign scores to the detected positive, neutral, and negative emotions as shown in Fig. 2. If any expression is detected, the basic score is 1 point. The positive emotion is 3 points, and the negative emotion gets 2 points. So time

sequence labelled with expression recognition score is obtained.

To verify the performance of the expression recognition module, a test experiment is conducted. The test experiment was on the Extended Cohn-Kanade (CK+) facial expression database, using 334 facial images from 30 people. The experimental results show that under 7-category emotion classification, accuracy of expression recognition module is 73.95%. Under 3-category emotion classification, the recognition accuracy rate achieves 84.13% (see Table II, III).

TABLE II: RESULT – 7-CATEGORY EXPRESSION RECOGNITION

anger	disgust	fear	happy	neutral	sad	surprise	
23.40%	75.47%	33.33%	87.93%	85.00%	80.00%	96.10%	**73.95%**

TABLE III: RESULT – 3-CATEGORY EXPRESSION RECOGNITION

positive	neutral	negative	
92.59%	85.00%	75.54%	**84.13%**

In addition, the Face++ Detect API is also used for head pose recognition, which will be briefly described here. For the student's video data, in addition to facial expression recognition, head pose recognition is also used as an aid. The return value of head pose recognition is a float number with a range of [-180, 180], which is accurate to 6 significant digits after the decimal point, and the unit is angle. There are three return values of head pose, namely pitch angle, roll angle, and yaw angle. When the face pose change is greater than $10\degree$ between adjacent time interval, a corresponding score of 0 to 3 points is assigned. The sequence of the face pose recognition score is thus obtained (see Table IV).

TABLE IV: SCORING METHOD FOR FACE POSE

| Δpitch_angle | $>10\degree$	| Δroll_angle | $>10\degree$	| Δyaw_angle | $>10\degree$
1	1	1

C. Speech Emotion Recognition Module

The speech features include semantic features and acoustic features. To implement speech emotion recognition in the teaching process, a method of extracting acoustic features is adopted, which is simple and sufficient to achieve the desired classification effect.

According to Russell's classic emotional model, emotions are divided into two dimensions, arousal dimension and valence dimension [20]. The degree of arousal dimension indicates whether the emotion is relaxed or intense, and the valence dimension indicates whether the emotion is pleasant or not. In general, speech emotions can be divided into six categories, namely, anger, happy, surprise, neutral, fear, and sad. Among them, anger, happy, and surprise are high-initiation emotions, whereas neutral, fear, and sad are low-initiation emotions.

In the recognition of teacher's audio, we are concerned about the performance of emotions in the dimension of arousal, which can well reveal teacher's effort in emotion arousal. In terms of implementation, the spectral-based correlation analysis feature "Mel Frequency Cepstral Coefficient (MFCC)" is adopted. MFCC is a commonly used speech emotion feature, which can accurately display the shape of the channel in the envelope of the short-term power spectrum of the speech. MFCC is also a frequency domain speech feature parameter based on human hearing characteristic and have a good robustness.

The top 13 MFCCs (i.e. numcep=13) are extracted each time in the process. The window length is set to 0.025 seconds (i.e. wintime=0.025s), and the time interval for extracting MFCC is 0.01 second (i.e. hoptime=0.01s).

MFCC feature belongs to the personalized speech emotion feature, which means unfortunately there remains great variance between different individuals. To solve this problem, in addition to extracting the MFCC feature, dynamic difference parameters of MFCC are also extracted as follows.

$$d_t = \begin{cases} C_{t+1} - C_t & , t < K \\ \dfrac{\sum_{k=1}^{K} k(C_{t+k} - C_{t-k})}{\sqrt{2\sum_{k=1}^{K} k^2}} & , K \le t < Q - K \\ C_t - C_{t-1} & , t \ge Q - K \end{cases} \tag{1}$$

where d_t is the first-order difference characterizing the change of speech emotions ; C_t is the cepstrum coefficient in time t; Q is the order of cepstrum coefficients; and K is the time interval of the first derivative. The value chosen for K is 2. The second-order difference parameter can be obtained by bringing the first-order difference result into (1) again, which characterizes the degree of speech emotion arousal.

Optimized MFCC parameters are obtained by averaging MFCC parameters, MFCC first-order difference parameters and MFCC second-order difference parameters. Optimized MFCC parameters can be considered as non-personalized speech emotion features.

In order to use the optimized MFCC parameters to score the emotional arousal dimension, a simple and clever standard deviation method is applied. First, the MFCC parameters are averaged over a period of time (set to 1 second in the designed system) according to (2).

$$x = (x_1 + x_2 + x_3 + \cdots + x_n)/n \tag{2}$$

Average operation can compensate for fluctuations caused by window truncation. Then the standard deviation of parameter x is calculated. The standard deviation represents emotion arousal level of speech. Likitha used to apply the MFCC standard deviation method to identify three voice emotions of happy, sad and anger, with a correct rate of about 80% [16].

In speech emotion recognition module, the score is defined according to the MFCC standard deviation, and the range here is still from 0 to 3 points (see Table V). The higher the score, the higher the emotional arousal level, which means that the teacher's mood is high. The threshold is determined based on actual course data in the following experiment.

TABLE V: SCORING METHOD FOR SPEECH EMOTION

std$\le$0.54	0.54<std$\le$0.575	0.575<std$\le$0.61	std$\ge$0.61
0	1	2	3

The performance of the speech emotion recognition module was also tested. The dataset is 1200 speech phrases in CASIA Chinese emotional speech corpus, including 4 people, 6 emotions, and 50 speech phrases each. Only high and low-initiation speech emotions are classified in the dimension of arousal level. The classification accuracy of common MFCC features is 76.40%, and the result of

optimized MFCC features is improved to 80.4% (see Table VI, VII).

anger	fear	happy	neutral	sad	surprise	
59.00%	55.50%	51.50%	48.00%	89.00%	79.00%	**76.40%**

anger	fear	happy	neutral	sad	surprise	
63.00%	61.50%	53.50%	57.00%	88.00%	79.00%	**80.40%**

D. Design of Evaluation Index

Through the above process, video data have been processed into emotion-related multiple scored time series.

In the evaluation index design module, which is the output layer of the system, frequency analysis and correlation analysis are applied to the time series to get three quantitative indicators: "Affective Frequency Index", "Affective Correlation Index" and "Affective Arousal Level".

"Affective Frequency Index" represents the level of emotional activity of student in class. This index is an absolute indicator, which does not take the order of sequences or teacher-student relationship into account. "Affective Correlation Index" represents the extent of emotional interaction between teacher and student. It shows whether teacher's and student's emotions are consistent in time order, supplementing the defects of the first indicator. For example, a student can be laughing at a comedy on his phone during class time when teacher is conveying his words in vain. In such situation, "Affective Frequency Index" can be meaningless high. And "Affective Correlation Index" will reveal the irrationality of the first index and do some revision to get the last index "Affective Arousal Level". "Affective Arousal Level" considers both emotion frequency and emotion correlation. It shows overall teaching quality assessment outcome in our designed system. For each of the three indexes, the higher the value, the better the quality of teaching.

The teacher's affective arousal sequence is the sum of the teacher's expression recognition score sequence and the speech emotion recognition score sequence, which is referred to as T sequence; the student's affective arousal sequence is the sum of the expression recognition score sequence and the face pose score sequence, which is referred to as S sequence. T sequence and S sequence both range from 0 to 6 points, and the adjacent time interval of each sequence is 1 second.

1) Definition of affective frequency index

The "Affective Frequency Index" is designed to assess the frequency of students' emotion arousal in the classroom. It is already clarified that students' emotions in class have a great impact on teaching outcomes, so "Affective Frequency Index" is designed to assess the extent of this impact through checking emotion arousal frequency.

$$f = \frac{Count\ of\ S\ sequence\ score\ greater\ than\ or\ equal\ to\ 3}{Total\ length\ of\ S\ sequence} \quad (3)$$

$$Freq = \begin{cases} 100 \cdot \log_{10}(9f/0.4+1) & ,\text{if } f \leq 0.4 \\ 100 & ,\text{if } f > 0.4 \end{cases} \quad (4)$$

where f indicates the percentage of the total duration when the student's emotion is arousal in the classroom. f and the subjective questionnaire survey result are not linear, so f is converted to the percentage score *Freq* by empirical formula to get the "Affective Frequency Index" value. The threshold here is set to 0.4.

2) Definition of affective correlation index

"Affective Correlation Index" is designed to assess the relevance of emotional arousal between teachers and students in class. Thus to assess the extent to which teachers' emotions and teaching guidance affect students' emotions.

In the correlation analysis, the sequence pairs are segmented. Window length of the sequence pair is set to be 60 seconds and the interval time is 30 seconds.

$$R = \frac{Length\ of\ significant\ correlated\ sequences\ (p < 0.05)}{Total\ length\ of\ paired\ sequences} \quad (5)$$

$$Corr = \begin{cases} 100 \cdot \log_{10}(9f/0.6+1) & ,\text{if } R \leq 0.6 \\ 100 & ,\text{if } R > 0.6 \end{cases} \quad (6)$$

where R represents the proportion of duration where the correlation of the teacher-student sequence is significant. Those correlation confidence smaller than 0.05 (i.e. $p < 0.05$) is defined as significant. The value is also translated to percentage score similar to the first index, at which the threshold is set to 0.6.

3) Definition of affective arousal level

In order to balance the frequency and correlation of emotions, the comprehensive index "Affective Arousal Level" is designed This index, along with the former two indexes, is used to evaluate teacher's teaching quality in the aspect of realizing affective arousal objective in teaching objectives. We call it "AAL" for short.

$$AAL = 100 \cdot (Freq/100)^{1.2 - R/100} \quad (7)$$

AAL is adjusted by *Corr* on the basis of *Freq* as a whole. If the score of the correlation index is low, the "Affective Arousal Level" score will be lowered accordingly, and vice versa. AAL is also a percentage score.

IV. EXPERIMENTS

A. Experimental Setup

The experiment materials include 3 pairs of online course videos (6 videos in total) of teachers and students with the duration of 10 minutes each. All the courses are taught in one-to-one form in Chinese. And all the subject is high school mathematics, the main content of which is elliptic problems in conic curves and equations. Our purpose is to implement and verify our prototype despite the small data scale.

One pair of the videos are selected for comparative experiments. The videos are manually labelled with expressions, speech emotions and head poses, and compared with the output of the designed evaluation system.

B. Experimental Results

In the comparative experiment, the results of the manually labeling of the first group are compared with the systematic extracting ones. The results are displayed by giving out quantitative indicators and visualizing T sequence and S sequence as shown in Fig. 3.
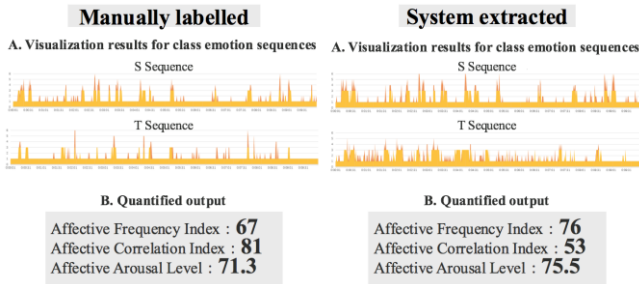

Fig. 3. Result of comparative experiment (left: manually labelled; right: system extracted).

Visualized T sequence and S sequence is superimposed by two parts, which are presented by yellow and orange color respectively. From the visualizing output, it can be seen that teacher's and student's emotional sequence in class are consistent whether it is manually labelled or systematic extraction.

The intermediate variables of manually labeling experiment is $f = 16.2\%$, $R = 36.8\%$, and $f = 21.3\%$, $R = 15.8\%$ in system extraction experiment.

After the comparative experiment, all three video pairs of teachers and students were then sent into the affective arousal evaluation system. The results are also illustrated by the form of quantitative indicators and visualization of T sequence and S sequence as shown in Fig. 4.
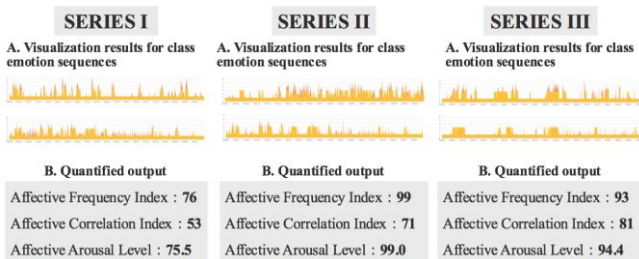

Fig. 4. Output of the system (series I, II, III).

The quantitative indicator output can be seen as a way of describing the class affective arousal sequences. So far, the system has been designed and implemented for affective arousal evaluation of education for the first time.

C. Verification Experiments

In section II, some researches verifying the relevance of teacher and student emotions in class have been introduced [9], [11], and empirical researches also proved that teacher's emotions explain the incremental variance of students' emotional changes [10], that is to say, teacher emotions and student emotions should have a time-order causal relationship. The results of our educational affective arousal evaluation system should share a consistency with these studies. So a variety of correlation tests and Granger causality tests were performed to verify the system reliability.

In the correlation analysis, Pearson correlation coefficient, Spearman correlation coefficient and Kendall correlation

coefficient are tested respectively on manually labelled sequences and system extracted sequences.

Segmentation correlation analysis is carried on in the case of an interval of 30 seconds and a window length of 60 seconds. From the results, it can be given that teacher's and student's sentiment sequence is partially significantly correlated, in line with our expectations (see Table VIII, IX).

TABLE VIII: CORRELATION ANALYSIS RESULT OF MANUALLY LABELLED SEQUENCES ($p <0.01$:**; $p <0.05$:*)

	1	2	3	4	5	6	7
Corr_Pearson	0.5696**	0.6234**	0.2909*	0.0262	0.0092	0.0407	0.0528
Corr_Spearman	0.5592**	0.3932**	0.2268	0.1070	0.0092	0.0571	0.0737
Corr_Kendall	0.5179**	0.3738**	0.2188	0.0999	0.0483	0.0517	0.0695
	8	9	10	11	12	13	14
Corr_Pearson	0.3439**	0.3796**	0.6471**	0.738**	-0.0781	-0.0948	-0.0949
Corr_Spearman	0.2266	0.3494**	0.6506**	0.5683**	-0.0832	-0.047	-0.047
Corr_Kendall	0.2119	0.3226**	0.6182**	0.5402**	-0.0823	-0.046	-0.0461
	15	16	17	18	19		
Corr_Pearson	0.0941	0.0724	-0.1187	-0.0639	0.1296		
Corr_Spearman	0.0914	0.0929	-0.0717	-0.0891	0.0859		
Corr_Kendall	0.0874	0.0894	-0.069	-0.0835	0.0827		

TABLE IX: CORRELATION ANALYSIS RESULT OF SYSTEM EXTRACTED SEQUENCES ($p <0.01$:**; $p <0.05$:*)

	1	2	3	4	5	6	7
Corr_Pearson	0.3597**	0.4354**	-0.0376	-0.2187	-0.0481	0.0648	-0.0323
Corr_Spearman	0.3771**	0.412**	-0.1058	-0.2618*	-0.0719	-0.0063	-0.0829
Corr_Kendall	0.3376**	0.3758**	-0.0971	-0.2372*	-0.0646	-0.0035	-0.0722
	8	9	10	11	12	13	14
Corr_Pearson	0.1233	0.103	-0.1097	0.0257	-0.1038	0.0923	-0.0487
Corr_Spearman	0.0976	0.0436	-0.0986	0.1286	-0.111	-0.0009	-0.108
Corr_Kendall	0.0872	0.0333	-0.0901	0.1213	-0.1083	0.0000	-0.0983
	15	16	17	18	19		
Corr_Pearson	-0.11	0.1061	-0.1742	-0.2728*	-0.0997		
Corr_Spearman	-0.1055	0.1484	-0.1493	-0.2979*	-0.1252		
Corr_Kendall	-0.0988	0.1401	-0.1383	-0.2708*	-0.1196		

In order to verify the contagious effect of teacher emotion on students' emotions, Granger causality test is performed on teacher's and student's affective arousal sequences in the case of manually labelling and system extraction. The results prove our hypothesis that the T sequence and the S sequence are Granger causality. Specifically, the teacher's emotional sequence is the Granger factor of the student's emotional sequence (see Table X, XI).

TABLE X: T IS THE GRANGER CAUSE OF S (MANUALLY LABELLED)

granger cause (alpha=0.05)	F	c_v	p	
T does not granger cause S	41.263	3.857	2.72E-10	reject
S does not granger cause T	3.053	3.857	0.081	accept

TABLE XI: T IS THE GRANGER CAUSE OF S (SYSTEM EXTRACTED)

granger cause (alpha=0.05)	F	c_v	p	
T does not granger cause S	1.657	1.646	0.198	reject
S does not granger cause T	0.351	1.646	0.554	accept

V. CONCLUSION

This paper proposes a prototype of affective arousal system for the self-system objective of education. We introduce artificial intelligence techniques to evaluates teaching quality specifically teacher's affective arousal level in the online education scene by the form of quantitative indicators.

With the incredible rapid development of AI techniques in emotion recognition, those techniques can be applied into different application fields. In this paper, latest AI techniques including facial and speech emotion recognition are integrated into the designed educational evaluation system, which is a pioneering work.

In this paper, a prototype of teaching quality evaluation for the emotional field of educational objectives is first proposed. At present, the application of emotional assessment in pedagogy focuses on the detection of students' attention only, ignoring the intercommunication between teachers and students. The design of this system focuses on evaluating the teaching quality by monitoring both teacher and student in pair, trying to establish affective evaluation of teachers' teaching. At the same time, system serves to aid process evaluation for students' learning procedure.

In addition, for the evaluation of teaching quality, this paper proposes a new process-based automatic evaluation system, which is completely different from the traditional method by investigating students' subjective feelings or by test results.

This paper presents the quantified output of the sample course sequences designed by the educational affective arousal evaluation system. It also needs to be pointed out that the system is a preliminary prototype and still has to be improved in teaching practice. The present absolute scores of the quantized output are relatively subjective scores, and there is no standard to compare at this stage. After a huge amount of teaching data is sent to the system, the threshold and score conversion formulas set in the scoring process can be adjusted and verified, so that the system can have better adaptability and greater application value.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Jingjing Liu and Xiaofeng Wu conducted the research, analyzed the data and wrote the paper; all authors had approved the final version.

REFERENCES

[1] R. J. Marzano and J. S. Kendall, *The New Taxonomy of Educational Objectives*, Corwin Press, 2006.

[2] T. M. E. Bloom's, "Bloom's taxonomy of educational objectives," *Longman*, 1965.

[3] A. Tix and M. Johnson, "If emotion aids learning, does it work online?" *Chronicle of Higher Education*, vol. 63, no. 9, 2016, pp. B34–B35.

[4] P. Lucey *et al.*, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," presented at 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2010.

[5] T. Götz, A. Zirngibl, R. Pekrun, and N. Hall, "Emotions, learning and achievement from an educational-psychological perspective," 2003.

[6] R. Pekrun, "The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice," *Educational Psychology Review*, vol. 18, no. 4, 2006, pp. 315-341.

[7] A. Wigfield, A. Battle, L. B. Keller, and J. S. Eccles, "Sex differences in motivation, self-concept, career aspiration, and career choice: Implications for cognitive development," *Biology, Society, and Behavior: The Development of Sex Differences in Cognition*, vol. 21, 2002, pp. 93-124.

[8] H. G. Ginott, "Teacher & child: A book for parents and teachers," *Avon Books*, 1975.

[9] A.C. Frenzel, T. Goetz, O. Lüdtke, R. Pekrun, and R. E. Sutton, "Emotional transmission in the classroom: exploring the relationship between teacher and student enjoyment," *Journal of Educational Psychology*, vol. 101, no. 3, 2009, p. 705.

[10] E. S. Becker, T. Goetz, V. Morger, and J. Ranellucci, "The importance of teachers' emotions and instructional behavior for their students' emotions — An experience sampling analysis," *Teaching and Teacher Education*, vol. 43, 2014, pp. 15-26.

[11] T. Goetz, O. Lüdtke, U. E. Nett, M. M. Keller, and A. A. Lipnevich, "Characteristics of teaching and students' emotions in the classroom: Investigating differences across domains," *Contemporary Educational Psychology*, vol. 38, no. 4, 2013, pp. 383-394.

[12] J. M. Sun, X. S. Pei, and S. S. Zhou, "Facial emotion recognition in modern distant education system using SVM," in *Proc. 2008 International Conference on Machine Learning and Cybernetics*, vol. 6, 2008.

[13] J. Jeon *et al.*, "A real-time facial expression recognizer using deep neural network," in *Proc. the 10th International Conference on Ubiquitous Information Management and Communication*, ACM, 2016.

[14] L. L. Xu, S. M. Zhang, F. X. Wang, and J. L. Zhao, "CNN expression recognition based on feature graph," *Chinese Automation Congress (CAC)*, IEEE, 2017.

[15] P. Tzirakis, J. Zhang, and B. W. Schuller, "End-to-end speech emotion recognition using deep neural networks," presented at 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018.

[16] M. Likitha, S. R. R. Gupta, K. Hasitha, and A. U. Raju, "Speech based human emotion recognition using MFCC," presented at 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), IEEE, 2017.

[17] A. J. Elliot and T. M. Thrash, "Approach-avoidance motivation in personality: approach and avoidance temperaments and goals," *Journal of Personality and Social Psychology*, vol. 82, no. 5, 2002, p. 804.

[18] L. C. Walsh, J. K. Boehm, and S. Lyubomirsky, "Does happiness promote career success? Revisiting the evidence," *Journal of Career Assessment*, vol. 26, no. 2, 2018, pp. 199-219.

[19] G. K. Suzanne and K. Vosburg, "'Paradoxical'Mood Effects on Creative Problem-solving," *Cognition & Emotion*, vol. 11, no. 2, 1997, pp. 151-170.

[20] A. R. James, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, 1980, p. 1161.

Jingjing Liu was born in Shanghai, China, on Feb. 1, 1995. She received dual degree, namely B.Eng. degree and B.B.A. degree from Fudan University, Shanghai, China, in 2017.

From July.2016 to Dec.2017, she was a co-founder of magic dreamer technology education star-up company. Since Sep.2017, she is currently working for a master's degree in EE at Fudan University. Her main research interests include signal & image processing, artificial intelligence, education evaluation system design and educational application.

Xiaofeng Wu is a senior lecturer of School of Information Science & Engineering and a director of School of Innovation & Entrepreneurship at Fudan University. He received the B. Sc., M. Sc. and PhD. degrees in electronic engineering from Fudan University.

He was a JSPS researcher in both Osaka University and Nara Institute of Science & Technology from 1999 to 2001, and a Postdoctoral researcher in Kyoto University from 2008 to 2009. He was a chief engineer of R&D division in Nirvana Technology Ltd. from 2001 to 2008 and responsible for the Robotics team in 2005 Aichi Expo and 2010 Shanghai Expo. His research interests include signal & image processing, biomedical electronics, multimedia techniques, culture computing, intelligent service robot, artificial intelligence.