Exploring Classroom Emotion with Cloud-Based Facial Recognizer in the Chinese Beginning Class: A Preliminary Study

Chinun Boonroungrut
Multi, Inter or Transdisciplinary program, University of South Africa, Pretoria, South Africa, 63243024@mylife.unisa.ac.za

Toe Toe Oo
Department of Educational Psychology, Sagaing University of Education, Sagaing, Myanmar, toe.oo@yahoo.com

Kim One
School of Foreign language, China University of Geosciences, Wuhan, PR. China, fitgato@gmail.com

The Facial Emotion Recognizer (FER) detection technology in the education field is in the early stage. Objective, to investigate the classroom emotion and the effectiveness of the Microsoft cloud-based FER interpretations The randomly selected 29 international students who enrolled the fundamental Chinese language course were investigated during five study weeks using the paper-based Student Outcome Survey which measured teaching, assessment, generic skill and learning experience. For FER, the 784 students’ facial images were extracted and uploaded into the cloud-based APIs. The students with GPA ≥ 80% showed higher significantly in Assessment components. No significant difference in FER emotion detection was found among five study weeks; however, natural and happiness emotion scores showed symmetric significantly in the opposite direction with $r = -0.82$, $P < 0.1$. The students’ overall outcomes were over average. FER indicated that the neutral emotion was the highest detected score. The APIs Reliability was still considerably questioned. Using updated quality and technology of photo and video recording were recommended in the further research.

Keywords: classroom, emotion, cloud-based, facial recognizer, Chinese, China

---

1 A basic Chinese course for international students, Central China Normal University (华中师范大学), Wuhan, Hubei, P.R. China

INTRODUCTION

Central China Normal University (Huazhong Normal University or CCNU) is one of the top ten normal universities (educational university) which was chosen by the international students around the world ranked by AT8069 and CWUR world university ranking in 2017, and one of the best 100 universities in Asia, 13rd in the nationwide ranked by World University Ranking 2018. The number of international students had increased every academic year. They have planned to study in Chinese program or English program in all majors (CSC, 2016; CWUR, 2017). The university has provided the Chinese foundation courses for ones who do not have any basis Chinese language skill approximately 2 semesters by the college of International Cultural Exchange. Students usually attend these courses before entering their majors taught in Chinese; however, in the English major program, students can register to the basic Chinese courses in the evening after their majors studying. Although most students showed good willing to learn a new language because the Chinese course was not the main major subject requirement, they had to drive their effort to learn Chinese as the new language which looks different entirely in letters, pronunciation including its grammar. So, it is interesting to investigate the emotions and overall satisfaction concerning teaching and learning experiences of those international students during Chinese language class as an extra class of new language during their scholarship program.

LITERATURE REVIEW

Emotion classroom and cloud-based facial emotion recognition APIs

Theoretically, to be successful in teaching and learning, several pieces of researches in this century indicated a positive relationship between emotion in the classroom and the students’ academic outcomes. Emotion and motivation played decision roles to describe components of a positive learning environment. The positive emotion associated with the supportive teaching and learning interaction including showing the teacher and student’s relations (Oo & Boonroungrut, 2017). Furthermore, learning engagement is positively related to classroom emotion; however, it is not an easy for instructors as the classroom moderators to observe students’ emotions in a million way of their expressions during teaching (Meyer & Turner, 2002). Interviewing or contributing any questionnaire may face bias from consideration of instructor’ authority or how the student recognize the previous lesson or events (Flowerdew & Martin, 2008). The evaluation of the face expression investigation from non-verbal communication becomes one of the new solutions. Nowadays, the detecting technology can reach the ease of split the task of emotion recognition (Yang, Alsadoon, Prasad, Singh, & Elchouemi, 2018). Infect, emotion detecting has extended from the facial emotion model since the early 1970s by Ekman and Friesen (Ekman & Friesen, 2003). Their series of studies showed given of intensity of facial emotion that some developers relied on that model to explain the human faces systematically among different groups of people and in various cultures (Groves et al., 2018).
Facial emotion recognition technology

The Facial Emotion Recognition (FER) technology with cloud-based facial emotion recognition application program interfaces (APIs) is the next step of feedback measuring tool in closely real-time social interaction (P Takáč, Sinčák, & Mach, 2016). It is the one of the present human-machine interaction (HMI) (Perveen, Ahmad, Khan, Khalid, & Qadri, 2016). As we know, the facial expression is the effective assessment of non-verbal tasks providing our fundamental understanding of emotional states that show individual’ intention, attitude and mindset (Perveen et al., 2016). Outstandingly, this technology has been utilized in marketing to measure and explore customers’ group satisfaction. Using a machine-based emotion shows benefits when capturing from the single to multiple faces recognition for visual information from camera studies (Cohen, Sebe, Garg, Chen, & Huang, 2003). This technology can perform in real time. It do not disturb while those people are working. It provides the true data from a facial expression that could be real-time collecting, analyzing and reporting. This technology makes a significant difference from taking the conservative paper-based questionnaire. Currently, some studies adapted the emotion recognizer technology in rapidly changing environment situation such as during a focus group discussion that required emotional recognizing of multiple people’s faces expressions (Weber, Cruz Rodriguez, & Mateus, 2016).

The cloud-based facial emotion API recognizer of Microsoft shares similarity of the modified version of Ekman’s emotion model (EEM) that includes two more components namely contempt and neutral (Peter Takáč, Mach, & Sinčák, 2016). This model dimension of emotions attempts to conceptualize human emotions into a multidimensional model (Knutas, Hajikhani, Salminen, Ikonen, & Porras, 2015). One of the psychological emotion circumflex models are distributed in the aspect of two circular areas: arousal and valence dimensions. In this explanation, Russell’s circumflex model can be described as the emotion specification shown in Figure 1. Running the APIs, it recognizes two positive emotions that are indicated by happiness and surprise. Another five expressions with anger, contempt, disgust, fear, and sadness are negative emotions. In this present preliminary study, it aimed to explore the role of cloud-based facial recognizer API in detecting different classroom emotions among the international students who studied in the first foundation Chinese language course in CCNU and addressed any concerned consideration points for adapting this FER technology in the further research.

International Journal of Instruction, January 2019 • Vol. 12, No. 1
Exploring Classroom Emotion with Cloud-Based Facial...

METHOD

Participants and facial images extraction

In this preliminary study, a paper-based survey and facial emotion recognition were applied. Firstly, a survey had been designed to explore the students’ study satisfaction outcomes in one of two Chinese beginning classes. The targeted class had begun in the 2nd semester from March to June in 2017. Twenty-nine international students registered. Note from Table 1; there were 17 students (58.6%) who participated over 80 percent of the course and declared GPA over 80% in their majors. Among them, 52.9% was male, 52.9% was African, 82.4% was under 35 years old, 76.5% was Master and Doctoral students, 58.8% studying in Education, Social Sciences and Psychology, and 88.2% never study the Chinese language before registered in this course.

Secondly, the total 784 facial images were extracted from two hundred and fourteen photos taken from 29 students during the course. All photos were captured randomly in the different times while the students were studying during 5 investigation weeks excepting the orientation week, mid-term and final examination weeks. The photo images were selected equally in the number of each sample per week. All cropped facial images were uploaded into the FER cloud-based database. Over facial expression and wrong facial angle images were screened by the first researcher and confirmed removing by the second author’s agreement before entering images to the APIs. The undetected images were also removed. The emotion estimation values were computed and presented in the provided form of APIs scores.

Measurements

(1) Students’ satisfaction was measured by the Student Outcome Survey. This questionnaire has been annually used for evaluation of the satisfaction of teaching, assessment and generic skill, learning experiences and overall satisfaction in those...
education training programs since 1995. The 20-item questionnaire contains 19 items for 3 components and 1 overall question item. Scores were computed from the 5-Likert scale: strongly disagree to strongly agree. There was a possibility to answer not applicable to all questions that would be counted and reported as missing. (Fieger, 2012).

(2) The Microsoft Emotion APIs is a cloud-based service under the umbrella of AZURE cloud service. It can be accessed from the standard HTTP application or Web APIs technology. The photos were uploaded into the database for its facial detection algorithm analyzing and interpreting. APIs showed 8 emotion labels based on the concept of Ekman’s emotional model; anger (A), contempt (C), disgust (D), fear (F), happiness (H), neutral or nature (N), sadness (SA) and surprise (SU). Each student’s faces were cropped for the better APIs recognizing and evaluating. Emotion APIs indicated scores ranged between 0.0000 – 1.0000.

Photo caption tools and data analysis
The camera quality was one of the current research concerns because the facial detecting technology-based images require enough pixel’s quality and the proper face position & degree to make verification in acquiring a face database. For the solution, two different cameras were used as the following: (1) ZTE Nubia UINX511J uses the 16-megapixel with its Sony IMX234 sensor and fast f/2.0 aperture. This camera was used to capture the photos randomly from the second week to the fourth week, and (2) Ricoh Theta S 360° degree camera uses F2.0 lens F value and ½.3 CMOS (X2) image sensor size. Its active pixel is approximately 12 megapixels with L:5376 x 2688, M:2048 x 1024. Object distance is approx. 10cm –∞(from the front of the lens). This camera captured the photos in the last week at the same time when the paper-based questionnaires were provided.

To verify the performance of the APIs, we compare the levels of emotions from APIs 4-digit decimal values with rounding off cut point at lower than 5 in the last digit. After removing some potential data specify problems including outliners, T-test, one-way, two-way ANOVA and Pearson’s correlation coefficient were applied. The scores from the survey and APIs were performed in SPSS mac OS.

Figure 2 - 3
Examples of the quality and angle of the Photos taken by; ZTE Nubia UINX511J (left) cameras and when Analysed by API (right)
FINDINGS

Preliminary results

In the preliminary study, a common method variance and a single common latent factor were run for each measure independently (Podsakoff, MacKenzie, & Podsakoff, 2012; Srichom & Lawthong, 2015). It revealed that those forms of bias influenced no score, and then we processed the next analysis. The results indicated that there were no significant differences in gender, age, race, education level, major and ones who had some basic Chinese background; however, students who got GPA from their major higher than 80% \((M = 4.25, SD = .53)\) and lower than 80% \((M = 2.90, SD = 1.27)\) differed significantly in Assessment subscale mean scores with condition \(t(15) = -2.948, p < 0.01\). Considerably, although several studies showed significance in gender differences, this present study did not find any significant differences in those 3 components. Continually, we explored the effect of GPA and gender differences on those components. The results of a factorial ANOVA indicated that all effects were significant as shown in Table 1. The GPA main effect indicated a significant difference between students’ GPA < 80% \((M = 2.90, SD = 1.27)\) and students’ GPA ≥ 80% \((M = 4.25, SD = .53)\). Genders yielded a significant difference between male \((M = 4.32, SD = .42)\) and female \((M = 3.88, SD = .91)\). The interaction effect was significant with condition \(F(1,63) = 3.628, P < 0.05\), partial \(\eta^2 = .218\).

Remarkably, the internal consistency correlation showed higher over acceptable levels. Assessment, and overall showed the significantly strongest correlation at \(r = .80, P < .01\). Moreover, Cronbach’s Alpha for satisfaction of teaching, assessment and generic skill were also highly acceptable at .97, .88 and .87 respectively as shown in Table 2.

Table 1
A 2-way ANOVA results for GPA and gender differences on Assessment

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F-Value</th>
<th>Partial (\eta^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>60.484</td>
<td>1</td>
<td>90.484</td>
<td>320.08**</td>
<td>.961</td>
</tr>
<tr>
<td>GPA</td>
<td>3.274</td>
<td>1</td>
<td>3.274</td>
<td>11.583**</td>
<td>.471</td>
</tr>
<tr>
<td>Genders</td>
<td>1.899</td>
<td>1</td>
<td>1.899</td>
<td>6.716*</td>
<td>.341</td>
</tr>
<tr>
<td>GPA x Genders</td>
<td>1.025</td>
<td>1</td>
<td>1.025</td>
<td>3.628*</td>
<td>.218</td>
</tr>
<tr>
<td>Error</td>
<td>3.675</td>
<td>13</td>
<td>.283</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: *; \(P < .05\), **; \(P < .01\)

Table 2
Inter-correlation and Reliability among 3 Sub-Scales of the Student Outcome Survey

<table>
<thead>
<tr>
<th></th>
<th>Teaching</th>
<th>Assessment</th>
<th>Generic Skill</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfaction of Teaching</td>
<td>(\alpha = .975)</td>
<td>(\alpha = .889)</td>
<td></td>
</tr>
<tr>
<td>Assessment</td>
<td>.644**</td>
<td>.647**</td>
<td>(\alpha = .877)</td>
</tr>
<tr>
<td>Generic Skill</td>
<td>.768**</td>
<td>.683**</td>
<td>.694**</td>
</tr>
<tr>
<td>Overall</td>
<td>.800**</td>
<td>.800**</td>
<td>.694**</td>
</tr>
</tbody>
</table>

Note: **; \(P < .01\) (2-tailed)
Facial emotion detection APIs results

Based on APIs scores from 784 images, the emotional main effect for each factor of emotions among the weeks was not significant, and no any detected emotion mean score shows difference in the reported amounts as shown in Table 4. Nonetheless, the scores on natural emotion were higher significantly in male (M=.857, SD=.25) than female (M=.720, SD=.35), t(96)=2.293, P =.024 (see mean differences between genders in Figure 5; axis x and y represented weeks and emotion APIs mean scores). Another emotion did not show differences at a significant level. After plotting a graph, natural and happiness emotion seemed perfectly symmetrical. Interestingly, we found that they were strongly negatively correlated significantly, r = -0.82, P <.001. Besides, angry showed the positive correlation with contempt and disgust significantly, r = .515 and .351, P < .001 respectively.

Table 4
Mean scores and (S.D.) Detected from 784 Photo Images based Emotion APIs

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Wk.1</th>
<th>Wk.2</th>
<th>Wk.3</th>
<th>Wk.4</th>
<th>Wk.5</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>0.054</td>
<td>0.020</td>
<td>0.050</td>
<td>0.033</td>
<td>0.011</td>
<td>0.715</td>
<td>0.584</td>
</tr>
<tr>
<td>Contempt</td>
<td>0.019</td>
<td>0.033</td>
<td>0.077</td>
<td>0.067</td>
<td>0.024</td>
<td>0.868</td>
<td>0.486</td>
</tr>
<tr>
<td>Disgust</td>
<td>0.004</td>
<td>0.019</td>
<td>0.017</td>
<td>0.035</td>
<td>0.023</td>
<td>0.512</td>
<td>0.727</td>
</tr>
<tr>
<td>Fear</td>
<td>0.009</td>
<td>0.006</td>
<td>0.005</td>
<td>0.012</td>
<td>0.008</td>
<td>0.331</td>
<td>0.586</td>
</tr>
<tr>
<td>Happiness</td>
<td>0.071</td>
<td>0.271</td>
<td>0.099</td>
<td>0.104</td>
<td>0.137</td>
<td>1.719</td>
<td>0.52</td>
</tr>
<tr>
<td>Natural</td>
<td>0.838</td>
<td>0.642</td>
<td>0.822</td>
<td>0.850</td>
<td>0.820</td>
<td>1.493</td>
<td>0.21</td>
</tr>
<tr>
<td>Sadness</td>
<td>0.011</td>
<td>0.021</td>
<td>0.007</td>
<td>0.019</td>
<td>0.029</td>
<td>1.384</td>
<td>0.245</td>
</tr>
<tr>
<td>Surprise</td>
<td>0.010</td>
<td>0.024</td>
<td>0.007</td>
<td>0.019</td>
<td>0.019</td>
<td>1.094</td>
<td>0.364</td>
</tr>
</tbody>
</table>

Note: Wk.; week

Table 5
Inter-correlation and Overall Means and SD among 8 APIs Emotional Scores

<table>
<thead>
<tr>
<th>Emotion</th>
<th>C</th>
<th>D</th>
<th>F</th>
<th>H</th>
<th>N</th>
<th>SD</th>
<th>SU</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>.515**</td>
<td>.351**</td>
<td>-.02</td>
<td>-.065</td>
<td>-.086</td>
<td>.044</td>
<td>.047</td>
<td>.0035</td>
<td>.0007</td>
</tr>
<tr>
<td>Contempt</td>
<td>.112</td>
<td>.029</td>
<td>.128</td>
<td>-.181</td>
<td>-.307</td>
<td>-.014</td>
<td>.0127</td>
<td>.0024</td>
<td></td>
</tr>
<tr>
<td>Disgust</td>
<td>.360**</td>
<td>.048</td>
<td>.048</td>
<td>-.101</td>
<td>.201*</td>
<td>-.099</td>
<td>.0019</td>
<td>.0007</td>
<td></td>
</tr>
<tr>
<td>Fear</td>
<td>.556</td>
<td>.119</td>
<td>.090</td>
<td>.69**</td>
<td>.008</td>
<td>.0002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Happiness</td>
<td>-.821**</td>
<td>-.099</td>
<td>-.075</td>
<td>.1293</td>
<td>.0260</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>-.035</td>
<td>.020</td>
<td>.8002</td>
<td>.2982</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sadness</td>
<td>-.053</td>
<td>.0175</td>
<td>.0376</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surprise</td>
<td>.0071</td>
<td>.0018</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Emo.; emotion, A; anger, C; contempt, D; disgust, F; fear, H; happiness, N; neutral, SA; sadness and SU; surprise, *, p < .05, **; p < .01
DISCUSSION

According to the present results, the student’s classroom outcome as the baseline showed strong and positive among the students who had the good academic achievement. The results indicated the same direction in students who got GPA higher than 80% with higher in classroom outcome. Curiously, Chinese classrooms in CCNU were recognized as a traditional teaching method with integration some teaching and communicating technologies. It could share a significant effect on classroom outcomes and also satisfied emotion. (El-Banna, Whitley, & McNelis, 2017; Missildine, Fountain, Summers, & Gosselin, 2013). With no surprise, this study showed high scores for all 3 components, especially in satisfaction of teaching sub-scale. Some education studies pointed out that teaching expertise and teachers’ goal orientation were the main indications of students’ classroom outcomes and satisfaction (Butt & ur Rehman, 2010; Wang, Hall, Goetz, & Frenzel, 2017) Interestingly, the communication between the teacher and students through the online platform after the class cordially and intimately was found. This factor could support why the students were positive in teaching sub-scale (Jaasma & Koper, 1999).

This preliminary study indicated various concerning points for the further research that is planned to apply the cloud-based facial emotion recognition technology. Absolutely, the challenge of facial extracted evaluation should be a good quality by means of dimensions, granulation, and lighting condition (Sariyanidi, Gunes, & Cavallaro, 2015) Directionally, the real situation cannot be controlled like in the laboratory. Some photos images could not be detected, and they had to remove from the analysis (see example in Figure 3 at least 6 students were not detected). Based on this result, the current facial recognition technology, the other known factors that, the better FER would be...
performed, the most of significant which were (1) Environment; head size, background, distance, lighting (2) Image Age; time elapsed including photo quality (3) Consistency Camera; optical characteristic and enrolment process (4) Gallery Size; the number of image enter to the gallery database (Introna & Nissenbaum, 2010). For a summary, the detecting technology associated with quality of photos, camera and database. Latterly, the development of image capturing technology has been rapidly improving, so choosing the camera with any photo capturing tool should be strongly concerned.

Human shows more than one modality to express emotions. It was expected that the performance of automatic multimodal systems that could be higher than automatic unimodal systems (Busso et al., 2004). From the APIs, natural and happiness emotion seemed against each other, and they showed a negative relationship higher than another emotion pairs. Anger in this study also stipulated no negative correlation range to sadness, happiness and neutral. Generally in FER coding, happiness and sadness were easier in detecting than those of fear and disgust (Smith, Cottrell, Gosselin, & Schyns, 2005), and anger is less with a chance to similar with happiness and sadness in detection (Durand, Gallay, Seigneuric, Robichon, & Baudouin, 2007).

LIMITATIONS
In this preliminary study, one of the limitations was our difficulty in using the automatic capturing cameras. Photos taken by the human or by the automation presented different outcomes. Many photos seemed overexpressed facial emotion especially in happiness with participants’ cheese smiling that caused misinterpretation from APIs. Moreover, the angle of camera position and capacity could not cover a whole classroom area, so some students were hidden by another. Also, the number of excluded participants was higher than our expectation, out of our control in the real classroom situation. In generalizing the results from the Student Outcomes Survey, absent and withdraw students (< 80% class participation) were excluded in the final stage; however, they were considered as a possibility to be a survival bias from the positive side. Lastly, a cross-sectional design limitation measured the only outcome at the end of the class. In detail, the use of the small group of participants was the present limitation. It limited the power of the test due to the maximum number of students in the selected classroom. Of note, there was a chance that those students who stopped learning Chinese before the end of the class were unsatisfied, so survivorship bias could be clarified and considered. Although the results of both student outcomes and APIs showed similar with several studies, these preliminary results should be implied with a caution that could affect to student’s emotion mean scores from other factors as the mediation effect. In addition, the APIs reliability is still questioned. It is not worth to address this gap as the further basic research.

CONCLUSION AND RECOMMENDATION
Camera quality should be mainly concerned. Consulting the photographer or audio-visual technicians for the best solution for the classroom setting is recommended. Video recording could be replaced to analyze the impact effect of time in the studying and to see how emotion changed in real time. Furthermore, the reliability of FER in emotion
APIs was still being a gap of knowledge. Analyzing with other statistic methods was required for finding more empirical findings.

In conclusion, the mean score differences between students who got GPA higher and lower than 80% were found at significantly in the assessment sub-scale. Eight emotion types did not show any difference in each week of studying as the main effect in this preliminary study; however, natural emotion showed the highest mean score and the easiest to be detected. Moreover, it seemed symmetric against happiness that showed strongly negative correlation with each other. FER detection from video recording was recommended for the further study instant of only photo image capturing.

ACKNOWLEDGEMENT
This preliminary study was supported voluntarily by Wang Tao, the class visiting instructor and Agril Musariz, the audio-visual technician.

CONFLICT OF INTEREST
This is an independent study, and no conflict of Interest was declared.

REFERENCES


