CASME Database: A Dataset of Spontaneous Micro-Expressions Collected From Neutralized Faces

Wen-Jing Yan, Qi Wu, Yong-Jin Liu, Su-Jing Wang and Xiaolan Fu*

Abstract— Micro-expressions are facial expressions which are fleeting and reveal genuine emotions that people try to conceal. These are important clues for detecting lies and dangerous behaviors and therefore have potential applications in various fields such as the clinical field and national security. However, recognition through the naked eye is very difficult. Therefore, researchers in the field of computer vision have tried to develop micro-expression detection and recognition algorithms but lack spontaneous micro-expression databases. In this study, we attempted to create a database of spontaneous micro-expressions which were elicited from neutralized faces. Based on previous psychological studies, we designed an effective procedure in lab situations to elicit spontaneous micro-expressions and analyzed the video data with care to offer valid and reliable codings. From 1500 elicited facial movements filmed under 60fps, 195 micro-expressions were selected. These samples were coded so that the first, peak and last frames were tagged. Action units (AUs) were marked to give an objective and accurate description of the facial movements. Emotions were labeled based on psychological studies and participants' self-report to enhance the validity.

I. INTRODUCTION

Micro-expression is a fast and brief facial expression which appears when people try to conceal their genuine emotions, especially in high-stake situations [1][2]. Haggard and Isaacs first discovered micro-expression (micro-momentary expression) and considered it as repressed emotions [3][4]. In 1969, Ekman analyzed a interviewing video of a patient stricken with depression who tried to commit suicide and found micro-expressions. From then on, several researches have been conducted in the field of micro-expression but few results were published. Micro-expression has gained popularity recently because of its potential applications in the process of diagnosis and national security. It is considered as one of the most effective clues to detect lies and dangerous behaviors [2]. The Transportation Security Administration in the USA has already employed Screening Passengers by Observation Techniques (SPOT), which was largely based on the findings from micro-expression studies [5]. In the clinical field, micro-expression may be used to understand genuine emotions of the patients and promote better therapies. However, micro-expression is considered so fleeting that it is almost undetectable and thus difficult for human beings to detect [2]. Matsumoto defined any facial expressions shorter than 500 ms as micro-expressions [6], which are much faster than conventional facial expressions and easily neglected. To better apply micro-expressions in detecting lies and dangerous behaviors, efficient micro-expression recognition system should be incorporated to greatly reduce the amount of work and time needed. Therefore, many researchers have tried to develop an automatic micro-expression recognition system to help people detect such fleeting facial expressions [7][8][9][10].

There are many facial expression databases [11] but micro-expression databases are rare. The following are the few micro-expression databases used in developing detection and recognition algorithms:

- USF-HD contains 100 micro-expressions, with the resolution of 720×1280 and frame-rate of 29.7 fps. Participants were asked to perform both macro- and micro-expressions. For micro-expressions, participants were shown some example videos containing micro-expressions prior to being recorded. The participant was then asked to mimic them [9].
- Polikovsky's database contains 10 university student subjects, who were instructed to perform 7 basic emotions with low facial muscles intensity and to go back to the neutral face expression as fast as possible, simulating the micro-expression motion. Camera settings are: 480×640 resolution, 200fps [12].
- YorkDDT contains 18 micro-expressions: 7 from emotional and 11 from non-emotional scenarios; 11 from deceptive and 7 from truthful scenarios. Micro-expressions were found in 9 participants (3 male and 6 female) [10][13].
- SMIC contains 77 spontaneous micro-expressions which were recorded by a 100fps camera. An interrogation room setting with a punishment threat and highly emotional clips were chosen to create a high-stake situation where participants undergoing high emotional arousal are motivated to suppress their facial expressions [10].

Previous micro-expression databases include some of the following problems:

- Unnatural micro-expressions. Some of those are created

Wen-Jing Yan, Su-Jing Wang and Xiaolan Fu are with State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, China. fuxl@psych.ac.cn

Qi Wu is with Department of Psychology, Hunan Normal University, 410000, China.

Yong-Jin Liu is with the TNList, Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, China

intentionally so they are different from the sponta- neous micro-expressions. According to Ekman, micro- expression cannot be intentionally controlled [1].

- Facial movements without emotions involved. Without careful analysis and research, it is easy to confuse unemotional facial movements as micro-expressions, such as blowing the nose, swallowing saliva or rolling eyes.
- Lack of precise emotion labeling. From a psychological perspective, some of these databases do not have correct emotion labels. Emotion labeling for micro-expressions is similar but not the same as that for conventional facial expressions.

As a result, we developed a database of micro-expressions to aid these researchers in their training and evaluation processes. The approaches of eliciting and analyzing micro-expressions were based on psychological studies. Here we provide a relatively effective and efficient way to create a spontaneous micro-expression database that includes the following advantages:

(1) The samples are spontaneous and dynamic micro-expressions. Before and after each micro-expression are baseline (usually neutral) faces, so the samples can also be used to evaluate detection algorithm.

(2) Participants were asked maintain a neutral face (neutralization paradigm) in the study. Therefore, micro-expressions captured in our database are relatively "pure and clear", without noises such as head movements and irrelevant facial movements.

(3) Action units were given for each micro-expression. AUs give detailed movements of facial expressions and help to give more accurate emotion labels [14][15].

(4) Two different cameras under different environmental configurations were used to increase the visual variability.

(5) The emotions that occur were carefully labeled based on psychological researches and participants' self-report. In addition, the unemotional facial movements were removed.

II. THE CASME DATABASE

The Chinese Academy of Sciences Micro-expression (CASME) database contains 195 micro-expressions filmed under 60fps. They were selected from more than 1500 elicited facial movements. These samples were coded with the onset, apex and offset frames[1], with action units (AUs) marked and emotions labeled. 35 participants (13 females, 22 males) were recruited with a mean age of 22.03 years (SD=1.60) in the study. All provided informed consent.

[1]The onset frame was the first one which changes from the baseline (usually neutral facial expressions). The apex-1 frame is the first one that reached highest intensity of the facial expression and if it keeps for a certain time, the apex-2 frame is coded.

TABLE I

DESCRIPTIVE STATISTICS FOR TOTAL DURATION AND ONSET DURATION OF MICRO-EXPRESSIONS IN CLASS A.

		Total duration		Onset duration	
	N	M(ms)	SD	M(ms)	SD
500 ms as the upper limit	83	289.96	82.72	130.32	51.48
All the samples*	100	/	/	142.17	51.47

*The fast-onset facial expressions (onset duration no more than 250 though total duration longer than 500 ms) were added.

TABLE II

DESCRIPTIVE STATISTICS FOR TOTAL DURATION AND ONSET DURATION OF MICRO-EXPRESSIONS IN CLASS B.

		Total duration		Onset duration	
	N	M(ms)	SD	M(ms)	SD
500 ms as the upper limit	65	299.24	81.93	123.99	47.42
All the samples*	95	/	/	137.37	51.47

*The fast-onset facial expressions (onset duration no more than 250 though total duration longer than 500 ms) were added.

Micro-expressions with the duration no more than 500 ms were selected for the database. In addition, facial expressions that last more than 500 ms but their onset duration less than 250 ms were also selected because fast-onset facial expressions is fundamentally characterized with micro-expressions as well. We recorded the facial expressions with two different environmental configurations and two different cameras. Therefore, we divide the samples into two classes: Class A and Class B.

A. Class A

The samples in Class A were recorded by BenQ M31 camera with 60fps, with the resolution set to 1280×720 pixels. The participants were recorded in natural light. The steps of the data analysis is in section III ACQUISITION AND CODING. Table I shows the basic information for the samples and Figure 1 shows an example.

B. Class B

The samples in Class B were recorded by Point Grey GRAS-03K2C camera with 60fps, with the resolution set to 640×480 pixels. The participants were recorded in a room with two LED lights. The steps of the data analysis were the same as those in Class A.

We selected 95 samples which last no more than 500 ms and another 30 samples with the onset phase[2] no more than 250 ms for this section(see Table II).

C. Distribution Fitting of the duration of micro-expressions

We used models of Normal, Gamma, Weibull and Birnbaum-Saunders to fit a curve to the duration of the micro-expressions and provided the distribution curves (see Figure II-C and Figure 3). By obtaining Akaike's information criterion (AIC) [16], Birnbaum-Saunders model best fits the total duration and Gamma model best fits the onset duration (Table III).
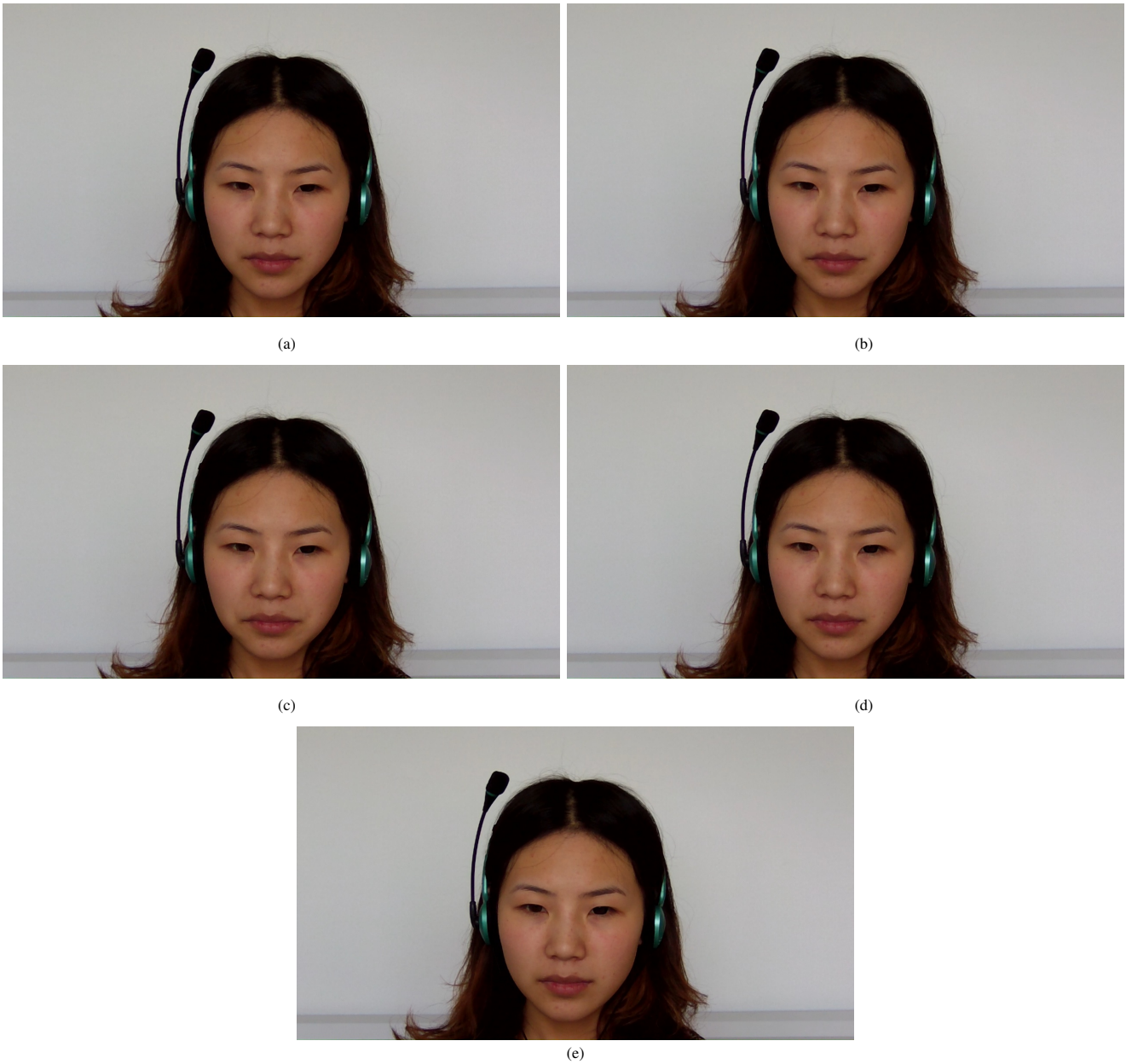
[2]the duration from the onset frame to apex frame

Fig. 1. An example of part of the frame sequence from Class A, including onset frame (a), apex frame (c) and offset frame (e).The AU for this micro-expression is 15, which is lip corner depressor. The movement is more obvious in video play (see supplementary material) but not in a picture sequence.

TABLE III
THE RESULTS OF KOLMOGOROV-SMIRNOV TEST AND AIC UNDER DIFFERENT FITTING MODELS.

Model	Total duration		Onset duration	
	LL	AIC	LL	AIC
Normal	-862.57	1729.1	-1058.8	2121.6
Gamma	-858.53	1721.1	-1048.74	2101.5*
Weibull	-863.5	1731	-1053.14	2110.3
Birnbaum-Saunders	-858.49	1721*	-1048.88	2101.8

*indicates the best choice in the test.

D. Action units and emotions

The action units (AUs) for every micro-expression were given (Table IV). Two coders coded independently then they arbitrated any disagreements. The reliability between the two coders is 0.83 [14]. The criteria for labeling the emotions were mainly based on Ekman's study[14]. Considering that the elicited micro-expressions in our study are mainly partial and with low intensity, we have to take into account participants' self-rating and the content of the video episodes when labeling the emotions. Besides the basic emotions, we also provide *repression* and *tense* because the six basic emotions do not cover all the configurations of AUs.

E. Baseline evaluation

These three dimensional data are easily expressed as 3rd-order tensor [17] mathematically. Thus, we used Multilinear
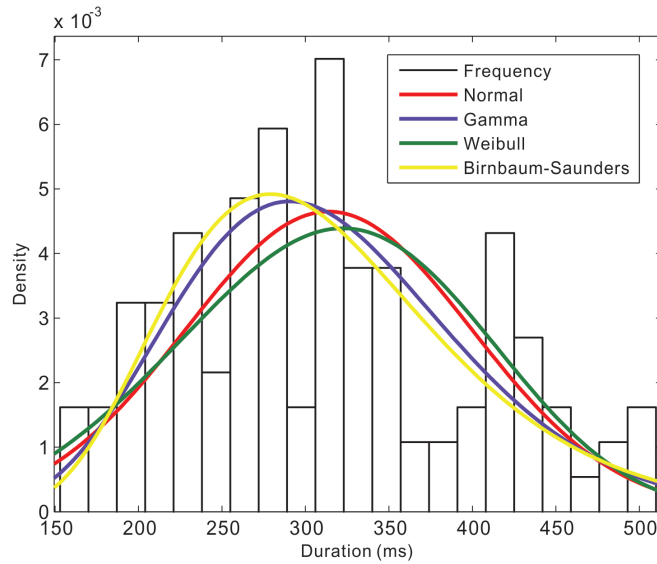
Fig. 2. The distribution fitting curves of four different models for the total duration of micro-expressions.
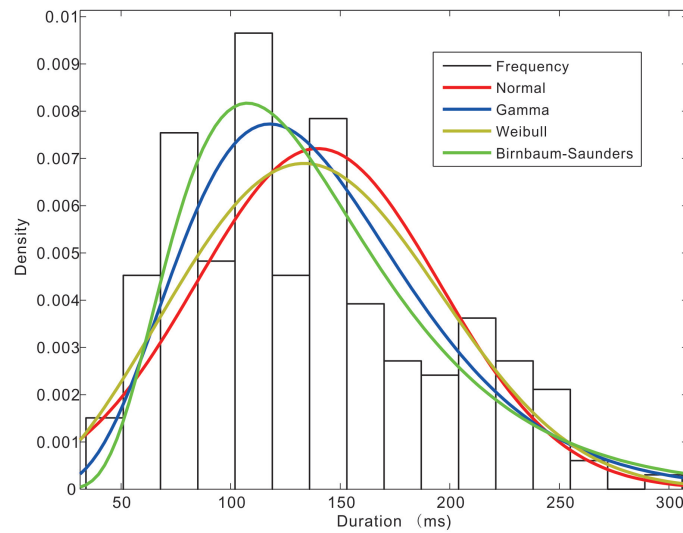

Fig. 3. The distribution fitting curves of four different models for the onset duration of micro-expressions.

TABLE IV

CRITERIA FOR LABELING THE EMOTIONS AND THE FREQUENCY IN THE DATABASE*.

Emotion	Criteria	N
Amusement	Either AU6 or AU12 must be present	5
Sadness	AU1 must be present	6
Disgust	At least one of AU9, AU10, AU4 must be present	88
Surprise	AU1+2, AU25 or AU2 must be present	20
Contempt	Either unilateral AU10 or unilateral AU12 be present	3
Fear	Either AU1+2+4 or AU20 must be present	2
Repression	AU14, AU15 or AU17 is presented alone or in combination	40
Tense	Other emotion-related facial movements	28

*The emotion labeling are just partly based on the AUs because micro-expressions are usually partial and in low intensity. Therefore, we also take account of participants' self-report and the content of the video episodes.

TABLE V
EXPERIMENTAL RESULTS ON MICRO-EXPRESSION DATABASE WITH MPCA.

	$G3$	$G6$	$G9$	$G12$	$G15$
$10 \times 10 \times 10$	0.3248	0.3370	0.3789	0.3784	0.4081
$20 \times 20 \times 20$	0.3286	0.3500	0.3801	0.3896	0.4101
$30 \times 30 \times 30$	0.3286	0.3500	0.3801	0.3896	0.4101
$40 \times 40 \times 40$	0.3282	0.3504	0.3817	0.3851	0.4035
$50 \times 50 \times 60$	0.3293	0.3511	0.3805	0.3851	0.4061
$60 \times 60 \times 60$	0.3293	0.3519	0.3813	0.3851	0.4061

Principal Component Analysis (MPCA) [18] as the baseline. From the CASME database, we selected emotions of *disgust*, *repression*, *surprise*, and *tense*. The micro-expression video set is partitioned into different galleries and probe sets. In this paper, Gm indicates that m samples per micro-expression are randomly selected for training and the remaining samples are used for testing. For each partition, we used 20 random splits for cross-validation tests. All samples are manually cropped and resized to $64 \times 64 \times 64$ pixels.

For Baseline evaluation, we conducted MPCA on the database. The convergence threshold η is set as 0.1. The optimal dimensionality are $10 \times 10 \times 10$, $20 \times 20 \times 20$, $30 \times 30 \times 30$, $40 \times 40 \times 40$, $50 \times 50 \times 50$ and $60 \times 60 \times 60$. Table V shows mean performances on these optimal dimensionality.

III. ACQUISITION AND CODING

In order to elicit "noiseless" micro-expressions, we employed the neutralization paradigm in which participants try to keep their faces neutralized when experiencing emotions. We used video episodes as the eliciting material with contents that are considered high in emotional valence. In this study, the participants experienced high arousal and strong motivation to disguise their true emotions.

A. Elicitation materials

We used video episodes with high emotional valence as the elicitation material. 17 video episodes were downloaded from the Internet, which was assumed to be highly positive or negative in valence and may elicit various emotions from the participants. The durations of the selected episodes ranged from about 1 minute to roughly 4 minutes. Each episode mainly elicited one type of emotion. 20 participants rated the main emotions of the video episodes and scores from 0 to 6 were given to each, where 0 is the weakest and 6 the strongest (see Table VI).

B. Elicitation procedure

To enhance the participants motivation of concealing their emotions, participants were firstly instructed that the purpose of the experiment was to test their ability to control emotions, which was highly related to their social success. The participants were also told that their payment was directly related to their performance. If they showed any facial expressions during the study, 5 Chinese Yuan (RMB) was deducted from

TABLE VI
THE FOLLOWING ARE THE PARTICIPANTS'S RATINGS ON THE 17 VIDEO EPISODES, THE MAIN EMOTION FOR EACH EPISODE, THE NUMBER OF PARTICIPANTS WHO FEEL SUCH AN EMOTION AND THEIR CORRESPONDING MEAN SCORE (FROM 0 TO 6).

Episode NO.	Main emotions	Rate of selection	Mean score
1	amusement	0.69	3.27
2	amusement	0.71	3.6
3	amusement	0.7	3.14
4	amusement	0.64	4.43
5	disgust	0.81	4.15
6	disgust	0.69	4.18
7	disgust	0.78	4
8	disgust	0.81	3.23
9	fear	0.63	2.9
10	fear	0.67	2.83
11*	/	/	/
12	disgust (fear)	0.60(0.33)	3.78(0.28)
13	sadness	0.71	4.08
14	sadness	1	5
15	anger (sadness)	0.69(0.61)	4.33(0.62)
16	anger	0.75	4.67
17	anger	0.94	4.93

*No single emotion word was selected by one third of the participants.

the payment as a punishment each time (though we actually offered similar payments for all the participants in the end). In addition, they were not allowed to turn their eyes or head away from the screen.

Each participant was seated in front of the 19-inch monitor. The camera (Point grey GRAS-03K2C or BenQ M31, with 60 frames per second) on a tripod was set behind the monitor to record the full-frontal face of the participants. The video episodes were presented by a computer which was controlled by the experimenter. The participants were told to closely watch the screen and maintain a neutral face.

After each episode was over, the participants were asked to watch their own facial movements in the recordings and indicate whether they produced irrelevant facial movements which could be excluded for later analysis.

C. Coding process

Two well-trained coders thoroughly inspected the recordings and selected the fast facial expressions. Afterwards, they independently spot the onset, apex and offset frames, and arbitrated the disagreement. The reliability (the agreement of frames) between the two coders is 0.78[19]. When they didn't agree on the location, the average of the two coders' numbers was taken. They processed the video recordings in the following steps:

Step 1. The first step is a rough selection. This procedure was to reduce the quantity of to-be-analyzed facial movements while not missing any target. The coders played the recordings at half speed and roughly spot the onset, apex and offset frames and then selected the facial expressions that last less than 1 s. It was also noticed that some of the leaked fast facial expressions in our study were characterized as fast onset with a slow offset. Thus, fast-onset facial expressions with the onset phases less than 500ms (though the total

duration is longer than 1 s) were selected for later analysis because of their special temporal features;

Step 2. The selected samples were then converted into pictures (one picture extracted from every two frames);

Step 3. Habitual movements (such as blowing the nose) or the movements caused by other irrelevant movements (such as pressing the lips when swallowing saliva, dilating nose because of inhaling, moving eyebrows because of the eyesight change) were removed. These irrelevant facial movements were confirmed by the participants after the experiments.

Step 4. By employing frame-by-frame approach, the coder stayed half a meter away from before a 17-inch monitor to spot the onset frames, apex frames and offset frames. Sometimes the facial expressions faded very slowly, and the changes between frames were very difficult to detect by eyes. For such offset frames, the coders only coded the last obvious changing frame as the offset frame while ignoring the nearly imperceptible changing frames.

IV. DISCUSSION AND CONCLUSION

A. *The intensity of the micro-expressions.*

Since the participants were trying to neutralize their facial expressions, the repression is strong. Thus the elicited facial expressions in the dataset are not low in intensity. They are not only fast but also subtle. Frame-by-frame scrutiny is usually more difficult than real-time observation to spot the micro-expressions. In another word, movement information is important in recognizing such type of micro-expressions.

B. *Criteria for labeling emotions.*

Unlike conventional facial expressions, micro-expressions in this database usually appear partially (either upper face or lower face). Moreover, these micro-expressions are low in intensity therefore the criteria for labeling emotions is somewhat different from conventional facial expressions. Though the criteria for labeling emotions are mainly based on Ekman's criteria[14], we still take participants's reports into account. For example, AU 14 and AU 17 were considered as repression. For facial expressions with no definite emotions but seemed tense, we defined them as *tense*.

C. *Fast-onset facial expressions.*

Due to the paradigm we used in eliciting micro-expressions, some of the facial expressions have fast onset but slow offset. These facial expressions, share the fundamental characteristics of micro-expressions, being involuntary, fast, and also revealing the genuine emotions that participants tried to conceal. Therefore, we include these samples into the database as well.

D. *Future work and availability*

The database is small for the moment. We are coding the remaining video recordings to create more samples. Because elicitation of micro-expressions is not easy and coding is time-consuming, this database can only be enlarged bit by bit. We will try to improve the elicitation approach and find more participants to enrich this database.

The full database file is available upon request to the corresponding author.

In summary, we try to provide a satisfying spontaneous micro-expression database for researchers to develop an micro-expression recognition algorithm. Based on the previous psychological studies on micro-expression, we improved the approaches of elicitation and data analysis. We removed the unemotional facial movements and make sure the selected micro-expressions are genuine ones. With multiple efforts, we provide a micro-expression database with validity and reliability, hoping that our work will help to develop an efficient micro-expression recognition system.

REFERENCES

[1] P. Ekman and W. Friesen, "Nonverbal leakage and clues to deception," DTIC Document, Tech. Rep., 1969.

[2] P. Ekman, "Lie catching and microexpressions," *The philosophy of deception*, pp. 118–133, 2009.

[3] E. A. Haggard and K. S. Isaacs, *Methods of Research in Psychotherapy*. New York: Appleton-Century-Crofts, 1966, ch. Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy, pp. 154–165.

[4] P. Ekman, "Darwin, deception, and facial expression," *Annals of the New York Academy of Sciences*, vol. 1000, no. 1, pp. 205–221, 2006.

[5] S. Weinberger, "Airport security: Intent to deceive," *Nature*, vol. 465, no. 7297, pp. 412–415, 2010.

[6] D. Matsumoto and H. Hwang, "Evidence for training the ability to read microexpressions of emotion," *Motivation and Emotion*, vol. 35, no. 2, pp. 181–191, 2011.

[7] Q. Wu, X. Shen, and X. Fu, "The machine knows what you are hiding: an automatic micro-expression recognition system," *Affective Computing and Intelligent Interaction*, pp. 152–162, 2011.

[8] M. Shreve, S. Godavarthy, V. Manohar, D. Goldgof, and S. Sarkar, "Towards macro-and micro-expression spotting in video using strain patterns," in *Applications of Computer Vision (WACV), 2009 Workshop on*. IEEE, 2009, pp. 1–6.

[9] M. Shreve, S. Godavarthy, D. Goldgof, and S. Sarkar, "Macro-and micro-expression spotting in long videos using spatio-temporal strain," in *IEEE Conference on Automatic Face and Gesture Recognition FG'11*. IEEE, 2011, pp. 51–56.

[10] T. Pfister, X. Li, G. Zhao, and M. Pietikainen, "Recognising spontaneous facial micro-expressions," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1449–1456.

[11] C. Anitha, M. Venkatesha, and B. Adiga, "A survey on facial expression databases," *International Journal of Engineering Science and Technology*, vol. 2, no. 10, pp. 5158–5174, 2010.

[12] S. Polikovsky, Y. Kameda, and Y. Ohta, "Facial micro-expressions recognition using high speed camera and 3d-gradient descriptor," in *Crime Detection and Prevention (ICDP 2009), 3rd International Conference on*. IET, 2009, pp. 1–6.

[13] G. Warren, E. Schertler, and P. Bull, "Detecting deception from emotional and unemotional cues," *Journal of Nonverbal Behavior*, vol. 33, no. 1, pp. 59–69, 2009.

[14] P. Ekman, W. Friesen, and J. Hager, "FACS investigators guide," *A human face*, 2002.

[15] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Automatic recognition of facial actions in spontaneous expressions," *Journal of Multimedia*, vol. 1, no. 6, pp. 22–35, 2006.

[16] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[17] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *Siam Review*, vol. 51, no. 3, pp. 455–500, 2009.

[18] H. P. Lu, N. P. Konstantinos, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Transactions on Neural Networks*, vol. 19, no. 1, pp. 18–39, 2008.

[19] S. Porter and L. ten Brinke, "Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions," *Psychological Science*, vol. 19, no. 5, pp. 508–514, 2008.