RECOGNIZING MICROEXPRESSION: AN INTERDISCIPLINARY PERSPECTIVE

EDITED BY: Xunbing Shen, Wenfeng Chen, Guoying Zhao and Ping Hu

PUBLISHED IN: Frontiers in Psychology

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88963-137-7
DOI 10.3389/978-2-88963-137-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

RECOGNIZING MICROEXPRESSION: AN INTERDISCIPLINARY PERSPECTIVE

Topic Editors:
Xunbing Shen, Jiangxi University of Traditional Chinese Medicine, China
Wenfeng Chen, Renmin University of China, China
Guoying Zhao, University of Oulu, Finland
Ping Hu, Renmin University of China, China

As a Chinese saying goes, “Look at the weather when you step out; look at men’s faces when you step in.” Recognizing expressions is a very common activity in daily life. People can infer someone’s inner emotions from his or her facial expressions. However, not everyone writes their emotion on their face; someone may suppress true emotion and express a false facial expression depending on politeness, context, culture, or status. The suppressed expressions can be expressed fleetingly in the form of microexpressions, which usually last only 1/25 to 1/5 second.

Microexpressions were of importance for many practical applications because it reflects the true inner feeling, such as national security, deception detection, clinical therapy, emotion analysis, and human-computer interaction. The recognition of microexpressions is the premise of application of microexpression and now the recognition of microexpressions are getting more and more attention. However, perceiving other’s microexpressions is not easy. The context, culture, and perceiver himself affect the recognition of microexpression.

There are considerable efforts in the field of psychology, neuroscience, and computer science to recognize facial microexpressions. This Researc Topic illuminates the latest advances in interdisciplinary understanding how microexpressions are perceived and recognized. The authors contribute from diverse perspectives in the current research topic by using behavioral experiment, EEG, fMRI, and computer vision techniques. They investigated how humans recognize macroexpressions and microexpressions in term of modulating factors (e.g., gender, duration) and the underlying neural mechanisms, and how machine recognition algorithms and models are developed and inspired by the human recognition data.

The Research Topic reveals that research on the recognition of microexpressions is diverse but progressing. This is not surprising given that this topic receives more and more attention due to its promising potential applications. As new techniques and theories develop, it is likely that efficient and effective algorithms for recognizing microexpression will become possible. We hope that these articles provide a look into that future.

Citation: Shen, X., Chen, W., Zhao, G., Hu, P., eds. (2019). Recognizing Microexpression: An Interdisciplinary Perspective. Lausanne: Frontiers Media. doi: 10.3389/978-2-88963-137-7

Table of Contents

frontiers in Psychology

EDITORIAL
published: 04 June 2019
doi: 10.3389/fpsyg.2019.01318

Editorial: Recognizing Microexpression: An Interdisciplinary Perspective

Xunbing Shen[1]*, *Wenfeng Chen*[2]*, *Guoying Zhao*[3] *and Ping Hu*[2]

[1] Department of Psychology, Jiangxi University of Traditional Chinese Medicine, Nanchang, China, [2] Department of Psychology, Renmin University of China, Beijing, China, [3] Center for Machine Vision and Signal Analysis, University of Oulu, Oulu, Finland

Keywords: macro-expression, micro-expression, recognition, computer vision, cognitive neuroscience

OPEN ACCESS

Edited and reviewed by:
Marina A. Pavlova,
University Hospital
Tübingen, Germany

****Correspondence:***
Xunbing Shen
shenxunbing@yahoo.com
Wenfeng Chen
wchen@ruc.edu.cn

Specialty section:
This article was submitted to
Emotion Science,
a section of the journal
Frontiers in Psychology

Received: *23 April 2019*
Accepted: *20 May 2019*
Published: *04 June 2019*

Citation:
Shen X, Chen W, Zhao G and Hu P (2019) Editorial: Recognizing Microexpression: An Interdisciplinary Perspective. Front. Psychol. 10:1318. doi: 10.3389/fpsyg.2019.01318

Editorial on the Research Topic

Recognizing Microexpression: An Interdisciplinary Perspective

As a Chinese saying goes, "Look at the weather when you step out; look at people's faces when you step in." Recognizing expressions is a very common activity in daily life. People can infer someone's inner emotions from his or her facial expressions. However, human do not always wear her heart on her sleeves; someone may suppress the expressions of true feelings and express a false facial expression depending on the context of cultural rule or his/her intention. The suppressed expressions can be leaked fleetingly in the form of micro-expressions, which usually last for only 1/25 to 1/5 s.

Microexpressions may reveal the genuine inner emotion and feelings, and are important for many practical applications, such as national security, deception detection, clinical therapy, consumer behavior analysis, and human-computer interaction. Microexpression recognition is an interdisciplinary field attracting a large amount of efforts from researchers in psychology, neuroscience, and computer science. This topic illuminates the latest advances in interdisciplinary understanding how microexpressions are perceived and recognized. The authors contribute from diverse perspectives in the current research topic by using behavioral experiment, EEG, fMRI, and computer vision techniques. They investigated how human recognize macroexpressions and microexpressions in term of modulating factors (e.g., gender, duration) and the underlying neural mechanism, and how machine recognition algorithms and models are developed and inspired by the human recognition data.

Gender will influence the recognition of macro-expressions, Liu et al. investigated the interaction between facial expressions and facial gender information during face perception by using EEG technique. They found that the processing of facial expressions could affect the processing of gender in the early and later stages, which indicated by the early (P1) and late (LPC). The results provide some insights for future work on the recognition of micro-expression.

Previous studies (Adolphs, 2002) showed that the perceiver would mimic the observed expressions while recognizing them; there are close relationships between the production of facial expressions of emotion and recognition of them. From the perspective of expression production, Qu et al. investigated the awareness of facial micro-expressions and macro-expressions (all expressions last for less than 4 s). They found awareness rates were 57.79% in the real-time condition and 75.92% in the video-review condition, and the awareness rate was influenced by the intensity and (or) the duration of facial expressions.

Microexpressions is characterized as dynamic, the features of dynamic expressions may be essential for the recognition of microexpressions. Pfister et al. (2011) started pioneering research on spontaneous micro-expression recognition with the first machine vision framework to recognize spontaneous micro-expressions and achieved very promising results that compare favorably with the human accuracy. Their most recent work integrating micro-expression recognition and detection has been also reported by MIT Technology Review (see http://www.technologyreview.com/view/543501/machine-vision-algorithm-learns-to-recognizehidden-facial-expression) and achieved increasing attention (Li et al., 2018). Guo et al. investigated the dynamic features of lip corners and their characteristics in genuine and posed smiles. They found that the genuine smiles have higher amount of onset, apex, offset, and total durations, as well as offset displacement compared to posed smiles; however, the amount of onset and offset speeds, and symmetry tended to be lower. Based on these results, Li et al. regarded the deep learning as a very promising method in the automatic recognition of micro-expression.

Is micro-expression recognition a variant of recognition of macro-expression, or is it a wholly distinctive neurological process? The answer may be the latter (Shen et al., 2016). To further reveal the neural mechanisms underlying the recognition of micro-expressions, Zhao et al. investigated the brain area activities while recognizing micro-expressions of fear and surprise, they found that fear micro-expression recognition evoked greater activities in the left precuneus, middle temporal gyrus, middle frontal gyrus, and right lingual gyrus; the right postcentral gyrus and left posterior insula were responsible for the recognizing surprise micro-expressions.

It is hard for naïve human to recognize micro-expression. Usually, researchers analyze the video clips containing micro-expressions by going through them frame by frame, which is time-consuming and inefficient. To find an effective algorithm for automatically recognizing micro-expression, Peng et al. developed a Dual Temporal Scale Convolutional Neural Network (DTSCNN) for spontaneous micro-expressions recognition. They used two micro-expression databases (CASME I/II, see Yan et al., 2014) to validate the algorithm and the results showed that the method achieved a recognition rate almost 10% higher than what other state-of-the-art methods can achieve.

Automatic facial micro-expression analysis has received increasing attention in the area of computer vision. However, the limitations of current literatures exist, e.g., microexpression database and effective algorithm are fewer. Oh et al. presented a comprehensive review of state-of-the-art databases and methods for micro-expressions recognition, and pointed out the challenges and future directions in the field of automatic facial micro-expression analysis.

During the interpersonal communication, other's facial expressions such as smiling can affect decision making. He et al. investigated the effects of smiling on the responses in ultimatum games, in which they found that smiling of the proposer can lead to a lower average rejection rate.

Together, the topic reveal that research on the recognition of microexpressions are diverse but progressing. This is not surprising given that it receives more and more attention due to the promising potential applications. As new techniques and theories develop, it is likely that efficient and effective algorithms for recognizing microexpression are promising. We hope that these articles provide a look into that future.

AUTHOR CONTRIBUTIONS

XS had primary writing responsibility. WC and GZ revised the manuscript. PH assisted with the preparation of the manuscript.

FUNDING

This project was partially supported by a grant from the Planed Project of Social Sciences in Jiangxi Province, China (Project No. 18JY24), Tekes Fidipro program (Grant No. 1849/31/2015) and Infotech Oulu, the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (18XNLG10) and the National Natural Science Foundation of China (31371031, 61632004) to WC.

REFERENCES

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behav. Cogn. Neurosci. Rev.* 1, 21–62. doi: 10.1177/1534582302001001003

Li, X., Hong, X., Moilanen, A., Huang, X., Pfister, T., Zhao, G., et al. (2018). Towards reading hidden emotions: a comparative study of spontaneous micro-expression spotting and recognition methods. *IEEE Trans. Affect. Comput.* 9, 563–577. doi: 10.1109/TAFFC.2017.2667642

Pfister, T., Li, X., Zhao, G., and Pietikäinen, M. (2011). "Recognising Spontaneous Facial Micro-expressions," in *Proceeding International Conference on Computer Vision (ICCV 2011)* (Barcelona).

Shen, X., Wu, Q., Zhao, K., and Fu, X. (2016). Electrophysiological evidence reveals differences between the recognition of microexpressions and macroexpressions. *Front. Psychol.* 7:1346. doi: 10.3389/fpsyg.2016.01346

Yan, W.-J., Li, X., Wang, S.-J., Zhao, G., Liu, Y.-J., Chen, Y.-H., et al. (2014). CASME II: An improved spontaneous micro-expression database and the baseline evaluation. *PLoS ONE* 9:e86041. doi: 10.1371/journal.pone.0086041

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 14 August 2017
doi: 10.3389/fpsyg.2017.01383

Symmetrical and Asymmetrical Interactions between Facial Expressions and Gender Information in Face Perception

Chengwei Liu[1,2], Ying Liu[2], Zahida Iqbal[2], Wenhui Li[3], Bo Lv[4] and Zhongqing Jiang[2]*

[1] School of Education, Hunan University of Science and Technology, Xiangtan, China, [2] School of Psychology, Liaoning Normal University, Dalian, China, [3] College of Preschool and Primary Education, Shenyang Normal University, Shenyang, China, [4] Collaborative Innovation Center of Assessment toward Basic Education Quality, Beijing Normal University, Beijing, China

To investigate the interaction between facial expressions and facial gender information during face perception, the present study matched the intensities of the two types of information in face images and then adopted the orthogonal condition of the Garner Paradigm to present the images to participants who were required to judge the gender and expression of the faces; the gender and expression presentations were varied orthogonally. Gender and expression processing displayed a mutual interaction. On the one hand, the judgment of angry expressions occurred faster when presented with male facial images; on the other hand, the classification of the female gender occurred faster when presented with a happy facial expression than when presented with an angry facial expression. According to the evoked-related potential results, the expression classification was influenced by gender during the face structural processing stage (as indexed by N170), which indicates the promotion or interference of facial gender with the coding of facial expression features. However, gender processing was affected by facial expressions in more stages, including the early (P1) and late (LPC) stages of perceptual processing, reflecting that emotional expression influences gender processing mainly by directing attention.

Keywords: facial expression, facial gender, interaction, ERP, face perception

Edited by:
Xunbing Shen,
Jiangxi University of Traditional Chinese Medicine, China

Reviewed by:
Shimin Fu,
Tsinghua University, China
Jianhui Wu,
Institute of Psychology (CAS), China

****Correspondence:***
Zhongqing Jiang
jzqcjj@hotmail.com

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *24 April 2017*
Accepted: *31 July 2017*
Published: *14 August 2017*

Citation:
Liu C, Liu Y, Iqbal Z, Li W, Lv B and Jiang Z (2017) Symmetrical and Asymmetrical Interactions between Facial Expressions and Gender Information in Face Perception. Front. Psychol. 8:1383. doi: 10.3389/fpsyg.2017.01383

INTRODUCTION

Facial expressions and gender information are always intertwined in human faces. We perceive a difference between a crying male and a crying female because there is an interaction between facial expression information and gender information. Previous studies have provided evidence to support this idea; for example, participants were usually faster and more accurate in detecting angry expressions on male faces and happy expressions on female faces (Becker et al., 2007), and gender classification occurred faster with happy female faces than angry female faces (Aguado et al., 2009). Previous studies have also provided neurophysiological evidence of an interaction between facial expression and gender. An evoked-related potential (ERP) study revealed an interaction between facial expressions and gender in the face-sensitive N170 component (Valdés-Conroy et al., 2014). A functional magnetic resonance imaging (fMRI) study revealed that the left amygdala in female participants was more active in successfully remembering fearful female faces, while the right amygdala in male participants was more involved in the memory of fearful male faces (Armony and Sergerie, 2007).

The following two different hypotheses regarding the interaction between facial expressions and gender have been proposed: bottom-up processing and top-down processing. The bottom-up processing hypothesis posits that the interaction between facial expressions and gender is a result of an overlap between two types of information (Becker et al., 2007; Hess and Anemarie, 2010; Zebrowitz et al., 2010; Slepian et al., 2011). For example, both a male face and an angry face have a smaller brow-to-lid distance; meanwhile, happy expressions could have an increase brow-to-lid distance, which is more similar to the female facial features (Slepian et al., 2011). The top-down processing hypothesis posits that top-down information (e.g., gender stereotypes, such as women tending to smile more than men, and men expressing anger more frequently than women) is the cause of the interaction between facial expressions and gender (Fabes and Martin, 1991; Lafrance et al., 2003; Neel et al., 2012). Although these two hypotheses are contradictory, the effect on people's responses are nearly identical. We named this effect the associated effect of facial expression and gender.

Although current theories of facial perception tend to agree that there is an interaction between facial expressions and gender processing, there are conflicting findings regarding the manifestation of this interaction. Gender information has been found to affect the categorization of emotional expressions, whereas emotional expressions did not affect the categorization of gender information (Atkinson et al., 2005; Karnadewi and Lipp, 2011). Gender classification was shown to be influenced by facial expression information, but expression classifications remain relatively unaffected by the facial gender (Wu et al., 2015). Some studies have shown no interaction between facial expressions and gender processing, supporting that independent routes exits for processing facial expressions and gender (Le and Bruce, 2002; Nijboer and Jellema, 2012).

Regarding the causes of the contradictory results regarding the interaction between facial expressions and gender, we speculated that in addition to the reasons noted by Karnadewi and Lipp (2011), e.g., expression type, experimental paradigm, stimuli, etc., the relative strength of the two types of information (e.g., expression vs. gender) could modulate their interaction. The intensity of the facial expression affected the accuracy of the expression recognition (Montagne et al., 2007; Hoffmann et al., 2010). Garner (1983) noted that, during a multiple dimensional stimuli processing, the dimension with slower speed of processing was more susceptible to the faster. Therefore, the asymmetric interaction between facial expressions and gender information might be due to a mismatch in their intensities. If the intensity of the two types of information was matched, their interaction would likely be symmetrical, which is one of the main hypotheses tested in the present study.

Although the mutual influence of gender and expression could be symmetrical if their intensities were matched, the precise stage of facial processing during which one type of information influences the other could be different because there are differences in the time course of gender and expression processing. Gender information was observed to be quickly and automatically processed using ERP technology, which was reflected by the N170 component, whereas during the later processing stages, gender information was no longer processed if it was irrelevant to the task (Mouchetant-Rostaing et al., 2000; Castelli et al., 2004; Tomelleri and Castelli, 2012). Emotion information processing is relatively faster than gender information processing in face perception processing (Wang et al., 2016); the effect of information processing appears as early as 100 ms from the onset of a stimulus, which is indexed on the P1 ERP component (Pourtois et al., 2004; Rellecke et al., 2012). Furthermore, the emotion effect was also observed in the late positive component (LPC) (Wild-Wall et al., 2008; Frühholz et al., 2009; Hietanen and Astikainen, 2013).

Using both an expression task and a gender task, the present study explores the mutual impact of expressions and gender when one type of information is task-relevant, while the other is task-irrelevant. Based on the above discussion, we hypothesize that during the gender classification task, the facial expression effect can occur as early as the P1 component, and the facial gender effect is hypothesized to occur during the N170 component in the expression classification task.

MATERIALS AND METHODS

Participants

Upon obtaining the approval of the Ethics Committee at University, a recruitment advertisement was posted at the entrance to the University, which is visibly accessible to all students. Twenty right-handed undergraduate participants (11 males, 9 females; aged 18–22 years; $M = 19.55$, $SD = 1.23$) were recruited for the experiment. The participants reported no history of brain diseases, or chronically taking any medicine affecting brain activity.

Material Evaluation and Selection

Twenty-three undergraduate participants (11 males, 12 females; aged 18–22 years; $M = 19.84$, $SD = 1.25$) were requested to rate gender and expression intensity information of 185 face images from CAPS (Chinese Affective Picture System) (Bai et al., 2005) on a 9-point scale. For the expression component, the participants were instructed to rate the faces according to how angry or happy the faces appeared (1 = *very angry*, 5 = *neither angry nor happy*, 9 = *very happy*). For the gender information, the participants rated how masculine or feminine the faces appeared (1 = *very masculine*, 5 = *neither masculine nor feminine*, 9 = *very feminine*). Although gender and expression information is different in nature, the evaluation of the intensity of the two types of information is comparable due to the use of the same participants and pictures.

According to on the above mentioned rating results, we selected 80 faces with a balanced gender and expression intensity. A paired samples *t*-test showed that there were no significant differences in the intensity between the two types of information

TABLE 1 | The intensity of the gender and expression information in each group of images.

	Happy face		Angry face	
Information type	**Female face**	**Male face**	**Female face**	**Male face**
Gender	8.14 ± 0.05	7.98 ± 0.05	7.88 ± 0.05	8.14 ± 0.05
Emotion	8.11 ± 0.05	8.06 ± 0.05	7.98 ± 0.05	8.01 ± 0.05

(gender and expression) in the happy face pictures, $t(39) = 0.73$, $p > 0.05$, and the angry face pictures, $t(39) = 0.38$, $p > 0.05$. An independent samples t-test revealed no significant difference in the intensity of the gender information between the happy and angry faces, $t(78) = 0.83$, $p > 0.05$, or in intensity of the expression information between the female and male faces, $t(78) = 0.20$, $p > 0.05$. Descriptions of these evaluations are shown in **Table 1**.

Procedures

The participants were seated in a quiet room in front of a computer at a distance of approximately 90 cm from the monitor screen. The face stimuli were presented in the center of the screen. All participants completed two tasks (expression discrimination: happy vs. angry; gender discrimination: male vs. female). Half of the subjects were first asked to discriminate between the facial expressions (happy vs. angry). The participants responded by pressing the right and left mouse buttons. The participants were provided 5 min of rest after the expression task was completed, and then the participants were asked to discriminate between male and female faces. The other half of participants were tested in the reverse order. Each stimulus combination (for example, happy female) was presented three times in each block, thus providing 240 trials per block for a total of 480 trials. A 2 × 2 × 2 within-subjects design was used, with gender (male vs. female), expression (angry vs. happy), and tasks (expression discrimination vs. gender discrimination) as the two levels.

The experiment included practice and formal sessions. During the practice session, the participants were presented with 16 pictures of faces and received feedback on their responses. Each trial began with a 500 ms fixation cross ("+") at the center of the computer screen, followed by 500~800 ms of a blank screen and the target face image. The face image remained on the screen until the participants responded or 1500 ms had passed (see **Figure 1**). The participants were instructed to judge the expression or gender of the face as quickly and accurately as possible. The participants responded by pressing keys. The assignment of the key mapping and task order was counterbalanced across the participants.

Electroencephalogram (EEG) Signal Acquisition and Analysis

The EEG signals were sampled at 500 Hz from 64 cap-mounted Ag/AgCl electrodes referenced to the left mastoid and placed according to the expanded international 10–20 system (Neuroscan Inc., United States). The impedance was below 5 KΩ. The EEG was amplified using a bandpass filter of 0.05–40 Hz. Due to the interference of ocular potentials, horizontal eye movements were monitored by electrodes placed on the outside of each eye, and vertical movements were monitored separately by electrodes located above and below the left eye.

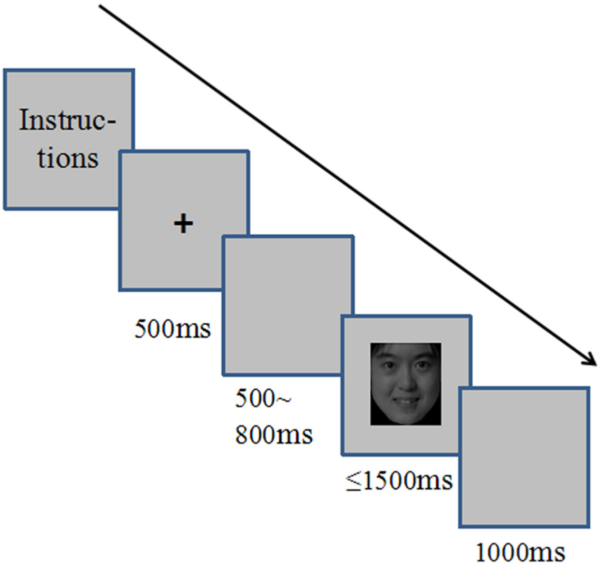

FIGURE 1 | The sequence of events during an experiment trial.

The EEG signals were re-referenced off-line to the common average of all scalp electrodes. Artifacts were rejected automatically if the signal amplitude exceeded ± 80 μV. Epochs of 1000 ms after the stimuli onset were computed with an additional 200 ms pre-stimulus baseline.

According to the ERP waveforms and previous studies (Itier and Taylor, 2002; Sato and Yoshikawa, 2007; Recio et al., 2011; Jiang et al., 2014), the amplitudes and latencies of each ERP component were derived from the averaged data obtained during the selected time windows over the electrode clusters as follows: P100 (100~160 ms) and N170 component (160~210 ms) over the electrode group including PO7, PO5, PO3, PO4, PO6, PO8, O1, OZ, and O2; LPC (350~800 ms) over the electrode group including CP1, CPZ, CP2, P1, PZ, and P2.

RESULTS

Behavioral Results

We tested the response accuracy using a 2 × 2 × 2 ANOVA, with task, expression and gender as the repeated-measures factors. The analysis did not find a significant main effect of task, $F(1,19) = 3.79$, $p > 0.05$, but a significant effect was found for facial expressions, $F(1,19) = 19.07$, $p < 0.01$, $\eta_p^2 = 0.50$, with a higher accuracy in the responses to the happy faces ($M = 0.96$, $MSE = 0.01$) than the responses to the angry faces ($M = 0.93$, $MSE = 0.01$). Importantly, a significant interaction was observed between facial expression and gender,

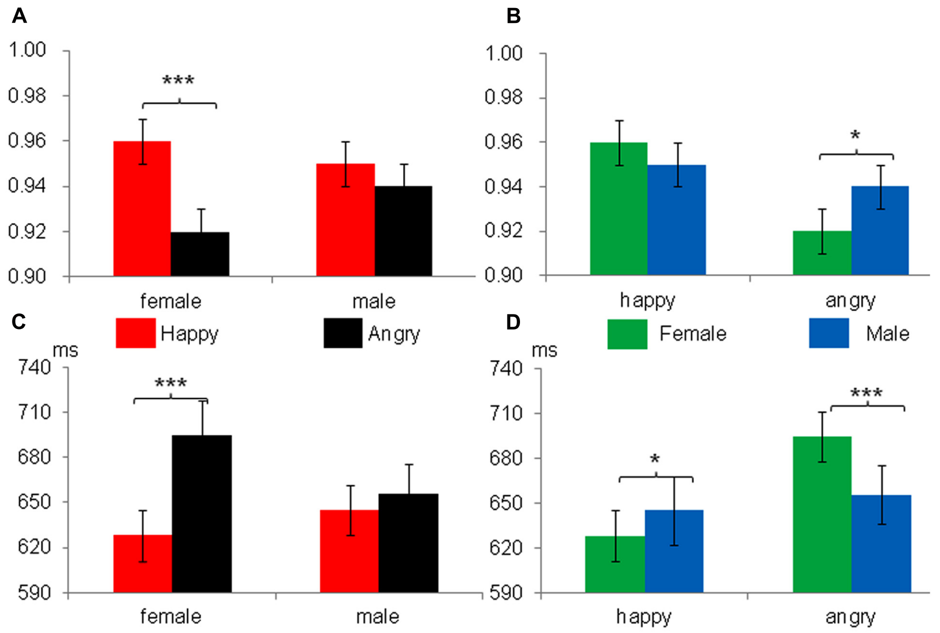

FIGURE 2 | Participants' accuracy **(A,B)** and response times **(C,D)** as a function of facial emotion and gender; the left images **(A,C)** reflect the effect of gender on expression processing; the right images **(B,D)** reflect the effect of expression on gender processing. *p < 0.05, ***p < 0.001.

$F(1,19) = 5.49$, $p < 0.05$, $\eta_p^2 = 0.22$. No task × facial expression × face gender interaction was found.

We further analyzed the interaction between facial expressions and gender from two perspectives. First, we explored the influence of expression on gender classification (see **Figure 2A**). The accuracy of judging the gender of a female face was significantly lower under the condition of angry faces ($M = 0.92$, $MSE = 0.01$) than under the condition of happy faces ($M = 0.96$, $MSE = 0.01$), $F(1,19) = 40.49$, $p < 0.001$, $\eta_p^2 = 0.68$. However, there was no significant difference in the recognition of male faces between the angry face ($M = 0.94$, $MSE = 0.01$) and happy face ($M = 0.95$, $MSE = 0.01$) conditions. Second, we explored the influence of gender on expression recognition (see **Figure 2B**), and the accuracy of judging an angry expression was significantly lower for female faces ($M = 0.92$, $MSE = 0.01$) than for male faces ($M = 0.94$, $MSE = 0.01$), $F(1,19) = 4.64$, $p < 0.05$, $\eta_p^2 = 0.19$. However, no significant differences were found in the accuracy of judging a happy expression between the female ($M = 0.96$, $MSE = 0.01$) and male ($M = 0.95$, $MSE = 0.01$) face conditions, $F(1,19) = 1.12$, $p > 0.05$, $\eta_p^2 = 0.06$.

A similar result was observed in the response time analysis. As shown in **Figure 1**, a significant main effect of facial expressions was found, $F(1,19) = 32.24$, $p < 0.001$, $\eta_p^2 = 0.63$, with faster RTs in response to happy expressions (636.51 ± 16.60 ms) than those in response to angry expressions (675.18 ± 21.29 ms). There was no significant effect of task, $F(1,19) = 2.15$, $p > 0.05$. There was a significant interaction between facial expressions and gender, $F(1,19) = 36.13$, $p < 0.001$, $\eta_p^2 = 0.66$. No task × facial expressions × gender interaction was found.

We further analyzed the interaction effect from two perspectives. First, regarding the influence of expression on gender recognition (see **Figure 2C**), the participants were slower to judge the gender of angry female faces ($M = 694.56$ ms, $MSE = 23.16$ ms) than they were to judge happy female faces ($M = 628.03$ ms, $MSE = 17.01$ ms), $F(1,19) = 43.43$, $p < 0.001$, $\eta_p^2 = 0.70$. However, there was no significant difference in judging the gender of male faces between angry ($M = 655.81$ ms, $MSE = 19.91$ ms) and happy expressions ($M = 644.97$ ms, $MSE = 16.88$ ms). Second, regarding the influence of gender on expression recognition (see **Figure 2D**), the participants were slower to classify the angry expressions on female faces ($M = 694.56$ ms, $MSE = 23.16$ ms) than those on male faces ($M = 655.81$ ms, $MSE = 19.91$ ms), $F(1,19) = 29.17$, $p < 0.001$, $\eta_p^2 = 0.61$. The results were opposite for the judgment of happy expressions as follows: the participants were slower to react to the male faces ($M = 644.97$ ms, $MSE = 16.88$ ms) than the female faces ($M = 628.03$ ms, $MSE = 17.01$ ms), $F(1,19) = 6.2$, $p < 0.05$, $\eta_p^2 = 0.25$.

ERP Results

We performed a repeated-measures ANOVA using task (2: expression discrimination vs. gender discrimination), gender (2: male vs. female), and expression (2: angry vs. happy) as the within-subjects factors to analyze the amplitudes of P1, N170, and LPC separately. The results of the analysis revealed a significant task × facial expressions × gender interaction [P1 component, $F(1,19) = 7.04$, $p < 0.05$, $\eta_p^2 = 0.27$; N170 component, $F(1,19) = 15.05$, $p < 0.05$, $\eta_p^2 = 0.44$; LPC component, $F(1,19) = 4.48$, $p < 0.05$, $\eta_p^2 = 0.19$]. Therefore, we further explored the relationship between facial expressions and gender separately under the different task conditions.

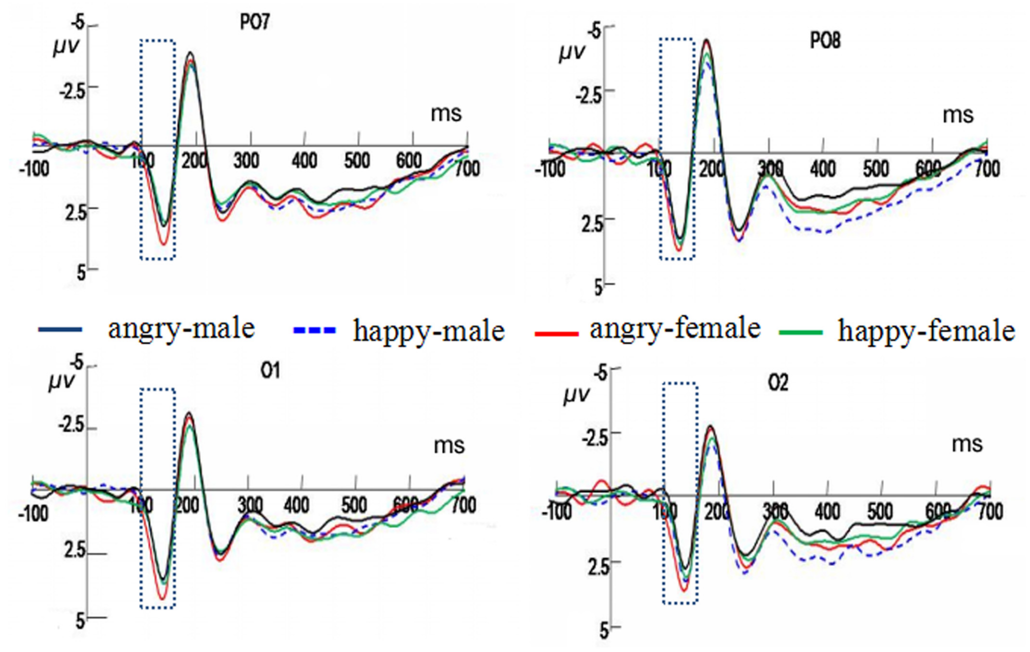

FIGURE 3 | Averaged evoked-related potential (ERPs) at PO7, PO8, O1, and O2 in response to angry male faces, angry female faces, happy male faces, and happy female faces under the gender classification conditions. The time window of P1 is shown by the rectangle.

Gender Classification Task

A 2 (Gender) × 2 (Expression) repeated-measures ANOVA of the mean amplitude values of P1 and LPC revealed a significant facial expression × gender interaction, but no significant interactions were observed in the N170 component.

P1 (100–160 ms)

There was a significant interaction between facial expression and gender, $F(1,19) = 5.48$, $p < 0.05$, $\eta_p^2 = 0.22$. Further analysis revealed that higher amplitudes were elicited by the angry female faces (4.79 ± 0.59 μV) than by the happy female faces (4.14 ± 0.48 μV), $F(1,19) = 4.49$, $p < 0.05$, $\eta_p^2 = 0.19$, but no significant difference was observed in the gender classification of the male faces between the angry (3.96 ± 0.60 μV) and happy expression (4.17 ± 0.49 μV) (see **Figure 3**) conditions.

LPC (350–800 ms)

The interaction between facial expression and gender was significant, $F(1,19) = 5.66$, $p < 0.05$, $\eta_p^2 = 0.23$. In the male faces, angry expressions elicited higher amplitudes (11.84 ± 1.03 μV) than the happy faces (11.08 ± 0.86 μV), $F(1,19) = 11.03$, $p < 0.01$, $\eta_p^2 = 0.37$. There was no significant difference in the female facial expressions (see **Figure 4**).

Expression Classification Task

In the expression classification task, a significant interaction between facial expression and gender was obtained only in the N170 component, $F(1,19) = 6.76$, $p < 0.05$, $\eta_p^2 = 0.26$. Further analysis revealed that judging happy expressions in male faces elicited more negative amplitudes (−3.35 ± 0.97 μV) than that judging female faces (−2.68 ± 0.89 μV), $F(1,19) = 4.59$, $p < 0.05$, $\eta_p^2 = 0.19$. No difference was found in judging angry expressions between the male and female faces (see **Figure 5**).

DISCUSSION

In the present study, we selected face pictures with equivalent intensities of gender and expression information to perform an experiment that required participants to judge both expression and gender. The behavioral results revealed a significant interaction between gender and expression information in both tasks. Interestingly, the ERP results showed that the interaction between facial expressions and gender occurred during different stages of face processing because of the different tasks. The effect of facial expressions on gender processing was mainly reflected during the P1 and LPC components, while gender affected expression processing only during the N170 component.

Symmetrical Interaction in Terms of the Existence of a Mutual Effect

The results of the behavior data in the present study revealed a symmetrical interaction between gender and facial expressions in face processing; thus, one type of information (i.e., gender or expression) processing was affected by the other (i.e., expression or gender). This result is inconsistent with previous studies (Atkinson et al., 2005; Karnadewi and Lipp, 2011) that reported that only gender information affects expression processing.

In a previous study (Atkinson et al., 2005; Karnadewi and Lipp, 2011), gender information unidirectionally affected expression information processing, which may have been due to the stronger intensity of the gender information relative to the emotional information because the gender classification was relatively faster than the expression classification. The present study pre-matched the intensity of the emotional information and gender information, which was also evidenced by the non-significant differences in the response time and accuracy of the participants' performance during the expression judgment task and the gender judgment task. Therefore, when the two types of information are matched in intensity, a bidirectional influence of expression on gender and of gender on expression was found; therefore, we hypothesize that their interaction is symmetrical.

Asymmetrical Interaction in Terms of Temporal Courses

The ERP data revealed that the interaction between facial expressions and gender differed along the time course of the face classification. The effect of gender was reflected in the N170 component, while the effect of facial expressions was mainly embodied in the P1 and LPC components.

After analyzing the details of these ERP results, we hypothesized that there were at least two underlying mechanisms. The first mechanism is the associated effect of gender and expression; that is, the congruence of the features of gender and expression (e.g., angry and male face vs. happy and female face) could facilitate their processing. Otherwise, if their features were incongruent (e.g., angry and female face vs. happy and male face), their processing could be hindered (Slepian et al., 2011). The second mechanism is the general effect of emotion, which usually appears as a negativity bias; thus, negative emotional stimuli could result in greater ERP components than positive stimuli (Huang and Luo, 2006). These two mechanisms could be added synergistically or cancel each other's effect.

In the expression classification task, the happy male face elicited more negative N170 responses than the happy female face, but there was no significant difference in the N170 responses between the male face and female face when both faces were angry. In the present study, the happy expression was congruent with the female faces instead of the male faces. The inconsistent relationship between facial expressions and gender could increase the difficulty of face processing, hinder the participants' performance, and increase the intensity of the responses in N170 (Rossion et al., 2000) in classifying happy expressions on male faces compared to classifying female faces. Regarding the classification of the angry expressions, although the participants' response times and accuracy were different between the male and female faces, there was no consistency in the N170 component, which may be due to the joint effect of the two mechanisms mentioned above. Considering the associated effect of gender and expression, the features of expression and gender in the angry female faces were incongruent, but they were congruent in the angry male faces (Slepian et al., 2011), which increased the difficulty of the expression classification task for the angry female face. Thus, this incongruency could increase the N170 response to angry female faces relative to that to angry male faces; on the other hand, the emotion of anger may itself increase N170 as previous studies have reported a negativity bias (Batty and Taylor, 2003; Caharel et al., 2005; Huang and Luo, 2006). Therefore, there could be a ceiling effect on N170 that masks the differences between the male and female angry faces.

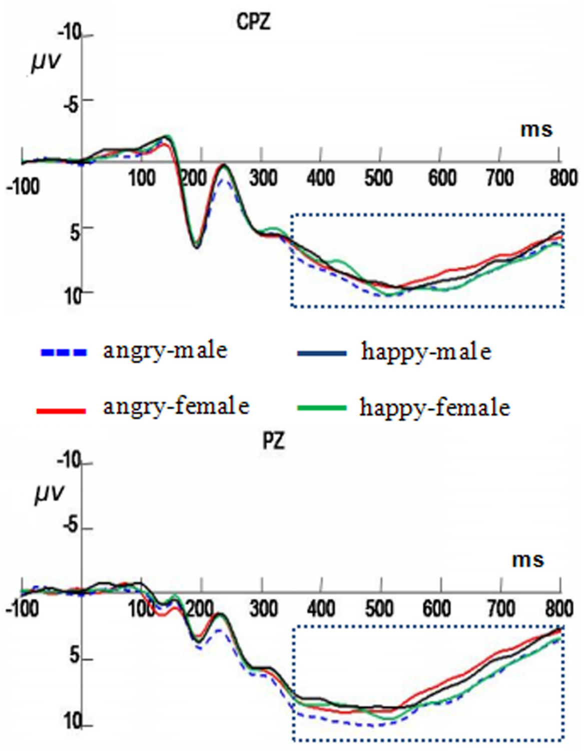

FIGURE 4 | Averaged ERP sat Cpz and Pz in response to the angry male faces, angry female faces, happy male faces, and happy female faces under the gender classification conditions. The time window of late positive component (LPC) is shown by the rectangle.

Regarding the gender classification task, the effect of expression first presented during the early ERP component of P1. The female face with an angry expression elicited more positive P1 than the happy expression, but no significant difference between the angry and happy faces was found for the male faces. These ERP results are consistent with the behavioral results, which revealed that the participants were slower and less accurate in classifying the gender when the expression was angry instead of happy only in female faces but not in male faces. P1 is considered to reflect the processing of low-level features in the extra-striatal visual cortex, and stimuli with special features usually induce more positive P1 amplitudes (Hillyard and Anllo-Vento, 1998). Considering that facial expressions produce distortions in the shape of individual facial features, such as lip raising or eye widening (Calder et al., 2001) and Slepian et al. (2011) noted that the female face naturally resembles a happy expression instead of an angry expression, we hypothesized that female faces would be more distorted by angry expressions than by happy expressions due to the incongruence, thus inducing larger P1 amplitudes in response to the angry female faces and increasing the difficulty of judging facial gender. Furthermore, the effect of emotion, which

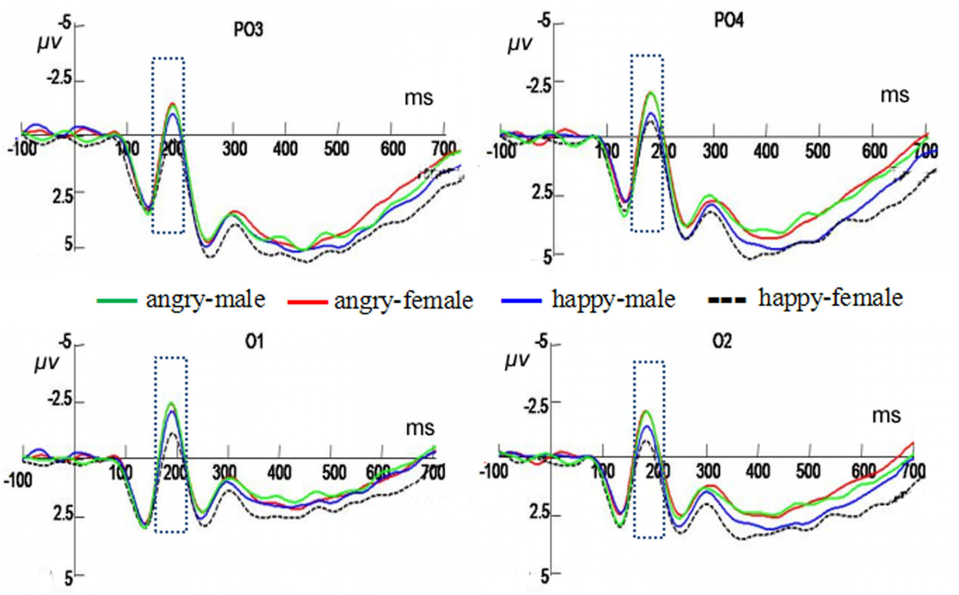

FIGURE 5 | Averaged ERPs at PO3, PO4, O1, and O2 in response to the angry male faces, angry female faces, happy male faces, and happy female faces under the expression classification conditions. The time window of N170 is shown by the rectangle.

appeared as a negativity bias in this study, could contribute to the larger P1 amplitude in response to the angry female face than that to a happy female face.

Similarly, in the male faces, a happy expression could cause more distortion in the facial features than an angry expression because the features of male faces are more congruent with anger (Calder et al., 2001; Slepian et al., 2011). Therefore, the P1 amplitude in response to happy male faces should be larger than that in response to angry male faces; however, considering the negativity bias (Huang and Luo, 2006), an angry male face could elicit a larger P1 amplitude than a happy male face. Therefore, these two effects could play contradicting roles in modulating the amplitude of P1 such that the comparison between the P1 amplitude in response to the happy male faces and angry male faces became non-significant.

During the second stage of the expression effect on gender classification, which was reflected by the LPC, male faces with an angry expression elicited higher amplitudes than happy faces, but there was no significant difference in the LPC between the angry and happy expressions on female faces. The difference between the two expressions on male faces are similar to those observed in previous studies and display a negativity bias (Cacioppo and Berntson, 1994; Cacioppo et al., 1997; Wild-Wall et al., 2008; Frühholz et al., 2009; Hietanen and Astikainen, 2013). Meanwhile, the effect of expression on female faces was non-significant. We suspect this might be due to the congruency of expression and gender information in an angry male face, which could emphasize the angry information such that its effects could also be reflected during the LPC even under the condition of implicit processing (gender classification task). Angry female faces, however, demonstrate atypical facial expression features and thus could not be reflected during this stage. This result also confirms that the LPC, unlike the early ERP components (e.g., P1 and N170), most likely reflects the psychological meaning rather than the physical features of the stimuli.

Limitations of the Present Study

Although the present study revealed differences in the interaction between gender and facial expressions using ERPs, there were certain confounding factors in the mechanism of the interaction. For example, the analysis could not directly distinguish the associated effect of gender and expression from the general effect of emotion, nor provide direct evidence differentiating the physical feature-based effects (i.e., through bottom-up processing) from the gender stereotypes-based effects (i.e., through top-down processing) in the interaction. The main cause of these limitations is that we did not separate the physical features from the concept of gender or expression in the stimuli. To resolve this confusion, a specific experimental paradigm (Fu et al., 2012) and stimuli (Ip et al., 2017) might be helpful.

CONCLUSION

In summary, the present study revealed a symmetrical interaction in terms of the existence of a mutual effect between gender and expression processing during face perception when the intensity of both types of information was matched.

Furthermore, the present study also revealed asymmetry in the psychological and physiological mechanisms underlying the interaction between gender and expression information. The ERP results provided evidence that facial expression affected gender processing mainly by attracting the participants' attention, which occurred during the early and late stages of face processing and was indexed by P1 and LPC; meanwhile, gender affected expression processing during the face structural encoding stage, as indexed by N170, by facilitating or interfering with facial expression structural information processing.

ETHICS STATEMENT

This study was performed in accordance with the recommendations of the "Experimental guidelines, Liaoning Normal University Ethics Committee." After being fully informed of the study, the participants provided written informed consent.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: CL, YL, and ZJ. Performed the experiments: CL. Analyzed the data: CL and ZJ. Wrote the paper: CL, ZJ, ZI, WL, and BL.

ACKNOWLEDGMENTS

This study was funded by the National Natural Science Foundation of China (Nos. 31271107 and 31400869), the Natural Science Foundation of Liaoning Province (No. 201602454), and the Open Platform Project of the Office of Education of Hunan Province (No. 16K034).

REFERENCES

Aguado, L., Garcíagutierrez, A., and Serranopedraza, I. (2009). Symmetrical interaction of sex and expression in face classification tasks. *Atten. Percept. Psychophys.* 71, 9–25. doi: 10.3758/APP.71.1.9

Armony, J. L., and Sergerie, K. (2007). Own-sex effects in emotional memory for faces. *Neurosci. Lett.* 426, 1–5. doi: 10.1016/j.neulet.2007.08.032

Atkinson, A. P., Tipples, J., and Burt, D. M. (2005). Asymmetric interference between sex and emotion in face perception. *Percept. Psychophys.* 67, 1199–1213. doi: 10.3758/BF03193553

Bai, L., Ma, H., Huang, Y., and Luo, Y. (2005). The development of native chinese affective picture system—a pretest in 46 college students. *Chin. Mental Health J.* 19, 719–722. doi: 10.3321/j.issn:1000-6729.2005.11.001

Batty, M., and Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Brain Res. Cogn. Brain Res.* 17, 613–620. doi: 10.1016/S0926-6410(03)00174-5

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., and Smith, D. M. (2007). The confounded nature of angry men and happy women. *J. Pers. Soc. Psychol.* 92, 179–190. doi: 10.1037/0022-3514.92.2.179

Cacioppo, J. T., and Berntson, G. G. (1994). Relationship between attitudes and evaluative space: a critical review, with emphasis on the separability of positive and negative subtrates. *Psychol. Bull.* 115, 401–423. doi: 10.1037/0033-2909.115.3.401

Cacioppo, J. T., Gardner, W. L., and Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: the case of attitudes and evaluative space. *Pers. Soc. Psychol. Rev.* 1, 3–25. doi: 10.1207/s15327957pspr0101-2

Caharel, S., Courtay, N., Bernard, C., Lalonde, R., and Rebaï, M. (2005). Familiarity and emotional expression influence an early stage of face processing: an electrophysiological study. *Brain Cogn.* 59, 96–100. doi: 10.1016/j.bandc.2005.05.005

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., and Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vis. Res.* 41, 1179–1208. doi: 10.1016/S0042-6989(01)00002-5

Castelli, L., Macrae, C. N., Zogmaister, C., and Arcuri, L. (2004). A tale of two primes: contextual limits on stereotype activation. *Soc. Cogn.* 22, 233–247. doi: 10.1521/soco.22.2.233.35462

Fabes, R. A., and Martin, C. L. (1991). Gender and age stereotypes of emotionality. *Pers. Soc. Psychol. Bull.* 17, 532–540. doi: 10.1177/0146167291175008

Frühholz, S., Fehr, T., and Herrmann, M. (2009). Early and late temporo-spatial effects of contextual interference during perception of facial affect. *Int. J. Psychophysiol.* 74, 1–13. doi: 10.1016/j.ijpsycho.2009.05.010

Fu, S., Feng, C., Guo, S., Luo, Y., and Parasuraman, R. (2012). Neural adaptation provides evidence for categorical differences in processing of faces and Chinese characters: an ERP study of the N170. *PLoS ONE* 7:e41103. doi: 10.1371/journal.pone.0041103

Garner, B. W. R. (1983). "Asymmetric interactions of stimulus dimensions in perceptual in formation processing," in *Perception, Cognition, and Development: Interactional Analyses*, eds T. J. Tighe and B. E. Shepp (Hillsdale, NJ: Erlbaum), 1–38.

Hess, and Anemarie, L. (2010). The relationship between gender, lifetime number of depressive episodes, treatment type, and treatment response in chronic depression. *J. Oral Maxillofac. Res.* 2, e1.

Hietanen, J. K., and Astikainen, P. (2013). N170 response to facial expressions is modulated by the affective congruency between the emotional expression and preceding affective picture. *Biol. Psychol.* 92, 114–124. doi: 10.1016/j.biopsycho.2012.10.005

Hillyard, S. A., and Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proc. Natl. Acad. Sci. U.S.A.* 95, 781–787. doi: 10.1073/pnas.95.3.781

Hoffmann, H., Kessler, H., Eppel, T., Rukavina, S., and Traue, H. C. (2010). Expression intensity, gender and facial emotion recognition: women recognize only subtle facial emotions better than men. *Acta Psychol.* 135, 278–283. doi: 10.1016/j.actpsy.2010.07.012

Huang, Y. X., and Luo, Y. J. (2006). Temporal course of emotional negativity bias: an ERPstudy. *Neurosci. Lett.* 398, 91–96. doi: 10.1016/j.neulet.2005.12.074

Ip, C., Wang, H., and Fu, S. (2017). Relative expertise affects N170 during selective attention to superimposed face-character images. *Psychophysiology* 54, 955–968. doi: 10.1111/psyp.12862

Itier, R. J., and Taylor, M. J. (2002). Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: a repetition study using erps. *Neuroimage* 15, 353–372. doi: 10.1006/nimg.2001.0982

Jiang, Z., Li, W., Recio, G., Liu, Y., Luo, W., Zhang, D., et al. (2014). Time pressure inhibits dynamic advantage in the classification of facial expressions of emotion. *PLoS ONE* 9:e100162. doi: 10.1371/journal.pone.0100162

Karnadewi, F., and Lipp, O. V. (2011). The processing of invariant and variant face cues in the Garner Paradigm. *Emotion* 11, 563–571. doi: 10.1037/a0021333

Lafrance, M., Hecht, M. A., and Paluck, E. L. (2003). The contingent smile: a meta-analysis of sex differences in smiling. *Psychol. Bull.* 129, 305–334. doi: 10.1037/0033-2909.129.2.305

Le, G. P., and Bruce, V. (2002). Evaluating the independence of sex and expression in judgments of faces. *Atten. Percept. Psychophys.* 64, 230–243. doi: 10.3758/BF03195789

Montagne, B., Kessels, R. P., De Haan, E. H., and Perrett, D. I. (2007). The emotion recognition task: a paradigm to measure the perception of facial emotional expressions at different intensities. *Percept. Motor Skills* 104, 589–598. doi: 10.2466/pms.104.2.589-598

Mouchetant-Rostaing, Y., Giard, M. H., Bentin, S., Aguera, P. E., and Pernier, J. (2000). Neurophysiological correlates of face gender processing in humans. *Eur. J. Neurosci.* 12, 303–310. doi: 10.1046/j.1460-9568.2000.00888.x

Neel, R., Becker, D. V., Neuberg, S. L., and Kenrick, D. T. (2012). Who expressed what emotion? Men grab anger, women grab happiness. *J. Exp. Soc. Psychol.* 48, 583–586. doi: 10.1016/j.jesp.2011.11.009

Nijboer, T. C., and Jellema, T. (2012). Unequal impairment in the recognition of positive and negative emotions after right hemisphere lesions: a left hemisphere bias for happy faces. *J. Neuropsychol.* 6, 79–93. doi: 10.1111/j.1748-6653.2011.02007

Pourtois, G., Grandjean, D., Sander, D., and Vuilleumier, P. (2004). Electrophysiological correlates of rapid spatial orienting towards fearful faces. *Cereb. Cortex* 14, 619–633. doi: 10.1093/cercor/bhh023

Recio, G., Sommer, W., and Schacht, A. (2011). Electrophysiological correlates of perceiving and evaluating static and dynamic facial emotional expressions. *Brain Res.* 1376, 66–75. doi: 10.1016/j.brainres.2010.12.041

Rellecke, J., Sommer, W., and Schacht, A. (2012). Does processing of emotional facial expressions depend on intention? Time-resolved evidence from event-related brain potentials. *Biol. Psychol.* 90, 23–32. doi: 10.1016/j.biopsycho.2012.02.002

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., et al. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport* 11, 69–74. doi: 10.1097/00001756-200001170-00014

Sato, W., and Yoshikawa, S. (2007). Enhanced experience of emotional arousal in response to dynamic facial expressions. *J. Nonverbal Behav.* 31, 119–135. doi: 10.1007/s10919-007-0025-7

Slepian, M. L., Weisbuch, M. R. A. Jr., and Ambady, N. (2011). Gender moderates the relationship between emotion and perceived gaze. *Emotion* 11, 1439–1444. doi: 10.1037/a0026163

Tomelleri, S., and Castelli, L. (2012). On the nature of gender categorization: pervasive but flexible. *Soc. Psychol.* 43, 14–27. doi: 10.1027/1864-9335/a000076

Valdés-Conroy, B., Aguado, L., Fernández-Cahill, M., Romero-Ferreiro, V., and Diéguez-Risco, T. (2014). Following the time course of face gender and expression processing: a task-dependent erp study. *Int. J. Psychophysiol.* 92, 59–66. doi: 10.1016/j.ijpsycho.2014.02.005

Wang, S., Wenhui, L., Bo, L., Xiaoyu, C., Ying, L., and Jiang, Z. (2016). ERP comparison study of face gender and expression processing in unattended condition. *Neurosci. Lett.* 618, 39–44. doi: 10.1016/j.neulet.2016.02.039

Wild-Wall, N., Dimigen, O., and Sommer, W. (2008). Interaction of facial expressions and familiarity: ERP evidence. *Biol. Psychol.* 77, 138–149. doi: 10.1016/j.biopsycho.2007.10.001

Wu, B., Zhang, Z., and Zhang, Y. (2015). Facial familiarity modulates the interaction between facial gender and emotional expression. *Acta Psychol. Sin.* 47, 1201–1212. doi: 10.3724/SP.J.1041.2015.01201

Zebrowitz, L. A., Kikuchi, M., and Fellous, J. M. (2010). Facial resemblance to emotions: group differences, impression effects, and race stereotypes. *J. Pers. Soc. Psychol.* 98, 175–189. doi: 10.1037/a0017990

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 30 May 2017
doi: 10.3389/fpsyg.2017.00832

"You Should Have Seen the Look on Your Face. . .": Self-awareness of Facial Expressions

Fangbing Qu[1,2,3]*, Wen-Jing Yan*[4]*, Yu-Hsin Chen*[4]*, Kaiyun Li*[5]*, Hui Zhang*[6] *and Xiaolan Fu*[2,3]*

[1] *College of Preschool Education, Capital Normal University, Beijing, China,* [2] *State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China,* [3] *Department of Psychology, University of Chinese Academy of Sciences, Beijing, China,* [4] *Institute of Psychology and Behavioral Sciences, Wenzhou University, Wenzhou, China,* [5] *School of Education and Psychology, University of Jinan, Jinan, China,* [6] *Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, United States*

The awareness of facial expressions allows one to better understand, predict, and regulate his/her states to adapt to different social situations. The present research investigated individuals' awareness of their own facial expressions and the influence of the duration and intensity of expressions in two self-reference modalities, a real-time condition and a video-review condition. The participants were instructed to respond as soon as they became aware of any facial movements. The results revealed that awareness rates were 57.79% in the real-time condition and 75.92% in the video-review condition. The awareness rate was influenced by the intensity and (or) the duration. The intensity thresholds for individuals to become aware of their own facial expressions were calculated using logistic regression models. The results of Generalized Estimating Equations (GEE) revealed that video-review awareness was a significant predictor of real-time awareness. These findings extend understandings of human facial expression self-awareness in two modalities.

Keywords: self-awareness, facial expression, awareness rate, duration, intensity

OPEN ACCESS

Edited by:
Ping Hu,
Renmin University of China, China

Reviewed by:
Linda Isaac,
Stanford University School of Medicine, United States
Chao Liu,
Beijing Normal University, China

****Correspondence:***
Xiaolan Fu
fuxl@psych.ac.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *09 February 2017*
Accepted: *08 May 2017*
Published: *30 May 2017*

Citation:
Qu F, Yan W-J, Chen Y-H, Li K, Zhang H and Fu X (2017) "You Should Have Seen the Look on Your Face. . .": Self-awareness of Facial Expressions. Front. Psychol. 8:832. doi: 10.3389/fpsyg.2017.00832

INTRODUCTION

At some point in our lives, we are usually confronted with a situation in which someone says, "You should have seen the look on your face. . .." One typically attempts to recall one's facial expression and ponders, "What was the look on my face?" to assess whether the facial expression expressed was appropriate in accordance with social norms, such as the feeling rules (Hochschild, 1979) and display rules (Ekman and Friesen, 1971). An accurate interpretation of one's facial expression is important in every interpersonal interaction because a considerable amount of information about one's affective state, status, attitude, cooperativeness, and competitiveness in social interactive situations is expressed and communicated to others through facial expressions (Ekman and Friesen, 1971; DePaulo, 1992; North et al., 2010, 2012). The misappraisal of facial expressions that we display to other people may have important consequences and may influence the course of the interaction. To prevent and mitigate the chances of misinterpreting our facial expressions, we need to possess a certain amount of emotional self-awareness, that is, what is expressed in our daily interactions with others (Hess et al., 2004).

Psychologists generally agree that individuals are experts at monitoring and perceiving their own emotional states and are capable of providing more accurate self-reports of their subjective experience of emotions and bodily experience than most other individuals could (Barrett et al., 2007; De Vignemont, 2014). Ansfield et al. (1995) noted an interesting dilemma in which we are rarely able to observe our own facial expressions, although others can see them. Hence, we often hear people say, "You should have seen the look on your face...." On many occasions, previous studies have noted discrepancies in humans' subjective experience of their facial expressions. Riggio et al. (1985) assessed individuals' self-perceived emotion-sending abilities by asking participants to express six basic emotions and rate their perceived success during the emotion-sending task. They observed that participants' self-perceived emotion-sending ability was not significantly correlated with their actual emotion-sending ability. Barr and Kleck (1995) first videotaped participants' facial expressions and asked them to rate their own facial expressiveness. Then, participants were shown the videotapes of their facial expressions. In their study, participants expressed surprise toward the inexpressiveness of their faces. This study inferred that people have stronger awareness of sensorimotor feedback but have weak facial display. These observations about humans' subjective experiences of facial expressions raise an interesting question that warrants further investigation regarding the extent of individuals' awareness of their own facial expressions. The literature offers very few studies in which researchers have directly investigated individuals' awareness of their own facial expressions, including real-time awareness (referring to participants' immediate self-reports of the occurrence of their facial expressions) and video-review awareness (referring to the extent to which participants can identify any facial movements in their face recordings).

Based on previous literature on human emotion, a crude conception of individuals' awareness of facial expressions can be proposed. Specifically, sensory feedback of sufficient strength is required for individuals to become aware of their facial expressions. Support for this notion is provided by a statement posited by Ekman: "While there are sensations in the face that could provide information about when muscles are tensing and moving, my research has shown that most people don't make much use of this information. Few are aware of the expressions emerging on their face until the expressions are extreme" (Ekman, 2009). Tomkins's (1962) facial feedback hypothesis posited that the responses of the affected motor and glandular targets (the face, primarily) supply sensory feedback to the brain, which is subjectively experienced as emotion if it reaches consciousness. It can be inferred from this passage that an individual may only become aware of his or her facial expression and emotion if and only if sensory feedback from the facial muscles is strong enough to reach consciousness. These findings and other studies further suggest that awareness of facial expressions may be influenced by factors such as facial expression intensity, duration, and frequency (Ekman et al., 1980; Adelmann and Zajonc, 1989). Although direct evidence in support of this claim is scarce at this time, numerous studies investigating facial expressions and their recognition may provide indirect evidence in support of this notion. In previous facial expression recognition studies, manipulation of facial expression duration was observed to influence individuals' recognition performance; specifically, the recognition rate decreased as a function of expression duration (Shen et al., 2012). Studies have also manipulated the intensity of facial expressions and observed that facial expressions with a higher intensity are recognized at a higher rate (Herba et al., 2006). Heuer et al. (2010) also investigated the detection and interpretation of emotional facial expressions by employing facial morphing paradigm and experimentally manipulated viewing conditions for emotion processing. Their results suggested as the facial expression intensity developing slowly, non-anxious controls and socially anxious individuals show different capacity of emotion onset perception, decoding accuracy, and interpretation. If facial expression duration and intensity influence one's recognition performance, we hypothesize that self-awareness of facial expressions may also be influenced by facial expression intensity and duration.

We attempt to study self-awareness of facial expressions and the influential factors from two modalities. We further propose a real-time monitoring and video-review paradigm to investigate the awareness rate of a subject based on information from different modalities, specifically, somatosensory feedback information from bodily sensory feedback and visual information from the individual's facial expression recordings in the video-review condition. In the real-time monitoring condition, participants were instructed to press a key on a keyboard the moment they felt facial movement and then return their facial expression to a neutral state. In the video-review condition, participants were asked to identify any facial movements in their video clips recorded in the real-time condition and press the pause button.

The present study investigated the extent to which people are self-aware of their facial expressions under real-time and video-review conditions and the influencing factors. We hypothesized that awareness of facial expressions would be influenced by both duration and intensity. The present study also sought to calculate the duration and intensity thresholds required for individuals to become aware of their own facial expressions. In addition, we intended to investigate the potential relationship between real-time and video-review awareness.

MATERIALS AND METHODS

Participants

Twenty-seven participants who were naïve to the study's objectives were recruited (16 females; mean age = 22.59 years, *SD* = 2.17). All participants signed an informed consent form and were told that they had the right to terminate the experimental procedure at any time. The data from four participants were excluded due to technical issues (i.e., camera failed to record because of insufficient memory space). Because we conducted the study during the last month of the academic school year, our data-collection aims were modest – to collect data from at least 25 students and finally collect at least 300 facial expression sample with emotional meanings. The institutional review board (IRB) of the Institute of Psychology, Chinese Academy of Sciences approved the study protocol.

Apparatus

An Open CV (Open Source Computer Vision Library)-based program was developed to record the time of participants' key press with relative accuracy and precision during the presentation of each emotional video while simultaneously controlling a high-speed video camera (Logitech Pro C920, recording at 60 fps) that captured participants' facial activities during each video.

After each emotional video was presented, the program stopped recording and saved the clip containing the participants' facial activities during the stimulus presentation to the hard drive.

Materials

Nine videos (5 meant to elicit happiness, 2 meant to elicit disgust, and 2 meant to elicit anger) were selected (see **Table 1**). Seven of the nine videos were chosen from a previous study (Yan et al., 2013), and the other 2 were chosen from the Internet. Twenty additional participants who did not participate in the formal experiment rated the videos by choosing one or two emotion keywords from a list and rated their intensity on a 7-point Likert scale (where 1 denoted not intense and 7 denoted very intense). If words belonging to a certain basic emotion (e.g., happiness) were chosen by one-third of the participants or more, that emotion was considered the main emotion(s) of the video (see **Table 1**) (Yan et al., 2013). We chose happiness, disgust, and anger as the target emotions because facial expressions related to these emotions have been observed to be readily elicited in studies that employed the neutralizing paradigm (Porter and ten Brinke, 2008; Yan et al., 2013). Each video lasted 1–2 min (see **Table 1**). Volume was fixed and controlled across participants.

Procedure

The participants sat at a table facing a 19-inch, color LCD monitor. A high-speed video camera on a tripod was placed behind the monitor to record the frontal view of the participant's face.

Phase 1: Real-Time Self-monitoring

All participants were tested individually. To obtain "uncontaminated" leaked fast facial expressions that are uncorrupted by various unemotional facial movements (e.g., speaking, blowing the nose and pressing the lips), a facial neutralization paradigm was used (Yan et al., 2013). Participants were motivated to neutralize their faces while they watched high-arousal emotional video clips.

TABLE 1 | Participant ratings of the nine video clips.

Clip no.	Duration (min:sec)	Main emotion	Selection rate	Mean intensity score
1	1:07	Disgust	0.86	4.14
2	1:35	Disgust	0.92	4.33
3	1:57	Anger	0.75	4.5
4	2:24	Anger	0.77	3.92
5	1:18	Happiness	0.92	4
6	1:32	Happiness	0.86	3.07
7	1:16	Happiness	0.86	3.28
8	1:48	Happiness	0.73	3.64
9	1:09	Happiness	0.71	3.17

We told the participants that they would view a series of short videos (some of which would be unpleasant) and that we were interested in their ability to control and be aware of their facial movements. The presentation sequence of the experimental videos was counterbalanced across participants. The participants watched the screen closely to maintain a neutral face and were instructed to avoid bodily movement. The participants were also instructed to press the spacebar as soon as they became aware of any facial movement and to return to a neutral face as quickly as possible.

The experimenter monitored the participants' faces on-line from another monitor. This setup helped the experimenter pre-define certain habitual movements to be verified by the participant after each video was shown. After each video, participants were prompted with a question that required them to rate the emotional valence and intensity of the video.

Phase 2: Video Review

The cue-review paradigm proposed by Rosenberg and Ekman (1994) requires participants to watch a replay of a stimulus film and report their emotions at points during the film when they remember having an emotion or an expression on a momentary basis, providing researchers with the opportunity to examine momentary changes in facial expressions of emotion rather than aggregated measures.

After Phase 1 ended, we applied the modified cued-review paradigm by asking participants to review their own facial expressions. Participants were asked to identify any facial movements in these clips and press the pause button. The experimenter recorded the time and asked the participants to recall and verbally report the emotion felt when they displayed the facial expression. These data were used to analyze the awareness rate of the participants' facial expressions in the video-review setting. Participants received only one chance to decide whether a facial expression had occurred, to match the procedure of Phase 1.

Coding

Prior to manual coding, all facial movements detected by participants in phase two were classified into two groups based on the participants' self-reports: (a) facial movements that participants could recall and for which they could report feeling a clear emotion and (b) those verified as habitual movements or facial movements that participants could not recall and for which they could not report feeling a clear emotion. According to the participants' self-reports, we filtered those frames (via Adobe Premiere 6.0) that were confirmed by the participants to be habitual movements or facial movements that participants could not recall and report feeling a clear emotion. These edited clips were given to two extensively trained coders who coded each frame of the clips for the presence and duration of the expressions on the participants' faces. The coding procedures employed were similar to a previous study (Yan et al., 2013).

Coding required the coders to code these expressions' onset times, apex times and offset times by applying the frame-by-frame approach. The apex frame showed the full expression that had a highest intensity for this facial expression. The offset frame was the frame right before a facial movement returned to baseline (Hoffmann et al., 2010). The total duration of these leaked facial expressions was calculated. Coders also rated the intensity of each facial expression from 1 to 7 (where 1 denoted not intense and 7 denoted very intense).

The reliability coefficient of the two coders was calculated according to the equation used in a previous study (Yan et al., 2013). The reliability coefficient of the two coders was 0.82 for all the samples.

RESULTS

All remaining facial expression samples with durations less than 4 s were included for further analysis.

Analysis of Real-Time Facial Expression Self-awareness

Descriptive Statistics of Real-Time Self-awareness

Of the 353 facial expressions detected by trained coders via the frame-by-frame approach, there were 204 facial expressions in which participants expressed awareness with a keyboard response. This result indicated that the participants' awareness rate of their facial expressions during the real-time monitoring phase was 57.79% (see **Table 2**). Among the 204 facial expressions for which participants expressed awareness during the real-time condition (hereafter referred to as "real-time aware facial expression"), there were 14 facial expressions that participants reported being unaware of during the video review phase.

The Relationship between Intensity/Duration of Expression and Real-Time Self-awareness

Forty-nine of the 353 expressions had an apex frame but no offset frame; hence, the total duration could not be obtained for these expressions. The remaining 304 expressions were included in the subsequent analysis.

TABLE 2 | Descriptive statistics for the real-time and post-awareness phases of leaked facial expressions.

		Real-time awareness		Post-awareness	
	N	Count	Rate (%)	Count	Rate (%)
All expressions	353	204	57.79	268	75.92

TABLE 3 | The logistic regression parameter estimations.

Awareness condition	Duration and intensity predict awareness	
	Intensity	Duration
Real-time	0.67***	0.01
Video-review	0.41***	0.03**

*$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

To further explore the relationship between factors such as expression intensity/duration and expression self-awareness in the real-time condition, a logistic regression was employed (McCullagh and Nelder, 1989). The response variable is binary (1/0): aware or unaware under the real-time or video-review condition. The predicting variables are the duration and intensity of the facial expression. **Table 3** shows the logistic regression coefficient estimations.

When both duration and intensity were included in the regression, interesting results arose and suggested that in the real-time condition, the intensity of facial expression is the only significant predictor of facial expression awareness ($\beta = 0.67$, $p < 0.001$), whereas duration is not ($\beta = 0.01$, $p > 0.05$). The results suggested that facial expression intensity is more important for expression self-awareness than duration is in the real-time condition.

Estimated Intensity Threshold in the Real-Time Awareness Condition

Following the logistic regression model fitting, we obtained different estimated equations to explore the variation of awareness on facial expression as a function of facial expression duration and intensity. The estimated logistic regression equation is as follows:

$$\log\left(\frac{p}{1-p}\right) = -2.86 + intensity^{*}0.67 + duration^{*}0.01 \quad (1)$$

Using the regression equations estimated above, we obtained intensity thresholds for different awareness rates (25, 50, 75, and 95%). The logistic regression presents various intensity thresholds needed by the participants for awareness of self-facial expression with changing duration in the real-time condition (**Figure 1**).

Analysis of Video-Review Facial Expression Self-awareness

Descriptive Statistics of Video-Review Facial Expression Self-awareness

Of the 353 facial expressions, there were 268 facial expressions in which participants expressed awareness, estimating the participants' awareness rate of their facial expressions during the video-review phase as 75.92%. Among the 268 facial expressions in which participants expressed awareness during the video-review phase (hereafter referred to as "video-review aware facial expression"), there were 91 facial expressions that participants reported being unaware of during the real-time monitoring phase.

The Relationship between Intensity/Duration of Expression and Video-Review Self-awareness

To investigate the relationship between the intensity/duration of expression and self-awareness in the video-review condition, a similar analysis was conducted on the remaining 304 expressions with a logistic regression (McCullagh and Nelder, 1989). The response variable was binary (1/0): aware or unaware. The predicting variables were the duration and rated intensity of the facial expression.

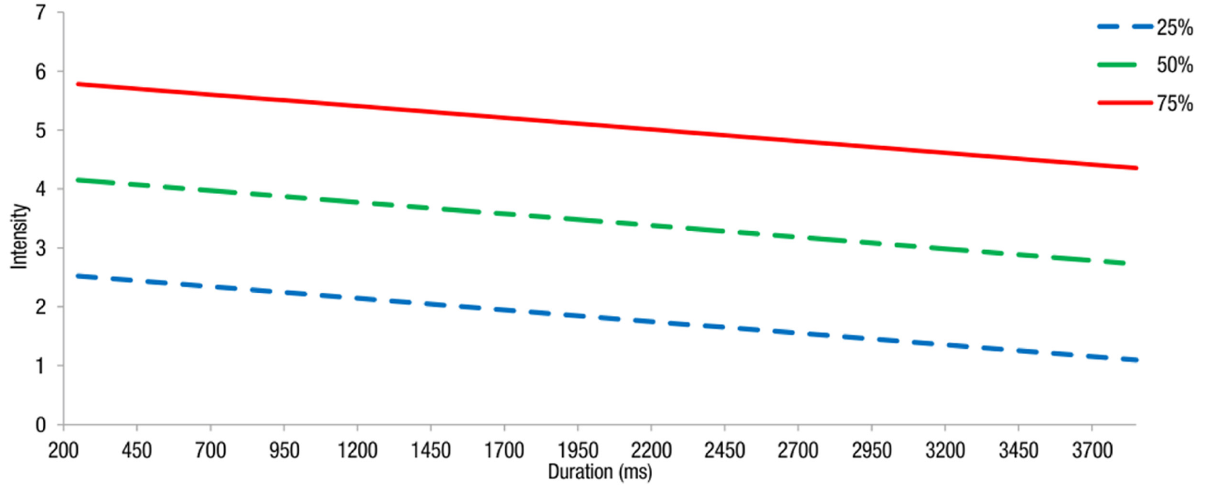

FIGURE 1 | Estimated intensity thresholds needed by the participants for awareness of self-facial expression with changing duration and different awareness rates (25, 50, and 75%) in the real-time condition. The expected intensity level for an estimated awareness rate of 95% was not included because it went beyond the maximum intensity setting (i.e., intensity = 7) in this experiment.

In contrast to the real-time condition, in the video-review condition, both the intensity ($\beta = 0.41$, $p < 0.001$) and the duration ($\beta = 0.03$, $p < 0.01$) were significant predictors of facial expression self-awareness. This result suggests that different mechanisms may exist between the real-time and video-review conditions.

Estimated Intensity Threshold in the Video-Review Awareness Condition

Following the logistic regression model fitting, similar logistic regression equation was obtained to estimate the probability of awareness on facial expression as a function of facial expression duration and intensity.

$$\log\left(\frac{p}{1-p}\right) = -1.004 + intensity^{*}0.41 + duration^{*}0.03 \quad (2)$$

Using the regression equations estimated above, we obtained intensity thresholds for different awareness rates (25, 50, 75, and 95%). The logistic regression presents the various intensity thresholds needed by participants for awareness of self-facial expression with changing duration in the video-review condition (**Figure 2**).

The Relationship between Real-Time and Video-Review Awareness of Facial Expression

Because both real-time and video-review awareness represent the capability of participants to be aware of their own facial expressions, a relationship may exist between these two types of awareness ability. However, no traditional statistical analysis method is available to describe the correlation between these two dichotomous observations, especially because of the inherent bonds of real-time and video-review awareness from the same subject. To address this issue, a novel analysis strategy was proposed. Generalized Estimating Equations (GEE), which was proposed in 1986 and was designed to model repeated measures, has shown great robustness for both the response distribution and various working correlation matrixes (Liang and Zeger, 1986). GEE has been widely and extensively used to model clustered responses, including binary responses (Zhang et al., 2011). Therefore, we applied GEE to investigate the relationship between real-time and video-review awareness. With the observations of each subject taken as a cluster, we fitted real-time awareness as the response variable with a logit link function to investigate to what extent video-review awareness is related to real-time awareness (Zhang et al., 2011).

The results showed a strong relationship between video-review and real-time awareness ($p < 0.0001$), suggesting that video-review awareness is a significant predictor of real-time awareness and that an individual's ability to be aware of facial expressions in the video-review awareness condition may predict one's performance in the real-time condition.

DISCUSSION

Real-Time Self-awareness of Facial Expressions

The accuracy of self-perception has been a long-standing area of study in psychology (Robins and John, 1997). Many theorists have been less than sanguine about people's ability to perceive their behavior objectively (Gosling et al., 1998). However, according to previous physiological studies on individuals' facial expressions (Rinn, 1984; Proske and Gandevia, 2012), some researchers are confident in individuals' ability to be aware of their facial expressions. Our results reveal that the average real-time awareness rate of all 353 leaked facial expressions was 57.79%. Individuals mainly rely on internally generated

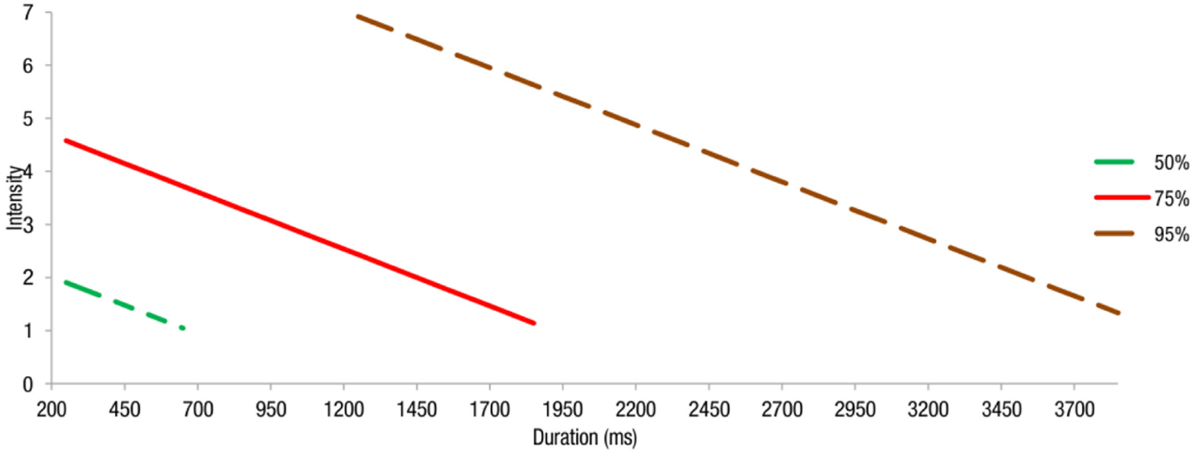

FIGURE 2 | Estimated intensity thresholds needed by the participants for awareness of self-facial expression with changing duration and different awareness rates (50, 75, and 95%) in the video-review condition. The expected intensity level for an estimated awareness rate of 25% was not included because it went beyond the minimum intensity setting (i.e., intensity = 1) in this experiment.

somatosensory feedback to form a subjective experience of facial expressions during a real-time phase.

From the perspective of self-perception theory, Laird (1984) suggested that the relation between facial expressions and emotional experience is a particular case of the general relation between behaviors and psychic states. It has been proposed that greater facial expressivity is uniformly associated with greater subjective experience both between and within subjects, indicating that the duration and intensity of facial expressions is associated with subjective experiences of self-produced facial expressions (Adelmann and Zajonc, 1989).

Our results are partly consistent with the above notion, revealing that the duration and intensity of facial expressions are associated with the subjective experience of self-produced facial expressions. To be more specific, in the real-time condition, only the intensity of facial expressions was a significant predictor of self-awareness, whereas duration was not. This result is consistent with Tomkins and Ekman's statements (Tomkins, 1962; Ekman, 2009). Furthermore, estimated intensity thresholds using logistic regression models show that at certain durations, a higher intensity threshold is needed for a higher awareness rate. This result may be explained by the information the participants employed. In the real-time condition, the participants mainly depended on somatosensory feedback information, such as muscle and skin sensations, to assist their awareness of facial expressions. Other indirect evidence has shown that facial expressions with high intensity are more easily recognized at a higher rate (Herba et al., 2006), which further indicates that intensity, not duration, is the only significant predictor of facial expression self-awareness.

Video-Review Self-awareness of Facial Expressions

The task in the video-review phase was actually a visual facial change detection task, which was mainly dependent on the visual information from one's own facial recordings. In the video-review condition, the awareness rate was 75.92%. Various factors have been demonstrated to be related to the accuracy of change detection, such as stimulus duration, interstimulus interval (ISI), intervening masks, and familiarity (Pashler, 1988). Other studies that have employed facial expression recognition have also found that the recognition accuracy rate was related to facial expression intensity (Herba et al., 2006) and duration (Shen et al., 2012).

In the video-review condition, our result was consistent with previous studies and showed that both the intensity and the duration of facial expressions were significant predictors of self-awareness. Furthermore, estimated intensity thresholds using logistic regression models show that at certain durations, a higher intensity threshold is necessitated for a higher awareness rate. Both intensity and duration information are important when the task is to detect changes that occur in participants' faces in videos.

The Relationship between Real-Time and Video-Review Awareness of Facial Expressions

Video-review awareness was relatively higher than real-time awareness (75.92% for video-review awareness and 57.79% for real-time awareness). This finding might indicate that awareness based on visual information is more advantageous. Several possibilities might produce the difference between these two conditions. First, in the real-time awareness condition, awareness of facial expressions was spontaneous and instant; the participants were required to monitor their facial expressions and to give their reports immediately when they felt their expression change due to the emotional movie. They possessed immediate somatosensory feedback, but visual information was not available; they could not see their own facial expressions. In the video-review condition, awareness was not spontaneous, and the participants watched their own facial expression videos without a sense of manipulation and muscle action. Second, in the real-time awareness condition, participants needed to watch the elicitation movies while monitoring their facial movements. In the video-review condition, participants only needed to watch their facial recordings and detect any changes in their faces visually. Subjects were more focused in the video-review condition, and the cognitive load was relatively lower. Third, the participants may have had post somatosensory feedback and sensory memory in the video-review condition, which may have facilitated awareness of the same expressions, thus outperforming the real-time condition.

Despite the differences addressed above, both of the awareness conditions represent individuals' ability to be aware of their own facial expressions. Thus, correlation between these two awareness conditions might exist. Our results suggest a strong relationship between the real-time and video-review awareness of facial expressions. Whether participants were aware or unaware of their facial expressions in the video-review condition significantly predicted their awareness in the real-time condition. As stated in the above passage, after the real-time awareness task, the participants may have had sensory memory of their facial movements. The participants may have used the integrated information from the post somatosensory feedback and visual information in the video-review condition, suggesting that the above two awareness abilities are correlated.

Several limitations should be noted when interpreting our findings. We used a self-report method to study individuals' self-awareness of their facial expressions. However, self-report measurements may be vulnerable to factors such as self-enhancement or other biases, as previous studies have shown (Gosling et al., 1998). Therefore, researchers should be cautious when interpreting and applying these findings. Furthermore, in the current study, we didn't include a baseline mood measure prior Phase 1. This measure could be used to better investigate its influence on the outcome in future research. In addition, participants in our study mainly consisted of university students with a limited age range and the number of participants recruited were rather limited. This somewhat limits the generalizability of the present results and it would be interesting to investigate participants from an older age group in future research, as elderly participants are generally associated with less negative face-emotion processing and physiological changes to emotion are also different to student cohort.

In the present study, considering the huge individual difference in the number of leaked facial expression and thus the awareness rate, we didn't analyze the data from the individual level, which might be one limitation. In the next study, we consider to separate the participants into different groups according to their facial expression expressivity, with subjects show relatively more numbers of facial expression in daily life into high expressivity group and less numbers of facial expression in daily life (e.g., people with poker face) into low expressivity group.

AUTHOR CONTRIBUTIONS

FQ and XF designed the experiment and wrote the manuscript. FQ, W-JY, Y-HC, and HZ performed the experiment and analyzed the collecting data. FQ, KL, Y-HC, and XF revised the manuscript.

FUNDING

This work was supported in part by grants from the National Natural Science Foundation of China (No. 61375009, 31500875), and the National Science Foundation of China (NSFC), and the German Research Foundation (DFG) in project Crossmodel Learning, NSFC 61621136008/DFG TRR-169.

ACKNOWLEDGMENTS

The authors wish to express sincere appreciation to Prof. Zijiang He (Department of Psychological and Brain Sciences, University of Louisville, Louisville) for his insightful suggestions and help for this study.

REFERENCES

Adelmann, P. K., and Zajonc, R. B. (1989). Facial efference and the experience of emotion. *Annu. Rev. Psychol.* 40, 249–280. doi: 10.1146/annurev.ps.40.020189.001341

Ansfield, M. E., DePaulo, B. M., and Bell, K. L. (1995). Familiarity effects in nonverbal understanding: recognizing our own facial expressions and our friends'. *J. Nonverbal Behav.* 19, 135–149. doi: 10.1007/BF02175501

Barr, C. L., and Kleck, R. E. (1995). Self-other perception of the intensity of facial expressions of emotion: do we know what we show? *J. Pers. Soc. Psychol.* 68, 608-618. doi: 10.1037/0022-3514.68.4.608

Barrett, L. F., Mesquita, B., Ochsner, K. N., and Gross, J. J. (2007). The experience of emotion. *Annu. Rev. Psychol.* 58, 373-403. doi: 10.1146/annurev.psych.58.110405.085709

De Vignemont, F. (2014). A multimodal conception of bodily awareness. *Mind* 123, 989–1020. doi: 10.1093/mind/fzu089

DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychol. Bull.* 111, 203-243. doi: 10.1037/0033-2909.111.2.203

Ekman, P. (2009). "Lie catching and microexpressions," in *The Philosophy of Deception*, ed. C. Martin (Oxford: Oxford University Press), 118–133. doi: 10.1093/acprof:oso/9780195327939.003.0008

Ekman, P., and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *J. Pers. Soc. Psychol.* 17, 124-129. doi: 10.1037/h0030377

Ekman, P., Freisen, W. V., and Ancoli, S. (1980). Facial signs of emotional experience. *J. Pers. Soc. Psychol.* 39, 1125-1134. doi: 10.1037/h0077722

Gosling, S. D., John, O. P., Craik, K. H., and Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *J. Pers. Soc. Psychol.* 74, 1337-1349. doi: 10.1037/0022-3514.74.5.1337

Herba, C. M., Landau, S., Russell, T., Ecker, C., and Phillips, M. L. (2006). The development of emotion-processing in children: effects of age, emotion, and intensity. *J. Child Psychol. Psychiatry* 47, 1098–1106. doi: 10.1111/j.1469-7610.2006.01652.x

Hess, U., Senecal, S., and Thibault, P. (2004). Do we know what we show? Individuals' perceptions of their own emotional reactions. *Curr. Psychol. Cogn.* 22, 247–266.

Heuer, K., Lange, W. G., Isaac, L., Rinck, M., and Becker, E. S. (2010). Morphed emotional faces: emotion detection and misinterpretation in social anxiety. *J. Behav. Ther. Exp. Psychiatry* 41, 418–425. doi: 10.1016/j.jbtep.2010.04.005

Hochschild, A. R. (1979). Emotion work, feeling rules, and social structure. *Am. J. Sociol.* 85, 551–575. doi: 10.1086/227049

Hoffmann, H., Traue, H. C., Bachmayr, F., and Kessler, H. (2010). Perceived realism of dynamic facial expressions of emotion: optimal durations for the presentation of emotional onsets and offsets. *Cogn. Emot.* 24, 1369–1376. doi: 10.1080/02699930903417855

Laird, J. D. (1984). The real role of facial response in the experience of emotion: a reply to Tourangeau and Ellsworth, and others. *J. Pers. Soc. Psychol.* 47, 909–917. doi: 10.1037/0022-3514.47.4.909

Liang, K.-Y., and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13–22. doi: 10.1093/biomet/73.1.13

McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models*, Vol. 37. Boca Raton, FL: CRC press. doi: 10.1007/978-1-4899-3242-6

North, M. S., Todorov, A., and Osherson, D. N. (2010). Inferring the preferences of others from spontaneous, low-emotional facial expressions. *J. Exp. Soc. Psychol.* 46, 1109–1113. doi: 10.1016/j.jesp.2010.05.021

North, M. S., Todorov, A., and Osherson, D. N. (2012). Accuracy of inferring self- and other-preferences from spontaneous facial expressions. *J. Nonverbal Behav.* 36, 227–233. doi: 10.1007/s10919-012-0137-6

Pashler, H. (1988). Familiarity and visual change detection. *Percept. Psychophys.* 44, 369–378. doi: 10.3758/BF03210419

Porter, S., and ten Brinke, L. (2008). Reading between the lies: identifying concealed and falsified emotions in universal facial expressions. *Psychol. Sci.* 19, 508–514. doi: 10.1111/j.1467-9280.2008.02116.x

Proske, U., and Gandevia, S. C. (2012). The proprioceptive senses: their roles in signaling body shape, body position and movement, and muscle force. *Physiol. Rev.* 92, 1651–1697. doi: 10.1152/physrev.00048.2011

Riggio, R. E., Widaman, K. F., and Friedman, H. S. (1985). Actual and perceived emotional sending and personality correlates. *J. Nonverbal Behav.* 9, 69–83. doi: 10.1007/BF00987139

Rinn, W. E. (1984). The neuropsychology of facial expression: a review of the neurological and psychological mechanisms for producing facial expressions. *Psychol. Bull.* 95, 52-77. doi: 10.1037/0033-2909.95.1.52

Robins, R. W., and John, O. P. (1997). Effects of visual perspective and narcissism on self-perception: is seeing believing? *Psychol. Sci.* 8, 37–42. doi: 10.1111/j.1467-9280.1997.tb00541.x

Rosenberg, E. L., and Ekman, P. (1994). Coherence between expressive and experiential systems in emotion. *Cogn. Emot.* 8, 201–229. doi: 10.1016/j.biopsycho.2013.09.003

Shen, X.-B., Wu, Q., and Fu, X.-L. (2012). Effects of the duration of expressions on the recognition of microexpressions. *J. Zhejiang Univ. Sci. B* 13, 221–230. doi: 10.1631/jzus.B1100063

Tomkins, S. S. (1962). *Affect, Imagery, Consciousness: The Positive Affects*, Vol. I. New York City, NY: Springer Publishing Company.

Yan, W.-J., Wu, Q., Liang, J., Chen, Y.-H., and Fu, X. (2013). How fast are the leaked facial expressions: the duration of micro-expressions. *J. Nonverbal Behav.* 37, 217–230. doi: 10.1007/s10919-013-0159-8

Zhang, H., Xia, Y., Chen, R., Gunzler, D., Tang, W., and Tu, X. (2011). Modeling longitudinal binomial responses: implications from two dueling paradigms. *J. Appl. Stat.* 38, 2373–2390. doi: 10.1080/02664763.2010.550038

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 21 February 2018
doi: 10.3389/fpsyg.2018.00202

The Dynamic Features of Lip Corners in Genuine and Posed Smiles

Hui Guo[1], *Xiao-Hui Zhang*[2], *Jun Liang*[2] *and Wen-Jing Yan*[2]*

[1] *Wenzhou 7th People's Hospital, Wenzhou, China,* [2] *Institute of Psychology and Behavior Sciences, Wenzhou University, Wenzhou, China*

The smile is a frequently expressed facial expression that typically conveys a positive emotional state and friendly intent. However, human beings have also learned how to fake smiles, typically by controlling the mouth to provide a genuine-looking expression. This is often accompanied by inaccuracies that can allow others to determine that the smile is false. Mouth movement is one of the most striking features of the smile, yet our understanding of its dynamic elements is still limited. The present study analyzes the dynamic features of lip corners, and considers how they differ between genuine and posed smiles. Employing computer vision techniques, we investigated elements such as the duration, intensity, speed, symmetry of the lip corners, and certain irregularities in genuine and posed smiles obtained from the UvA-NEMO Smile Database. After utilizing the facial analysis tool OpenFace, we further propose a new approach to segmenting the onset, apex, and offset phases of smiles, as well as a means of measuring irregularities and symmetry in facial expressions. We extracted these features according to 2D and 3D coordinates, and conducted an analysis. The results reveal that genuine smiles have higher values for onset, offset, apex, and total durations, as well as offset displacement, and a variable we termed Irregularity-b (the *SD* of the apex phase) than do posed smiles. Conversely, values tended to be lower for onset and offset Speeds, and Irregularity-a (the rate of peaks), Symmetry-a (the correlation between left and right facial movements), and Symmetry-d (differences in onset frame numbers between the left and right faces). The findings from the present study have been compared to those of previous research, and certain speculations are made.

Keywords: dynamic features, genuine smiles, posed smiles, lip corners, OpenFace

OPEN ACCESS

Edited by:
Xunbing Shen,
Department of Psychology, Jiangxi University of Traditional Chinese Medicine, China

Reviewed by:
Lynden K. Miles,
University of Aberdeen, United Kingdom
Zizhu Fan,
East China Jiaotong University, China

***Correspondence:**
Wen-Jing Yan
yanwj@wzu.edu.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *14 September 2017*
Accepted: *06 February 2018*
Published: *21 February 2018*

Citation:
Guo H, Zhang X-H, Liang J and Yan W-J (2018) The Dynamic Features of Lip Corners in Genuine and Posed Smiles. Front. Psychol. 9:202. doi: 10.3389/fpsyg.2018.00202

INTRODUCTION

Among the various interpersonal social signals, facial expressions are one of the most frequently used to express social intentions. In human social interactions, the smile is the most common. A smile typically reflects a happy mood (i.e., a genuine smile), but people often disguise their smiles according to the situation. For example, in greetings and conversations, people may deliberately smile out of politeness (Ambadar et al., 2009; Hoque et al., 2011; Shore and Heerey, 2011). In situations where an individual intends to deceive another, the liar may deliberately display a pleasant expression to mask their ill intent and enhance their credibility, in order to appear amiable (Hoque et al., 2011). Smiles that are not elicited via the genuine subjective experience of happiness are often called deliberate, posed, false, social, polite, or masking smiles

(Ekman and Friesen, 1982; Scherer and Ellgring, 2007; Krumhuber and Manstead, 2009; Mavadati et al., 2013; Gutiérrezgarcía and Calvo, 2015). For the remainder of this article, we will refer to masked or deliberate smiles as "posed" smiles, while spontaneous and authentic smiles will be referred to as "genuine." This will assist us in better clarifying the numerous concepts we will address.

Dynamic Feature Differences between Genuine and Posed Smiles

Previous studies have investigated the differences between genuine and posed smiles, and reported several key indicators and features that may help to differentiate between the two, such as the Duchenne marker, duration, intensity, symmetry, and smoothness.

Duchenne Markers

Duchenne markers are important features of genuine smiles. According to the Facial Action Coding System (Ekman et al., 2002), a Duchenne smile consists of Action Units (AU) AU6 and AU12. AU6 indicates a contraction of the orbicularis muscle, which represents the lifting of the cheek muscles and leads to the formation of crow's feet (Ekman et al., 1988). AU12 indicates a contraction of the zygomaticus muscle, which extends the corners of the mouth sideways and lifts the lip corners up to form a prominent U-shape or happy face. According to previous studies, a smile should only be considered genuine when both AU6 and AU12 occur simultaneously (Ekman, 2006). Frank and Ekman (1993) speculated that most people are able to voluntarily control AU12; in contrast, only a small percentage of people (20%) are able to voluntarily regulate AU6. Hence, it was concluded that AU6 would be a better indicator of a genuine smile, and this later became known as the "Duchenne marker."

Researchers have often observed these Duchenne markers in individuals presented with pleasant stimuli (Ekman et al., 1990; Soussignan and Schaal, 1996) and/or in those providing self-reports that indicate a pleasant state of mind (Ekman et al., 1988; Frank et al., 1993). However, there is still significant debate regarding whether AU6 can be considered the gold standard, because other researchers have observed individuals behaviorally controlling their facial expressions and voluntarily expressing Duchenne smiles, even in the absence of pleasant or happy emotions. Krumhuber et al. (2009) found that in genuine smiles, the ratio of Duchenne smiles to non-Duchenne smiles is 70–30%, respectively, while for posed conditions it was 83–17%, respectively. Several other studies reported observing high proportions of posed smiles that fit the Duchenne smile criteria. For example, it was argued in one study that 56% of smiles fitting the posed smile condition met the criteria for Duchenne smiles (Schmidt et al., 2006). Other researchers have observed up to 60% (Gosselin et al., 2002), 67% (Schmidt and Cohn, 2001), and 71% (Gunnery et al., 2013) of posed smiles fitting the Duchenne smile criteria. These findings suggest that the standards for differentiating and recognize genuine smiles may indeed require more than Duchenne markers.

Other studies have also questioned whether Duchenne smiles actually indicate whether an individual is experiencing a pleasurable or happy emotion. It was reported that smiles fitting the Duchenne smile criteria were observed when participants watched videos designed to elicit negative emotions (Ekman et al., 1990), as well as when they failed in a game (Schneider and Josephs, 1991). These findings suggest that AU6 is more likely to reflect emotional intensity than pleasant or positive emotions, a speculation that is supported by studies that observed negative expressions such as sadness and pain also incorporating AU6 (Bolzani-Dinehart et al., 2005). Krumhuber and Manstead (2009) suggested that Duchenne markers might indicate intensity rather than pleasant or positive emotions. In their study, participants were asked to score the force of Duchenne and non-Duchenne smiles. The results revealed that the intensity score for a "Duchenne smile" (on a scale of 1 to 5, with 1 representing very weak and 5 indicating very strong) was significantly higher ($M = 3.11$) than a "non-Duchenne smile" ($M = 0.97$). Based on these and other findings, it would seem that Duchenne markers alone might not be a sufficient indicator of genuineness in a smile; hence, practitioners and layperson alike have begun to seek other indicators that could better aid in differentiating between genuine and posed smiles.

Duration

Genuine and posed smiles tend to differ in duration. According to the Investigator's Guide for FACS, onset time is defined as the length of time from the start of a facial expression to the moment the movement reaches a plateau where no further increase in muscular action can be observed. Apex time is the duration of that plateau, and offset time is the length of time from the end of the apex to the point where the muscle is no longer in action (Ekman et al., 2002). The total duration of a genuine smile can range from 500 to 4,000 ms, while posed smiles can be either longer or shorter (Ekman and Friesen, 1982). Previous studies have reported differences in duration for the various phases of genuine and deliberate expressions. Accordingly, an expression can typically be divided into the onset, apex, and offset phases, wherein the subjective experiences of emotions elicit and form the onset. If the emotional experience is sufficiently intense, it creates and possibly prolongs the apex phase. As the subjective experience of the emotion subsides, the activated facial muscles gradually return to a relaxed state or neutral expression, which marks the end of the offset phase and typically signals the end of the facial expression. Genuine smiles tend to have a slower onset speed and longer onset duration than posed smiles (Hess and Kleck, 1990; Schmidt et al., 2006, 2009). According to Ekman (2009), very brief (<0.5 s) or very long (>5 s) durations of expression occur more often in deliberate rather than spontaneous expressions. As for onset phase duration, genuine smiles can range between 0.5 and 0.75 s. Solitary spontaneous smiles have an average onset duration of 0.52 s, and spontaneous smiles produced in a social context average an onset duration of 0.50, 0.59, and 0.67 s (Schmidt et al., 2003, 2006, 2009; Tarantili et al., 2005).

Speed is yet another commonly investigated parameter. In previous research, the smile samples investigated were generally of high intensity or fully expressed. Schmidt et al. (2006) found that onset and offset speeds and offset duration were all greater in posed smiles. Likewise, Cohn and Schmidt (2004) reported that posed smiles had faster onsets. Finally, Schmidt et al. (2006) observed greater onset and offset speeds, amplitude (displacement) of movement, and offset duration in posed smiles.

Intensity

According to FACS and numerous previous studies, the intensity of a facial expression ranges from A (a trace) to E (a full-blown expression); an alternative scale ranges from 1 (of weak intensity) to 5 (of very strong intensity). Accordingly, Krumhuber et al. (2009) investigating the intensity of Duchenne markers in Duchenne and non-Duchenne smiles, reporting that the intensity ($M = 3.07$) of the Duchenne smiles was greater than that of the non-Duchenne smiles ($M = 1.77$). This finding was replicated in research conducted by Krumhuber and Manstead (2009), where participants either smiled spontaneously in response to an amusing stimulus (a spontaneous condition) or were instructed to pose a smile (a deliberate condition). Coders were then tasked to rate and record the highest intensity AU12 motions for each facial expression. The results revealed that Duchenne smiles were rated as more intense ($M = 3.11$) than non-Duchenne smiles ($M = 0.97$). Interestingly, this study also indicated that deliberate Duchenne smiles were rated as more intense ($M = 3.37$) than spontaneous Duchenne smiles ($M = 2.85$). These finding suggests that individuals who fake smiles are capable of behaviorally controlling their facial movements, and tend to express exaggerated smiles that are more intense than genuine smiles. In line with these findings, Schmidt et al. (2006) observed that the amplitude (displacement) of movement was greater in deliberate smiles.

Symmetry

Genuine and posed smiles may also have different features with regards to symmetry (Frank and Ekman, 1993; Frank et al., 1993). When asymmetries occurred in posed smiles, they were usually stronger on the left side of the face. Ekman et al. (1981) videotaped children spontaneously making happy faces that were elicited by jokes or encouragement; these were then compared to posed happy faces. Smiles formed in response to watching an amusing film were nearly always symmetrical (96%); expressions in response to negative emotions from watching unpleasant films were, for the most part, also symmetrical (75%). A meta-analysis revealed that this asymmetry was stronger for posed than spontaneous emotional expressions (Skinner and Mullen, 1991). A later review (1998) of 49 experiments shows that posed and spontaneous expressions did not differ in the direction of facial asymmetry, unlike clinical observations indicating that spontaneous expressions showed more bilaterality Borod et al. (1998). A more recent review (Powell and Schirillo, 2009) described non-clinical studies and suggested that there actually is facial asymmetry, with emotions being expressed more on the left side of the face than on the right in both spontaneous and posed expressions. However, several studies (Schmidt et al., 2006, 2009; Ambadar et al., 2009) utilized computer vision techniques to measure the displacement of action units associated with smiling, and observed no differences in the asymmetry of intensity (i.e., the amplitude) between genuine and posed expressions.

Regarding temporal asymmetries, Ross and Pulusu (2013) employed high-speed cameras to isolate time features and examine asymmetries in genuine and posed expressions. Posed expressions overwhelmingly originated on the right side of the face, whereas spontaneous expressions began most often on the left. In the upper half of the expressions, this pattern was particularly stable. In another study, however, Schmidt et al. (2006) found no differences between genuine and posed smiles in terms of asymmetries in onset or offset duration. A number of other researchers also found no difference.

Irregularity (Smooth)

The degree of irregularity in genuine and posed smiles may also differ. Some facial expressions are very irregular; an apex may be steady or there may be noticeable changes in intensity before the offset phase begins (Ekman et al., 2002). The degree of irregularity refers to whether there are pauses or discontinuous changes in the phases of the expression (e.g., onset, apex, offset). Although this varies with the particular social circumstances, the onset of a deliberate expression will often be more abrupt than that of a spontaneous expression (Ekman, 2006). Hess and Kleck (1990) defined irregularity as the number of onsets and offsets throughout the entirety of the expression, and found that genuine expressions are more regular than those that are posed. Frank et al. (1993) used a different definition, smoothness, in their study. Smooth refers to the degree of positive correlations among the durations of the onset, apex, offset, and complete expression. This study found that genuine smiles (those with AU6) were smoother than posed expressions.

Analyzing Facial Expressions Using Computer Vision Techniques

Because an historic lack of easy-to-use quantitative analysis tools (Frank et al., 2009), only a handful of studies on the dynamic characteristics of expression (such as duration, velocity, smoothness, motion symmetry, synchronization of different parts, etc.) have been conducted. With the development of computer vision and pattern recognition techniques, researchers have begun to employ new analysis tools to further study facial expressions. For example, by tracking the various parts of the face over time, they have been able to witness gradual changes in intensity for each phase, the symmetry of synchronization of left and right movements, and so on.

Considering the difficulties in manual coding using FACS, computer researchers have been working on developing new and better face analysis tools. The analysis of facial expressions generally involves three steps: detecting the face in a picture or video, extracting the facial features, and recognizing and classifying those features. The field of computer vision focuses on how to accurately classify different expressions (Pantic and Patras, 2006; Sebe et al., 2007) and AU (Cohn and Sayette, 2010; Littlewort et al., 2011; Wu et al., 2012; Mavadati et al., 2013). From a psychological perspective, researchers are more interested in how particular dynamic features distinguish different smiles. Therefore, the focus is on the feature extraction method and quantitative analysis of the facial movement. Here, feature extraction refers to the use of computers to extract image information, in order to determine whether the points in each image belong to a particular image feature. In other words, this process looks for image information (such as edges, corners, textures, etc.) that is specific to the original feature (a number of pixels). Feature extraction methods are mainly divided into two categories: geometric feature-based approaches and appearance-based methods (Tian et al., 2011). A system based on geometric features extracts the shape and position of the facial composition (such as the mouth, eyes, eyebrows, nose, etc.); an appearance-based methodology uses visual features to represent the object. These two types of processes have different levels of performance in extracting different features, but the merits of the performances are uncertain.

Schmidt et al. (2006) used the CMU / Pitt Automatic Facial Analysis System (AFIA) to measure the movement characteristics of the large zygomatic muscle during genuine and posed smiles. This system automatically fits the landmarks on the first frame of the video clip, and then uses the Lucas-Kanade optical flow (OF) algorithm to track the feature points, after correcting for head motion. The algorithm tracks a pixel on the first frame of the image and determines the position of that pixel on subsequent images, in order to determine the pixel's coordinate changes. The intensity is defined by the moving distance of the feature points, divided by the width of the mouth. With further calculations, the duration, displacement, and velocity of the mouth movement can all be quantified. The results of Schmidt et al. (2006) indicate that compared to posed expressions, mouth movement during onset and offset were shorter and faster in genuine smiles, but there was no difference in symmetry between the two.

Dibeklioğlu et al. (2012) extracted 25 descriptors (features) to train a classifier, as is common practice with computer vision researchers, in order to distinguish genuine from posed smiles; these descriptors included duration, duration ratio, maximum and mean displacements, the SD of the amplitude, total and net amplitudes, amplitude ratio, maximum and mean speeds, maximum and mean accelerations, net amplitude, duration ratio, and left/right amplitude difference for three different face regions (eyes, cheeks, and mouth). After the feature extractions, the researchers trained a classifier that attempted to recognize whether a given video was genuine or posed.

Yan and Chen (in press) tried to quantify micro-expressions using the Constraint Local Model (CLM) and Local Binary Pattern (LBP) methods. The CLM process detects 66 feature points for each face image and tracks the movement and distance for each of these landmarks. These feature points are distributed on the contours of the head, eyes, nose, and mouth. The dynamic features of these landmarks are then described by calculating their position changes over time. Based on the feature points, Yan and Chen (in press) divided the facial area into 16 areas of interest (such as the insides of the eyebrows, which is Interest Area 1), and extracted their texture features using LBP. By comparing the correlations among the textures of the first and subsequent frames, the motion features could be described. The researchers then tested the effects of these two feature extraction methods on 50 micro-expressions, finding that they were similar to manual coding when determining the peak frame.

The Aim of This Work

Previous research has considered the Duchenne mark (AU6), duration, symmetry, irregularity, and other clues in order to investigate the differences between genuine and posed smiles. However, while some indicators have inspired a fairly stable consensus, others have continued to be controversial. We employed a newly-developed analytical tool to investigate specific movements of the mouth and lip corners, which are the most prominent and easily posed in a smile. In Dibeklioğlu et al. (2012), 25 features were considered. Many of these features were difficult to explain from a psychological perspective. Based on this previous research, we extracted duration, speed, intensity, symmetry, and irregularity.

We conducted feature extractions to produce 2D and 3D coordinates with OpenFace, in order to investigate how certain dynamic features of the lip corners differed between genuine and posed smiles. Overall, we hypothesized that genuine smiles would be of longer duration, slower speed, and lower intensity; we also explored the differences in irregularity and symmetry between the two types of smiles.

METHODS

Materials

We used the UvA-NEMO Smile Database (Dibeklioğlu et al., 2012) to analyze the dynamics of genuine and posed smiles of enjoyment. The database consists of 1,240 smile videos (597 spontaneous and 643 posed) obtained from 400 subjects (185 female and 215 male), making it the largest smile database in the literature, to date. The ages of the subjects varied from 8 to 76 years, with 149 subjects being younger than 18 years (offering 235 spontaneous and 240 posed smiles). Of the total, 43 subjects did not have spontaneous smile samples and 32 had no posed smiles. The videos are in RGB color and were recorded at a resolution of 1,920 × 1,080 pixels, at a rate of 50 frames per second, and under controlled illumination conditions.

For the posed smiles, each subject was asked to posture an enjoyment smile as realistically as possible, after being shown a sample video of a prototypical smile. This differed from the samples in Schmidt et al. (2006), where the spontaneous smiles were not the result of any specific elicitation procedure. These genuine smiles of enjoyment were elicited by a set of short, funny video segments shown to each subject for approximately 5 min. The mean duration of the spontaneous and posed smile segments was 3.9s ($\sigma = 1.8$), and the average interocular distance from the database was approximately 200 pixels. The segments all began and ended with neutral or near-neutral expressions. This is considered a well-established database that not only contains a large sample size, but also offers a well-designed lab situation and well-set elicitation approach.

TABLE 1 | Definitions of the features extracted for a single facial expression.

Feature	Definition
Duration	$\frac{F(S)}{r}, \frac{F(S^{+})}{r}, \frac{F(S^{=})}{r}, \frac{F(S^{-})}{r}$
Duration Ratio	$\frac{F(S^{+})}{F(S)}, \frac{F(S^{=})}{F(S)}, \frac{F(S^{-})}{F(S)}$
Displacement	$max(D), D_{offset}$
Speed	$\frac{\sum D^{+}}{F(D^{+})}, \frac{\sum D^{-}}{F(D^{+})}$
Irregularity	$\frac{P}{F(S)/r}, SD(D^{=})$
Symmetry	$Cor(D_L, D_R), M(D_R^{=}) - M(D_L^{=}), F_{R-onset} - F_{L-onset}, \|F_{R-onset} - F_{L-onset}\|$

Analysis Tool: Openface

OpenFace (Baltrusaitis et al., 2016) is not only the first open source tool for facial behavior analysis, it demonstrates state-of-the art performance in facial landmark detection (Baltrusaitis et al., 2013), head pose tracking, AU recognition (Wood et al., 2015) and eye gaze estimation (Wood et al., 2015). The source code can be downloaded here[1]. OpenFace 0.3.0 provides 2D and 3D spatial landmarks for analyzing faces. In this study, the results from a variety of different landmark systems are examined and discussed.

Design

The independent variable for this research was authenticity: genuine / posed. The dependent variables included: duration, speed, intensity, symmetry, and irregularity (see **Table 1** for details). In the database, each participant provided at least one trial for a genuine or posed condition, so this was considered a within-subject design.

In **Table 1**, S indicates a complete smile. The signals are symbolized with a super-index, and (+), (=), and (–) denote the segments of onset, apex, and offset, respectively. For example, S^{+} pools the onset segments for one smile, N defines the number of frames in a given signal, and r is the frame rate of the video. D defines the displacement (the difference in amplitudes between the fiducial and selected frames) of a given signal. D_L and D_R are the displacements of the left and right lip corners, respectively. P defines the number of peaks (an onset and offset form one peak, but only displacement differences larger than 10% D_{max} between adjacent peaks and valleys were filtered out). We measured the offset displacement (D_{offset}) because from our observations it seemed that spontaneous smiles usually ended with the trace of a smile. Therefore, we hypothesized that the displacement in offset frames would be larger in spontaneous smiles than in those that were posed.

In addition to conventional features such as duration, displacement, and speed, we also examined other dynamic elements such as irregularity and symmetry. For irregularity, we used two indicators: Irregularity-a and Irregularity-b. Irregularity-a defined the number of peaks per second. This was similar to Hess and Kleck's (1990) method, where the onsets and offsets for each facial expression were counted as irregularities. Irregularity-b defined the standard deviation (SD) values for the apex displacements. SD was used to quantify the amount of variation or dispersion of a set of data values. We also used SD to measure changes in the apex phase. If the apex phase was just a plateau, the SD was close to zero; when there was substantial fluctuation in the apex phase, the SD was large. However, it would not have been appropriate to use SD to measure the onset and offset phases, because it would have made it difficult to find any psychological meaning.

For symmetry, we explored four different methods. For simplicity, we labeled them a, b, c, and d. In terms of the lip corners, Symmetry-a, Cor (D_L, D_R), defined the Pearson correlation coefficient, Symmetry-b described the mean displacement differences for the apex phase, and Symmetry-c reflected the onset frame differences from a temporal perspective. Symmetry-d denoted the absolute value of Symmetry-c. Here Symmetry-c and Symmetry-d are the "reversed scoring" index, the larger, the more asymmetric.

Procedure for Using 3D Landmarks

3D Pose Correction

OpenFace uses the recently proposed Conditional Local Neural Fields (CLNF) (Baltrusaitis et al., 2013) for facial landmark detection and tracking. Sixty-eight 3D landmarks are detected in each frame, and the 3D coordinates for each landmark are generated. In addition, OpenFace provides 3D head pose estimations, as well as roll (θz), yaw (θy), and pitch (θx) rotations.

Since head movements may occur along with facial movements, it was essential remove (or control for) the influence of the head pose. In this research, head pose estimations for three directions (or rotations) were transformed into a rotation matrix, and each landmark in the 3D space was corrected by post-multiplying the corresponding rotation matrix.

$$l_i^{t'} = l_i^t R_x(-\theta_x)\, R_y\left(-\theta_y\right) R_z(-\theta_z) \tag{1}$$

where l_i is the aligned landmark and R_x, R_y, and R_z denote the 3D rotation matrices for the given angles. Moreover, to control for the influence of head translation (left–right or up–down movement), we selected one stable point, landmark 34 (indicating the inner nostril), and subtracted the other landmark coordinates from it. In previous studies, inner eye corners were often considered stable and used as the reference. However, in most face alignment tools, inner eye corner landmarks change sharply when eyes blink. Therefore, such a reference is actually unsuitable when there are eye blinks in the facial expression.

$$l_i^{t''} = l_i^{t'} - l_{34}^{t'} \tag{2}$$

Displacement Measurement

After correcting the head pose rotation and translation, we proceeded to divide the smiles into the onset, apex, and offset phases. This phase segmentation relied on the pattern of movement in the lip corners, which is quite conspicuous in a smile (the lip corners pull backward and upward). By tracking the lip corners over time, we were able to gain the lip corner

[1] https://github.com/TadasBaltrusaitis/OpenFace.

coordinates in the world coordinates for each frame. We could then calculate the displacement of the lip corners.

There were several possible ways to describe the displacement of lip corner movements: (1) the initial center point could be calculated as the midway position between the lip corners in the initial frame. This initial center point is then recalculated automatically in each frame, relative to the stable inner eye corner feature points, allowing for accurate measurement in cases of small head movements. The pixel coordinates of the right lip corner in subsequent frames, relative to the initial center point of the lip corners in the initial frame, are then automatically obtained using the Lucas–Kanade algorithm for feature tracking (Lien et al., 2000). The displacement of the right lip corner is considered the indicator of the lip movement. The displacement is standardized for the initial width of each participant's mouth in the initial image.

$$D_{lip}(t) \ = \rho\left(\frac{l_r^t + l_l^t}{2}, l_r^t\right) \tag{3}$$

where l_r^t and l_l^t denote the coordinates of the right and left lip corners in frame *t*, respectively, and ρ is the Euclidean distance between the given points.

(2) Dibeklioğlu et al. (2012) estimated the smile amplitude as the mean amplitude of the right and left lip corners, normalized by the length of the lip. Let *D_{lip}(t)* be the value of the mean displacement signal of the lip corners in frame *t*. This can be estimated as:

$$D_{lip}\left(t\right) = \frac{\rho\left(\frac{l_{49}^1 + l_{55}^1}{2}, l_{49}^t\right) + \rho\left(\frac{l_{49}^1 + l_{55}^1}{2}, l_{55}^t\right)}{2\rho\left(l_{49}^1 \ , l_{55}^1\right)} \tag{4}$$

where l_i^t denotes the 3D location of the *i-th* point (in this research, points 49 and 55 indicated the right and left lip corners, respectively) in frame *t*, and ρ is the Euclidean distance between the given points.

(3) Lip corners movements can be calculated according to changes in each landmark location, across time.

$$D_{lip}\left(t\right) = \frac{\rho\left(l_{49}^1, l_{49}^t\right) + \rho\left(l_{55}^1, l_{55}^t\right)}{2} \tag{5}$$

We used this simple calculation in this research because: (1) the initial middle point was unsteady, due to the movement of the lip corners (e.g., when the two lip corners experienced unbalanced changes, the middle point of the lips deviated); (2) we compared genuine with posed smiles using a within-subject design, which meant that the length of the mouth for the given participant was the same, and thus there was no need for normalization; (3) with head pose (rotation and translation) corrections, the faces from different frames were aligned, and thus the displacement of the lip corners could be calculated according to the coordinates of the lip corners.

Phase Segmentation

Based on the lip corner displacement, we attempted to segment the smiles into three phases: onset, apex, and offset. As Dibeklioğlu et al. (2012) mentioned, the longest continuous increase in D_{lip} is defined as the onset phase. Similarly, the offset phase is the longest continuous decrease in D_{lip}. The phase between the last frame of the onset and the first frame of the offset is the apex. This is a very easy and effective way of segmenting the different phases. However, when the smile movement is not very regular (usually displayed as small peaks in the curves, as seen in **Figure 1**), the segmentation method (the longest continuous increase / decrease) may not be sufficiently accurate. According to FACS, an apex may be steady or there may be noticeable fluctuations in intensity before offset begins. Hess and Kleck (1990) calculated all onset, apex, and offset durations for a single facial expression by the naked eye, with subjective definitions of all three phases. If the facial expression was not smooth (i.e., regular), there were several possible onset durations. In this research, we used the UvA-NEMO Smile Database, where only a single smile was contained in each video episode. We segmented only single smiles with one set of onset, apex, and offset phases.

This study attempted to simplify phase segmentation while considering the irregularity of facial movements. To do so, we proposed a new method for segmenting the phases, as follows:

(1) Smooth the displacement of the lip corners across time using a moving average filter, at a window length of 3.
(2) Find all peaks (there are still peaks or bulges, even after smoothing), including the highest (maximum displacement of lip corners in a given smile, D_{max}).
(3) Define the apex phase as the regions between peaks that are higher than 70% of the D_{max}. Sometimes there is only one peak that is higher than 70% of the D_{max} (this is actually the highest peak); in such cases, the apex phase consists of only one frame.
(4) Define the onset phase as the region between the onset frame and onset-apex boundary. For all of the valleys before the apex phase, the valley that is nearest the apex phase where the displacement is less than 20% of the D_{max} is the onset frame. If there is no such valley, the lowest displacement frame before the apex phase is the onset frame.
(5) Define the offset phase as the region between the offset frame and apex-offset boundary. For all of the valleys after the apex phase, the valley nearest the apex phase where the displacement is less than 20% of the D_{max} is the offset frame. If there is no such valley, the lowest-displacement frame after the apex phase is the offset frame.

Peaks: $D_{(i-1)}–D_i < 0$ and $D_{(i+1)}–D_i \geq 0$
Valleys: $D_{(i-1)}–D_i > 0$ and $D_{(i+1)}–D_i \leq 0$
where D_i indicates the displacement of the *i-th* frame.

Procedure for Using 2D Landmarks

The procedure for the 2D landmark system is quite similar to that of the 3D landmark system. Sixty-eight 2D landmarks were detected in each frame. We selected one stable landmark, landmark 34 (indicating the inner nostril) and subtracted the other landmark coordinates from it. Lip corner movements were calculated according to Equation (3). Based on the level of lip corner displacement, we attempted to segment the smiles into three phases: onset, apex, and offset. Then, the dynamic features of the lip corners during the smiles were extracted.

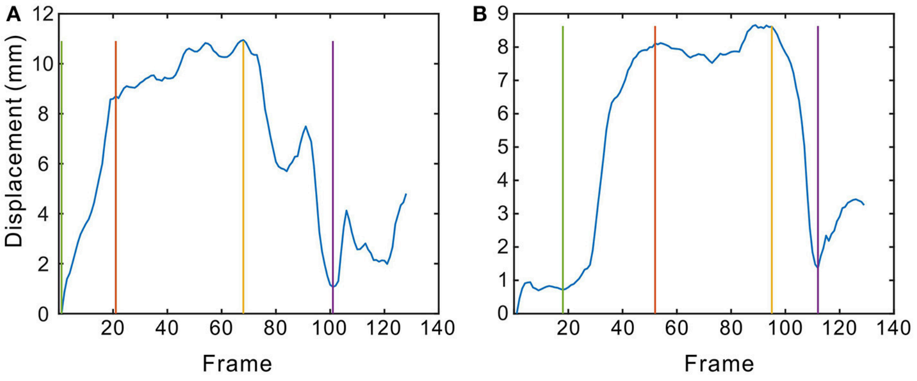

FIGURE 1 | The displacement of the lip corner movements across time. **(A,B)** are two examples. The blue curve indicates that the lip corner displacement changes across time, while the green line denotes onset, the red and yellow lines reference the boundary of the apex phase, and the purple line highlights the offset. The steps taken to complete this segmentation are described in section Phase Segmentation.

DATA ANALYSIS AND RESULTS

Using the 3D Landmark System

After analyzing 1,240 video episodes of smiles from 400 subjects, we extracted 22 features for each smile. Certain smile video episodes were removed based on the following two conditions: (1) the offset phase displacement was larger than that of the apex phase; and (2) the offset phase was less than 0.2 s. These smile video episodes were removed because they tended to exhibit complex facial expressions or be prone to incorrect manual segmentation. As a result, 124 samples were excluded. We aggregated genuine and posed conditions for each subject. Those with only spontaneous or deliberate smiles were also removed. In the end, 297 subjects exhibiting both smile conditions were included for further analysis.

The *p*-value is highly affected by the sample size. In particular, when the sample size approaches 250, the difference / effect is statistically significant regardless of the alpha level (Figueiredo Filho et al., 2013). Also, as Hair et al. (1998) said, "by increasing [the] sample size, smaller and smaller effects will be found to be statistically significant, until at very large sample sizes almost any effect is significant." Due to the large sample size in this study, the F-value could have been inflated, and thus the *p*-values easily influenced. Therefore, we set a strict cut-off point at $p \leq 0.01$, and placed more emphasis on the effect size. Richardson (2011) argued that *partial* η^2 values of 0.0099, 0.0588, and 0.1379 could serve as benchmarks for small, medium, and large effect sizes, respectively.

A repeated measures multivariate analysis of variance (MANOVA) was conducted to examine the effect of the independent variable (authenticity: genuine / posed) on the combined dependent variables (facial features).

A one-way (authenticity: genuine / posed) repeated measures MANOVA was conducted for each participant's facial features. These analyses confirmed that there was a significant multivariate effect for authenticity [$F_{(15,282)} = 37.126$, $p < 0.001$, partial $\eta^2 = 0.664$]. Within-group univariate analyses indicated no differences between the genuine and posed conditions for the following five dependent variables: onset, offset, and apex ratios, Symmetry-b, and Symmetry-c. Significant differences between genuine and posed conditions were observed for the remaining 12 dependent variables [$F_{(1,296)} \geq 8.343$, $p \leq 0.004$, partial $\eta^2 \geq 0.027$]. The onset, offset, apex, and total durations, as well as the offset and standard apex displacements were observed to be significantly higher in the genuine condition. Onset and offset speeds, irregularity-a (rate of peaks), Symmetry-a, and Symmetry-d (smaller values means less asymmetric) were all observed to be significantly lower in the genuine condition. The means, *SD*, *F*, *p*, and partial η^2 values are all shown in **Table 2**.

TABLE 2 | Descriptive and inferential statistics for genuine and posed smiles from the 3D approach.

	Genuine		Posed		*F*	*P*	Partialη^2
	M	*SD*	*M*	*SD*			
Onset Duration	0.93	0.55	0.57	0.24	130.375	<0.001	0.306
Offset Duration	1.10	0.89	0.68	0.34	60.481	<0.001	0.170
Apex Duration	2.97	1.55	1.84	0.74	138.446	<0.001	0.319
Total Duration	5.00	1.98	3.09	0.82	271.829	<0.001	0.479
Onset Ratio	0.20	0.09	0.19	0.08	2.889	0.090	0.010
Offset Ratio	0.23	0.12	0.22	0.10	0.259	0.611	0.001
Apex Ratio	0.57	0.15	0.59	0.12	1.966	0.162	0.007
Offset Displacement	3.04	1.58	2.62	1.22	13.752	<0.001	0.044
Max Displacement	12.19	3.43	12.97	3.05	10.339	0.001	0.034
Onset Speed	13.63	7.43	23.73	10.34	244.854	<0.001	0.453
Offset Speed	9.94	6.35	16.83	9.94	116.720	<0.001	0.283
Irregularity-a	0.78	0.43	0.86	0.36	8.343	0.004	0.027
Irregularity-b	0.89	0.48	0.66	0.38	46.802	<0.001	0.137
Symmetry-a	0.92	0.11	0.96	0.06	32.834	<0.001	0.100
Symmetry-b	2.80	1.98	2.74	2.00	0.209	0.648	0.001
Symmetry-c	−1.28	5.93	−0.49	3.85	3.997	0.046	0.013
Symmetry-d	3.78	5.04	2.23	3.35	22.154	<0.001	0.070

*The order of the variables corresponds to the variable definitions in **Table 1**.*

Using the 2D Landmark System

The data removal criteria were discussed in section Using the 3D landmark system. As a result of this procedure, 96 samples were excluded. All subjects who only had either genuine or posed smiles in the database were removed. The result was that 302 subjects with both conditions were included for further analysis. A repeated measures multivariate analysis of variance (MANOVA) was conducted to examine the effect of the independent variable (authenticity: genuine / posed) on the combined dependent variables (the facial features).

One-way (authenticity: genuine / posed) repeated measures MANOVA analysis was conducted for the participants' facial features. These analyses confirmed that there was a significant multivariate effect for authenticity [$F_{(15,287)} = 43.636$, $p < 0.001$, partial $\eta^2 = 0.695$]. Within-group univariate analyses indicated no differences between the genuine and posed conditions for the following six dependent variables: onset, offset, and apex ratios, max displacement, Symmetry-b, and Symmetry-c. Significant differences between genuine and posed conditions were observed for the remaining 11 dependent variables [$F_{(1,296)} > 11.418$, $p \leq 0.001$, partial $\eta^2 \geq 0.037$]. Onset, offset, apex, and total durations, offset displacement, and Irregularity-b were observed to be significantly higher in genuine smiles. However, onset and offset speeds, Irregularity-a, Symmetry-a, and Symmetry-d were observed to be significantly lower for genuine smiles. The means, *SD*, F, *p*, and partial η^2 values are all shown in **Table 3**.

TABLE 3 | Descriptive and inferential statistics for genuine and posed smiles from the 2D approach.

	Genuine		Posed		*F*	*P*	Partialη^2
	M	*SD*	*M*	*SD*			
Onset Duration	1.16	0.74	0.63	0.26	152.126	< 0.001	0.336
Offset Duration	1.23	0.91	0.79	0.42	61.767	< 0.001	0.170
Apex Duration	2.60	1.44	1.66	0.76	110.330	< 0.001	0.268
Total Duration	5.00	2.00	3.08	0.82	259.848	< 0.001	0.463
Onset Ratio	0.25	0.11	0.21	0.08	22.448	< 0.001	0.069
Offset Ratio	0.26	0.13	0.26	0.12	0.439	0.508	0.001
Apex Ratio	0.50	0.16	0.52	0.15	5.842	0.016	0.019
Offset Displacement	6.97	4.74	5.36	3.85	24.125	< 0.001	0.074
Max Displacement	31.46	9.61	31.87	9.32	0.376	0.540	0.001
Onset Speed	28.93	13.73	53.90	23.53	322.262	< 0.001	0.517
Offset Speed	24.00	17.82	38.12	22.54	82.401	< 0.001	0.215
Irregularity-a	0.59	0.30	0.74	0.31	43.528	< 0.001	0.126
Irregularity-b	2.09	1.31	1.48	0.91	51.397	< 0.001	0.146
Symmetry-a	0.81	0.24	0.87	0.22	11.418	0.001	0.037
Symmetry-b	0.49	8.12	−0.46	9.5	1.951	0.164	0.006
Symmetry-c	0.07	5.28	0.40	3.20	0.956	0.329	0.003
Symmetry-d	3.21	4.42	1.71	2.81	23.419	< 0.001	0.072

The order of the variables corresponds to the variable definitions listed in ***Table 1****.*

DISCUSSION

Lip corner movement (AU12) is the core action unit of a smile; it is easily controlled and posed. This study considered the dynamic features of this action unit in both genuine and posed smiles. We extracted features using 2D and 3D coordinates, and found that the results were quite similar between the two approaches. Only the maximum amplitude was determined to be significant in the 2D method; this value was insignificant in the 3D approach. This research revealed that genuine smiles' onset, offset, apex, and total duration times were significantly longer than those of posed smiles. Genuine smiles also had higher offset displacement and Irregularity-b values (the *SD* of the apex phase) than did posed smiles. In contrast, posed smiles had faster onset and offset speeds. Furthermore, dynamic feature analyses of the left and right lip corners revealed that posed smiles were more asymmetrical than genuine smiles. These findings are discussed below, from a psychological perspective.

Duration

Previous studies have reported that the duration of a facial expression can range from 0.5 to 4 s, and researchers have speculated that the duration of a posed smile is either longer or shorter (Ekman and Friesen, 1982) than one that is genuine. The facial expressions analyzed here lasted about 5 s for genuine smiles inspired by enjoyment, and approximately 3 s for smiles that were posed. Overall, the durations of spontaneous smiles were much longer during the onset, apex, and offset phases. These results are in line with previous findings (Hess and Kleck, 1990; Schmidt et al., 2006, 2009) that reported genuine expressions having slower onset speeds and, in general, longer total durations.

People seemed to be unaware of the longer durations of their spontaneous smiles, because even when asked to pose a smile as naturally as possible, the duration was nearly always much shorter. Thus, we inferred that in their minds, the prototypical pattern of a genuine smile was also much shorter. However, as perceivers, people are able to judge the authenticity of a smile by its duration. Duchenne smiles with longer onset and offset durations were judged to be more authentic than their shorter counterparts (Krumhuber and Kappas, 2005). Yet these researchers determined that the genuineness rating tended to decrease as a function of how long the smile was held at its apex. This conclusion contradicts our findings. One possible reason for this conflict may be that the stimuli in their study were synthesized faces, which could make the lasting-static apex phase wired.

Another explanation for people's inability to accurately simulate genuine smiles may be that they have a static pattern for smiling and ignore the more dynamic features. It's possible that they then pay more attention to morphological features such as the Duchenne marker, a notion that has repeatedly surfaced in previous research and been popularized by mass media (such as the BBC online test[2]). There may also be yet another explanation: the subjects can't hold their muscles in position for the proper length of time without the fuel provided by emotions.

[2] http://www.bbc.co.uk/science/humanbody/mind/surveys/smiles/index.shtml.

To appear more genuine in times of enjoyment, people should begin at a slower pace, hold the smile longer, and fade at a reduced speed in terms of the Duchenne marker.

Intensity

On the maximum displacement of smiles, the 3D approach showed that the intensity was higher in posed smiles, but the difference was insignificant with the 2D approach. It is important to note that this intensity was for lip corner movement and not for the smile itself. It seems counterintuitive that facial smiles elicited from strong emotions would be no more intense than those that are posed. In a posed condition, subjects must expend effort to pull their lips, perhaps even more than is actually needed and especially when displaying large smiles. This may be because they believe that large smiles feature wide lips and humans excel at lip control. Therefore, in a posed condition the intensity of the lip corners appears even larger than in a genuine smile. In genuine smiles and laughter, humans don't appear to pull their lip corners to their fullest possible extent. This conclusion echoes the common experiencing of cheek pain when laughing for an extended period of time, instead of pain emanating from the lip corners.

In this study, we also measured offset displacement, finding higher levels in genuine smiles. This verified our observation that spontaneous smiles usually ended with the trace of a smile. The expression of emotion involves a short and intense process that seems to have an additional, later influence on the expresser's mood. Previous researchers found that presenting emotional stimuli had an effect on subsequent behavior or processing (Barrett et al., 2016). Though strong emotions fade, relevant feelings or moods may linger. For example, after exultation, one does not instantly return to emotional neutrality; instead, a sense of happiness may remain. This type of trace or residual facial expression has not been properly studied. We hypothesize that emotionally elicited facial expressions generally leave a slight trace or a "long tail" after ending. This hint of genuine emotion may appear on the face for some time, even though it is too weak to discern. However, when compared with a neutral face, the subtle differences become obvious. This finding provides a new perspective from which to observe subtle emotional facial expressions.

Symmetry

In this study, we attempted to measure symmetry from four indicators; two were based on intensity, and two on time features. We used a Pearson correlation coefficient to measure the intensity differences between the left and right lip corners (i.e., Symmetry-a). This was a pilot attempt at using a correlation coefficient as an indicator. The advantage was that this indicator considered all of the phases at once. Higher values reflected that the left and right corners moved more synchronically; however, such values didn't reveal intensity differences between the two sides. The asymmetries between genuine and posed smiles were found to be different. Though these results echoed much previous research (e.g., Ekman et al., 1981; Borod et al., 1998; Powell and Schirillo, 2009), Symmetry-a employed a different approach (i.e., the correlation coefficient of the intensity of the lip corners), suggesting that they actually cannot be compared directly with one another.

We also used Symmetry-b, similar to what was employed in previous studies, to calculate the mean displacement of the apex phase of the left and right lip corners. We employed intervals instead of a single point in order to keep this indicator reliable. However, the results showed no differences in this respect. For Symmetry-b, our results closely resembled those of a previous study (Schmidt et al., 2006), wherein no differences were observed between genuine and posed smiles with regards to the asymmetry of intensity (amplitude). It should be noted that different approaches were employed in measuring intensity; the other study used the maximum value of the amplitude (a single point) to indicate the intensity of a smile, while we took the mean value of the apex phase as the indicator. Considering that output values from computer vision algorithms are usually unsteady (i.e., they may feature a certain amount of noise), we used the mean instead of a single value.

Symmetry-c and Symmetry-d measured the onset frame differences from a temporal perspective. Symmetry-c reflected the onset frame gap between different smile types. Positive values indicated that the right side was stronger than the left, and negative values designated the reverse. In this study, no difference was found between genuine and posed smiles; this was unlike the results reported by Ross and Pulusu (2013), where posed expressions overwhelmingly originated on the right side of the face and spontaneous expressions on the left. If the positive–negative (left–right) direction was not considered (i.e., if the absolute value of the difference, Symmetry-d, was used) the results revealed that the genuine and posed smiles did differ, with the posed smiles being more asymmetrical (larger time gaps were observed between the left and right lip corners). Posed smiles were more asymmetrical; this echoed the results of many studies, though most observations were of intensity rather than onset time. Different from Ross and Pulusu's (2013) work that considered different types of asymmetry for different types of smiles and Schmidt et al.'s (2006) study that analyzed assymetrical differences between genuine and posed smiles, we observed another type of assymetry (the significance of Symmetry-d). This considerably complicated the temporal asymmetry.

Irregularity

Only a very few studies have addressed irregularities in facial expressions. In our research, we used two indicators to describe irregularity. However, they did not support one another. Genuine smiles had greater values for Irregularity-a (the rate of peaks) and lesser values for Irregularity-b (the *SD* value for the apex phase). Irregularity-a was similar to the indicator used in Hess and Kleck (1990), and the results were consistent with their findings; posed smiles were more irregular.

Irregularity-b was a pilot indicator for this study. The values for Irregularity-b were larger for genuine smiles, indicating that during the apex phase, more changes were seen in genuine than in posed smiles. Because genuine emotions are not the same every time, genuine smiles tend to vary in strength, duration, and type. Even for smiles attributable to enjoyment, expressions may often differ in various ways. Conversely, in posed smiles, individuals may follow a prototypical pattern that results in the expressions being similar. Thus, the notion of irregularity requires further research.

Some Considerations Regarding Facial Dynamics Analysis with Computer Vision Techniques

Over the past few years, there has been an increased interest in automatic facial behavior analysis. There are many algorithms currently available for analyzing facial movements, and specifically, lip corners. The current study is one of many that applies computer vision techniques to the analysis of nonverbal behavior.

OpenFace provided the 2D and 3D approaches we employed here. Traditionally, facial tracking has primarily been based on 2D methods, and these algorithms have matured. The 3D model in OpenFace is actually based on 2D images and does not actually include depth information from the camera. Instead, OpenFace uses a 3D representation of facial landmarks and projects them onto the image using orthographic camera projection. Therefore, the reliability and validity of the model used here needed to be further verified. The 3D model was able to extract head pose information (translation and orientation), in addition to detecting facial landmarks. This allowed us to accurately estimate the head pose once the landmarks were detected. With these considerations, we were able to employ two approaches that could be compared with one another.

The techniques are improving and as a result, accuracy will be enhanced. However, there are a variety of conflicting ideas regarding how to define features, not only in terms of the boundaries of phases, but also dynamic elements such as duration, speed, symmetry, and irregularity. Different groups have different definitions, which makes for inconsistencies in the literature. Therefore, we should be cautious when comparing results produced by different research groups.

AUTHOR CONTRIBUTIONS

HG proposed the idea and gave suggestions in writing the paper; X-HZ designed the experiment and analyzed the data; JL analyzed the data; W-JY proposed the idea, designed the experiment, analyzed the data and wrote the paper.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (No. 31500875), Zhejiang Provincial Natural Science Foundation of China (No. LQ16C090002), and Wenzhou Social Development Foundation (2017Y525).

REFERENCES

Ambadar, Z., Cohn, J. F., and Reed, L. I. (2009). All smiles are not created equal: morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. *J. Nonverbal Behav.* 33, 17–34. doi: 10.1007/s10919-008-0059-5

Baltrusaitis, T., Robinson, P., and Morency, L. P. (2013). "Constrained local neural fields for robust facial landmark detection in the wild," in *Paper Presented at the IEEE International Conference on Computer Vision Workshops* (Sydney, NSW).

Baltrusaitis, T., Robinson, P., and Morency, L. P. (2016). "OpenFace: an open source facial behavior analysis toolkit," in *Paper Presented at the IEEE Winter Conference on Applications of Computer Vision* (Lake Placid, NY).

Barrett, L. F., Lewis, M., and Haviland-Jones, J. M. (2016). *Handbook of Emotions.* New York, NY: Guilford Publications.

Bolzani-Dinehart, L. H., Messinger, D. S., Acosta, S. I., Cassel, T., Ambadar, Z., and Cohn, J. (2005). Adult perceptions of positive and negative infant emotional expressions. *Infancy* 8, 279–303. doi: 10.1207/s15327078in08035

Borod, J. C., Koff, E., Yecker, S., Santschi, C., and Schmidt, J. M. (1998). Facial asymmetry during emotional expression: gender, valence, and measurement technique. *Neuropsychologia* 36, 1209–1215.

Cohn, J. F., and Sayette, M. A. (2010). Spontaneous facial expressions in a small group can be automatically measured: an initial demonstration. *Behav. Res. Methods* 42, 1079–1086. doi: 10.3758/BRM.42.4.1079

Cohn, J. F., and Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *Int. J. Wavelets Multires. Inform. Process.* 2, 121–132. doi: 10.1142/S021969130400041X

Dibeklioğlu, H., Salah, A. A., and Gevers, T. (2012). *Are You Really Smiling at Me? Spontaneous versus Posed Enjoyment Smiles*. Berlin; Heidelberg: Springer.

Ekman, P. (2006). Darwin, Deception, and Facial Expression. *Ann. N.Y. Acad. Sci.* 1000, 205–221. doi: 10.1196/annals.1280.010

Ekman, P. (2009). "Lie catching and microexpressions," in *The Philosophy of Deception*, ed C. Martin (Oxford: Oxford University Press), 118–133.

Ekman, P., Davidson, R. J., and Friesen, W. V. (1990). The Duchenne smile: emotional expression and brain physiology: II. *J. Pers. Soc. Psychol.* 58:342. doi: 10.1037/0022-3514.58.2.342

Ekman, P., Friesen, W., and Hager, J. (2002). *FACS Investigator's Guide (The Manual on CD Rom)*. Salt Lake: Network Information Research Corporation.

Ekman, P., and Friesen, W. V. (1982). Felt, false, and miserable smiles. *J. Nonverbal Behav.* 6, 238–252. doi: 10.1007/BF00987191

Ekman, P., Friesen, W. V., and O'Sullivan, M. (1988). Smiles when lying. *J. Person. Soc. Psychol.* 54, 414–420. doi: 10.1037/0022-3514.54.3.414

Ekman, P., Hager, J. C., and Friesen, W. V. (1981). The symmetry of emotional and deliberate facial actions. *Psychophysiology* 18, 101–106. doi: 10.1111/j.1469-8986.1981.tb02919.x

Figueiredo Filho, D. B., Paranhos, R., Rocha, E. C. D., Batista, M., da Silva J. A. Jr., Santos, M. L. W. D., et al. (2013). When is statistical significance not significant? *Braz.. Polit. Sci. Rev.* 7, 31–55. doi: 10.1590/S1981-38212013000100002

Frank, M. G., and Ekman, P. (1993). Not all smiles are created equal: the differences between enjoyment and nonenjoyment smiles. *Humor Int. J. Humor Res.* 6, 9–26. doi: 10.1515/humr.1993.6.1.9

Frank, M. G., Ekman, P., and Friesen, W. V. (1993). Behavioral markers and recognizability of the smile of enjoyment. *J. Person. Soc. Psychol.* 64:83. doi: 10.1037/0022-3514.64.1.83

Frank, M. G., Maccario, C. J., and Govindaraju, V. (2009). "Behavior and security," in *Protecting Airline Passengers in the Age of Terrorism,* ed P. Seidenstat (Santa Barbara, CA: Greenwood Pub Group), 86–106.

Gosselin, P., Beaupré, M., and Boissonneault, A. (2002). Perception of genuine and masking smiles in children and adults: sensitivity to traces of anger. *J. Genet. Psychol.* 163, 58–71. doi: 10.1080/00221320209597968

Gunnery, S. D., Hall, J. A., and Ruben, M. A. (2013). The deliberate Duchenne smile: individual differences in expressive control. *J. Nonverbal Behav.* 37, 29–41. doi: 10.1007/s10919-012-0139-4

Gutiérrezgarcía, A., and Calvo, M. G. (2015). Discrimination thresholds for smiles in genuine versus blended facial expressions. *Cogent Psychol.* 2:1064586. doi: 10.1080/23311908.2015.1064586

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., and Tatham, R. L. (1998). *Multivariate Data Analysis, Vol. 5*. Upper Saddle River, NJ: Prentice hall, 207–219.

Hess, U., and Kleck, R. E. (1990). Differentiating emotion elicited and deliberate emotional facial expressions. *Eur. J. Soc. Psychol.* 20, 369–385. doi: 10.1002/ejsp.2420200502

Hoque, M., Morency, L. P., and Picard, R. W. (2011). Are you friendly or just polite?: analysis of smiles in spontaneous face-to-face interactions," in *Paper presented at the Affective Computing and Intelligent Interaction International Conference, ACII 2011* (Memphis, TN).

Krumhuber, E. G., and Manstead, A. S. (2009). Can Duchenne smiles be feigned? New evidence on felt and false smiles. *Emotion* 9, 807–820. doi: 10.1037/a0017844

Krumhuber, E., and Kappas, A. (2005). Moving smiles: the role of dynamic components for the perception of the genuineness of smiles. *J. Nonverbal Behav.* 29, 3–24. doi: 10.1007/s10919-004-0887-x

Krumhuber, E., Manstead, A. S., Cosker, D., Marshall, D., and Rosin, P. L. (2009). Effects of dynamic attributes of smiles in human and synthetic faces: a simulated job interview setting. *J. Nonverbal Behav.* 33, 1–15. doi: 10.1007/s10919-008-0056-8

Lien, J. J.-J., Kanade, T., Cohn, J. F., and Li, C.-C. (2000). Detection, tracking, and classification of action units in facial expression. *Robot. Autonom. Syst.* 31, 131–146. doi: 10.1016/S0921-8890(99)00103-7

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., et al. (2011). The computer expression recognition toolbox (CERT)," in *Paper Presented at the Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference* (Santa Barbara, CA).

Mavadati, S. M., Mahoor, M. H., Bartlett, K., Trinh, P., and Cohn, J. F. (2013). Disfa: a spontaneous facial action intensity database. *IEEE Trans. Affect. Comp.* 4, 151–160. doi: 10.1109/T-AFFC.2013.4

Pantic, M., and Patras, I. (2006). Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Trans. Syst. Man Cybernet. Part B Cybernet.* 36, 433–449. doi: 10.1109/TSMCB.2005.859075

Powell, W. R., and Schirillo, J. A. (2009). Asymmetrical facial expressions in portraits and hemispheric laterality: a literature review. *Laterality* 14, 545–572. doi: 10.1080/13576500802680336

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educ. Res. Rev.* 6, 135–147. doi: 10.1016/j.edurev.2010.12.001

Ross, E. D., and Pulusu, V. K. (2013). Posed versus spontaneous facial expressions are modulated by opposite cerebral hemispheres. *Cortex* 49, 1280–1291. doi: 10.1016/j.cortex.2012.05.002

Scherer, K. R., and Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion* 7:113. doi: 10.1037/1528-3542.7.1.113

Schmidt, K. L., Bhattacharya, S., and Denlinger, R. (2009). Comparison of deliberate and spontaneous facial movement in smiles and eyebrow raises. *J. Nonverbal Behav.* 33, 35–45. doi: 10.1007/s10919-008-0058-6

Schmidt, K. L., Cohn, J. F., and Tian, Y. (2003). Signal characteristics of spontaneous facial expressions: automatic movement in solitary and social smiles. *Biol. Psychol.* 65, 49–66. doi: 10.1016/S0301-0511(03)00098-X

Schmidt, K. L., Ambadar, Z., Cohn, J. F., and Reed, L. I. (2006). Movement differences between deliberate and spontaneous facial expressions: Zygomaticus major action in smiling. *J. Nonverbal Behav.* 30, 37–52. doi: 10.1007/s10919-005-0003-x

Schmidt, K. L., and Cohn, J. F. (2001). Human facial expressions as adaptations: evolutionary questions in facial expression research. *Am. J. Phys. Anthropol.* 116, 3–24. doi: 10.1002/ajpa.20001

Schneider, K., and Josephs, I. (1991). The expressive and communicative functions of preschool children's smiles in an achievement situation. *J. Nonverbal Behav.* 15, 185–198. doi: 10.1007/BF01672220

Sebe, N., Lew, M. S., Sun, Y., Cohen, I., Gevers, T., and Huang, T. S. (2007). Authentic facial expression analysis. *Image Vis. Comput.* 25, 1856–1863. doi: 10.1016/j.imavis.2005.12.021

Shore, D. M., and Heerey, E. A. (2011). The value of genuine and polite smiles. *Emotion* 11, 169–174. doi: 10.1037/a0022601

Skinner, M., and Mullen, B. (1991). Facial asymmetry in emotional expression: a meta-analysis of research. *Br. J. Soc. Psychol.* 30, 113–124. doi: 10.1111/j.2044-8309.1991.tb00929.x

Soussignan, R., and Schaal, B. (1996). Forms and social signal value of smiles associated with pleasant and unpleasant sensory experience. *Ethology* 102, 1020–1041. doi: 10.1111/j.1439-0310.1996.tb01179.x

Tarantili, V. V., Halazonetis, D. J., and Spyropoulos, M. N. (2005). The spontaneous smile in dynamic motion. *Am. J. Orthod. Dentofacial Orthop.* 128, 8–15. doi: 10.1016/j.ajodo.2004.03.042

Tian, Y., Kanade, T., and Cohn, J. F. (2011). "Facial expression recognition," in *Handbook of Face Recognition,* eds S. Z. Li and A. K. Jain (London: Springer), 487–519.

Wood, E., Baltruaitis, T., Zhang, X., Sugano, Y., Robinson, P., and Bulling, A. (2015). Rendering of eyes for eye-shape registration and gaze estimation," in *Paper presented at the 2015 IEEE International Conference on Computer Vision (ICCV)* (Santiago).

Wu, T., Butko, N. J., Ruvolo, P., Whitehill, J., Bartlett, M. S., and Movellan, J. R. (2012). Multilayer architectures for facial action unit recognition. *IEEE Trans. Syst. Man Cybernet. Part B Cybernet.* 42, 1027–1038. doi: 10.1109/TSMCB.2012.2195170

Yan, W. J., and Chen, Y. H. (in press). Measuring dynamic micro-expressions via feature extraction methods. *J. Comput. Sci.* doi: 10.1016/j.jocs.2017.02.012

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

GENERAL COMMENTARY
published: 25 September 2018
doi: 10.3389/fpsyg.2018.01610

Commentary: The Dynamic Features of Lip Corners in Genuine and Posed Smiles

Yingqi Li[1], *Zhongyong Shi*[2,3], *Honglei Zhang*[4,5], *Lishu Luo*[4,5] and *Guoxin Fan*[5,6]*

[1] School of Humanity, Tongji University, Shanghai, China, [2] Psychiatry Department, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai, China, [3] Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States, [4] School of Management and Economics, Tianjin University, Tianjin, China, [5] Surgical Planing Lab, Radiology Department, Brigham and Women's Hospital, Boston, MA, United States, [6] School of Medicine, Tongji University, Shanghai, China

Keywords: facial recognition, deep learning, dynamic features, smiles, lip corners

A Commentary on

The Dynamic Features of Lip Corners in Genuine and Posed Smiles
by Guo, H., Zhang, X.-H., Liang, J., and Yan, W.-J. Front. Psychol. 9:202. doi: 10.3389/fpsyg.2018.00202

OPEN ACCESS

Edited by:
Xunbing Shen,
Jiangxi University of Traditional Chinese Medicine, China

Reviewed by:
Lynden K. Miles,
University of Aberdeen, United Kingdom

****Correspondence:***
Guoxin Fan
gfan@tongji.edu.cn;
gfan1@bwh.harvrad.edu

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *06 May 2018*
Accepted: *13 August 2018*
Published: *25 September 2018*

Citation:
Li Y, Shi Z, Zhang H, Luo L and Fan G (2018) Commentary: The Dynamic Features of Lip Corners in Genuine and Posed Smiles. Front. Psychol. 9:1610. doi: 10.3389/fpsyg.2018.01610

For thousands of years of human history, we have learned how to fake or hide our genuine feelings and emotions to people around us intentionally or unconsciously. It is, indeed, an irony that this is what we view as emotional intelligence, and which we practice to win people over, display our politeness, tackle dilemmas, and deal with other complicated situations. Posed smiles are one of the most common faked expressions in our daily life. Indeed, it is a challenge for the computer vision system to recognize the genuine smile apart from posed smiles of an individual, and this may be difficult to interpret by humans too sometimes. Recently, an interesting work by Guo et al. (Guo et al., 2018) employed computer vision techniques to investigate the potential differences in the duration, intensity, speed, symmetry of the lip corners, and certain irregularities between genuine and posed smiles based on the UvA-NEMO Smile Database. The results are quite rewarding since they found that genuine smiles were correlated with higher onset, offset, apex, and total duration, as well as offset displacement and irregularity-b, compared with posed smiles. In addition, posed smiles were correlated with higher onset and offset speeds, irregularity-a, symmetry-a, and symmetry-d.

We cannot agree with the saying that only a handful of studies on the dynamic features of facial expressions have been conducted due to the lack of user-friendly analytic tools. On the contrary, in the past decades, hundreds of studies have focused on the dynamic features of facial expressions (Sandbach et al., 2012; Ko, 2018). Valstar et al. (2006) differentiated spontaneous brow actions from posed ones focusing on velocity, duration, and order of occurrence. Littlewort et al. (2009) distinguished fake pain from real pain by analyzing facial actions based on Gabor features. Dibeklioglu et al. (2012) analyzed the dynamics of eyelids, cheeks, and lip corners to tell genuine smiles from posed ones, and extracted 25 features, which were also cited by the author. Guo et al. (2018) said that not all these 25 features could be explained from a psychological perspective; hence, they extracted the duration, speed, intensity, symmetry, and irregularity aspects in their study. The question is why do all of the potential features need to be explained by psychological theory. It is possible that in this manner we may lose a lot of useful information to help distinguish genuine smiles from posed ones. Obviously, we still have great limited knowledge in psychology itself.

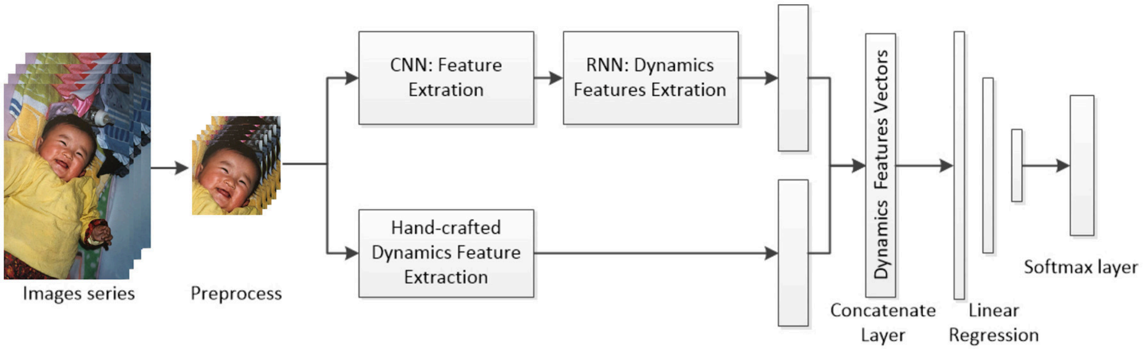
Images series
Preprocess
CNN: Feature Extration
RNN: Dynamics Features Extration
Hand-crafted Dynamics Feature Extraction
Dynamics Features Vectors
Concatenate Layer
Linear Regression
Softmax layer

FIGURE 1 | Hybrid model combining CNN and hand-crafted dynamic features (the smile picture belongs to the first author, and informed consent was obtained from the first author).

Indeed, the value of all the above-mentioned pioneering works should be appreciated, as they helped improve the recognition of posed smiles from spontaneous expressions over time. However, the hand-crafted features built by rules may lead to inadequate abstraction and representations. We are wondering whether the 25 features encompass the whole story to tell genuine smiles from posed ones, and how many of these extracted features would help the computer vision system to recognize posed smiles from genuine facial expressions. Obviously, there is still a lot of work left for us to consider and all of the features identified by different studies and extracted from different datasets need to be analyzed to help conduct the recognition performance. We cannot tell how much the dynamic features of the lip corners would help to differentiate genuine smiles from posed smiles from the diagnostic data presented in the current study.

Recently, deep learning has led to overwhelming performances in image or video processing over conventional methods such as facial recognition and classification (Peng et al., 2017; Rodriguez et al., 2017; Majumder et al., 2018; Yu et al., 2018). Many start-up companies have already built their businesses displaying outstanding performance in the field of facial recognition in security. It is not surprising that researchers have already adopted convolutional neural networking (CNN) to differentiate genuine smiles from posed ones, and the recognition performances have been promising (Kumar et al., 2017; Mandal et al., 2017). However, another question arises as to whether deep learning will take over this area and wipe out the necessity of studying hand-crafted features.

In reality the recognition performances, as consequences of deep learning in classifying the genuine smiles and posed smiles may rely heavily on the size of the training data. Unfortunately, datasets containing labeled genuine smiles and posed smiles are limited (Xu et al., 2017). However, the good news is that hand-crafted features combined with deep learning may have the potential to improve the recognition performances compared with deep learning alone supported by limited data (Pesteie et al., 2018). It is possible to build a hybrid model by inputting features from deep learning along with well-known features obtained from conventional methods into a classifier (**Figure 1**). We admit that deep learning has also been criticized for its level of interpretability, known as the black box. However, many researchers have realized the importance of solving the problem of the black box associated with deep learning, and solutions have been proposed to tackle the same (Gunning, 2017; Samek et al., 2017; Shwartz-Ziv and Tishby, 2017).

Considering outstanding recognition performance, we do believe that deep learning will dominate the area of image recognition and classification, including discriminating genuine smiles from posed ones. As for the black box, we should regard it as an accompanying aspect of deep learning, instead of being a mere limitation. It would be better if we can solve the problem of the black box similar to how Newton figured out why apples always fell to the ground. When that day comes, deep learning will have a greater impact than it has today, though we admit that more efforts are needed to solve the problem of the black box associated with deep learning.

AUTHOR CONTRIBUTIONS

YL designed and wrote the manuscript. ZS revised the manuscript. HZ and LL gave critical comments. GF reviewed and approved the manuscript.

ACKNOWLEDGMENTS

This work is funded by the China Scholarship Council (201706260169). I would like to thank Jolin Tsai, whose songs have inspired me throughout my whole life. The study is also funded by another two grants: Key Project of Philosophy and Social Science Research of the Ministry of Education (15JZD026); National Natural Science Foundation (71673200).

REFERENCES

Dibeklioglu, H., Salah, A. A., and Gevers, T. (2012). "Are you really smiling at me? Spontaneous versus posed enjoyment smiles," in *Computer Vision – ECCV 2012*, eds A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid (Berlin; Heidelberg: Springer Berlin Heidelberg), 525–538.

Gunning, D. (2017). *Explainable Artificial Intelligence (xai)*, Arlington, VA: Defense Advanced Research Projects Agency (DARPA), nd Web.

Guo, H., Zhang, X. H., Liang, J., and Yan, W. J. (2018). The dynamic features of lip corners in genuine and posed smiles, *Front. Psychol.* 9:202. doi: 10.3389/fpsyg.2018.00202

Ko, B. C. (2018). A brief review of facial emotion recognition based on visual information. *Sensors (Basel)* 18:E401. doi: 10.3390/s18020401

Kumar, G. A. R., Kumar, R. K., and Sanyal, G. (2017). "Discriminating real from fake smile using convolution neural network," in *2017 International Conference on Computational Intelligence in Data Science (ICCIDS)* (Chennai: IEEE), 1–6.

Littlewort, G. C., Bartlett, M. S., and Lee, K. (2009). Automatic coding of facial expressions displayed during posed and genuine pain. *Image Vis. Comput.* 27, 1797–1803. doi: 10.1016/j.imavis.2008.12.010

Majumder, A., Behera, L., and Subramanian, V. K. (2018). Automatic facial expression recognition system using deep network-based data fusion. *IEEE Trans. Cybern.* 48, 103–114. doi: 10.1109/TCYB.2016.2625419

Mandal, B., Lee, D., and Ouarti, N. (2017). "Distinguishing posed and spontaneous smiles by facial dynamics," in *Computer Vision – ACCV 2016 Workshops*, eds C.-S. Chen, J. Lu, and K.-K. Ma (Cham: Springer International Publishing), 552–566.

Peng, M., Wang, C., Chen, T., Liu, G., and Fu, X. (2017). Dual temporal scale convolutional neural network for micro-expression recognition. *Front. Psychol.* 8:1745. doi: 10.3389/fpsyg.2017.01745

Pesteie, M., Lessoway, V., Abolmaesumi, P., and Rohling, R. N. (2018). Automatic localization of the needle target for ultrasound-guided epidural injections. *IEEE Trans. Med. Imaging* 37, 81–92. doi: 10.1109/TMI.2017.2739110

Rodriguez, P., Cucurull, G., Gonalez, J., Gonfaus, J. M., Nasrollahi, K., Moeslund, T. B., et al. (2017). Deep pain: exploiting long short-term memory networks for facial expression classification. *IEEE Trans Cybern.* 1–11. doi: 10.1109/TCYB.2017.2662199

Samek, W., Wiegand, T., and Müller, K.-R. (2017). *Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models*. arXiv:1708.08296 [Preprint].

Sandbach, G., Zafeiriou, S., Pantic, M., and Yin, L. (2012). Static and dynamic 3D facial expression recognition: a comprehensive survey. *Image Vis. Comput.* 30, 683–697. doi: 10.1016/j.imavis.2012.06.005

Shwartz-Ziv, R., and Tishby, N. (2017). *Opening the Black Box of Deep Neural Networks via Information*. arXiv:1703.00810 [Preprint].

Valstar, M. F., Pantic, M., Ambadar, Z., and Cohn, J. F. (2006). "Spontaneous vs. posed facial behavior: automatic analysis of brow actions," in *Proceedings of the 8th international conference on Multimodal Interfaces*, ACM, Banff, Alberta, Canada, 162–170.

Xu, C., Qin, T., Bar, Y., Wang, G., and Liu, T. Y. (2017). "Convolutional neural networks for posed and spontaneous expression recognition," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, 769–774.

Yu, Z., Tan, E. L., Ni, D., Qin, J., Chen, S., Li, S., et al. (2018). A deep convolutional neural network-based framework for automatic fetal facial standard plane recognition. *IEEE J. Biomed. Health Inform.* 22, 874–885. doi: 10.1109/JBHI.2017.2705031

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 31 August 2016
doi: 10.3389/fpsyg.2016.01346

Electrophysiological Evidence Reveals Differences between the Recognition of Microexpressions and Macroexpressions

*Xunbing Shen[1,2], Qi Wu[2,3], Ke Zhao[2] and Xiaolan Fu[2]**

[1] Department of Psychology, Jiangxi University of Traditional Chinese Medicine, Nanchang, China, [2] State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China, [3] Department of Psychology, Hunan Normal University, Changsha, China

Microexpressions are fleeting facial expressions that are important for judging people's true emotions. Little is known about the neural mechanisms underlying the recognition of microexpressions (with duration of less than 200 ms) and macroexpressions (with duration of greater than 200 ms). We used an affective priming paradigm in which a picture of a facial expression is the prime and an emotional word is the target, and electroencephalogram (EEG) and event-related potentials (ERPs) to examine neural activities associated with recognizing microexpressions and macroexpressions. The results showed that there were significant main effects of duration and valence for N170/vertex positive potential. The main effect of congruence for N400 is also significant. Further, sLORETA showed that the brain regions responsible for these significant differences included the inferior temporal gyrus and widespread regions of the frontal lobe. Furthermore, the results suggested that the left hemisphere was more involved than the right hemisphere in processing a microexpression. The main effect of duration for the event-related spectral perturbation (ERSP) was significant, and the theta oscillations (4 to 8 Hz) increased in recognizing expressions with a duration of 40 ms compared with 300 ms. Thus, there are different EEG/ERPs neural mechanisms for recognizing microexpressions compared to recognizing macroexpressions.

Keywords: microexpression, macroexpression, recognition, EEG/ERPs, ERSP, sLORETA

OPEN ACCESS

Edited by:
Marco Tamietto,
Tilburg University, Netherlands

Reviewed by:
Jan Van Den Stock,
Katholieke Universiteit Leuven,
Belgium
Frank A. Russo,
Ryerson University, Canada

****Correspondence:***
Xiaolan Fu
fuxl@psych.ac.cn

Specialty section:
This article was submitted to
Emotion Science,
a section of the journal
Frontiers in Psychology

Received: *30 April 2016*
Accepted: *23 August 2016*
Published: *31 August 2016*

Citation:
Shen X, Wu Q, Zhao K and Fu X
(2016) Electrophysiological Evidence
Reveals Differences between
the Recognition of Microexpressions
and Macroexpressions.
Front. Psychol. 7:1346.
doi: 10.3389/fpsyg.2016.01346

INTRODUCTION

Facial expressions serve important social functions, and the recognition of emotional facial expressions is vital for everyday life (Niedenthal and Brauer, 2012). However, emotion is not necessarily displayed on the face at all times. In a number of interpersonal situations, people hide, disguise, or inhibit their true feelings (Ekman, 1971), leading to partial or very rapid production of expressions of emotion, which are called microexpressions (Ekman and Friesen, 1969; Bhushan, 2015).

A microexpression is a facial expression that lasts between 1/25 and 1/5 of a second, revealing an emotion that a person is trying to conceal (Ekman and Friesen, 1969; Ekman, 1992, 2003; Porter and ten Brinke, 2008). A microexpression resembles one of the universal emotions: disgust, anger, fear, sadness, happiness, or surprise. Microexpressions usually occur in high-stakes situations in which people have something valuable to gain or lose (Ekman et al., 1992). According to Ekman (2009), microexpressions are believed to reflect a person's true intent, especially one of a hostile nature. Therefore, microexpressions can provide an essential behavioral clue for lie detection and can be employed to detect a dangerous demeanor (Metzinger, 2006; Schubert, 2006; Weinberger, 2010).

Little is known regarding the characteristics that differentiate microexpressions and macroexpressions. The most important difference between microexpressions and macroexpressions is their duration (Svetieva, 2014). However, there are different estimates of the duration of microexpressions (Shen et al., 2012). According to Shen et al. (2012), there are at least six estimates of the duration of microexpressions, and 200 ms duration can be used as a boundary for differentiating microexpressions and macroexpressions. However, it is unclear whether there are different neural indicators for recognizing expressions with durations of less than 200 ms and those with durations of longer than 200 ms. If microexpressions and macroexpressions are qualitatively different (from the viewpoint of the perceiver, they should be recognized as different objects), then it can be expected that there are different brain mechanisms for processing facial expressions with a duration shorter than or longer than the duration boundary (200 ms).

Thus, we aimed to investigate the neural mechanisms for recognizing expressions with different durations, which can aid in the evidence-based separation of microexpressions from macroexpressions (i.e., to determine the boundary between microexpressions and macroexpressions). In other words, if differences in the neural characteristics that recognize one group of expressions with one kinds of duration and another group of expressions with other kinds of duration exist, we can say that the two groups of expressions are different. If we can find the discrepancy in the electroencephalogram (EEG)/event-related potentials (ERPs) between expressions with different durations, we can divide expressions with different EEG/ERPs characteristics into two groups. One group can be called microexpressions (with a short duration), and the other can be called macroexpressions (with a longer duration). Given the behavioral difference in recognition of microexpressions and macroexpressions and the disagreement regarding the conceptual definition for the duration of a microexpression, we seek to find electrophysiological evidence of the boundary (200 ms) that separates microexpressions from macroexpressions.

The EEG can indicate the characteristic temporal, spatial, and spectral signatures of specific cognitive processes. We explored the EEG activities during recognizing expressions with different duration (40, 120, 200, and 300 ms) to examine whether there is a turning point near 200 ms as indicated by the EEG measurements. Two different objects or ideas should only be thought of as separate entities when they have a number of differing characteristics. If neural differences are present before and after the turning point (e.g., 200 ms), then we can safely say that the duration of the conceptual definition of a microexpression is less than the turning point (the upper limit of microexpression duration). As there are different behavioral characteristics in the recognition of expressions with a duration of less than 200 ms and expressions with a duration longer than 200 ms (Shen et al., 2012), we hypothesized that recognizing expressions with a duration of less than 200 ms and expressions with a duration of longer than 200 ms will show different EEG characteristics (i.e., amplitude, oscillatory dynamics, and source location). Consequently, there should be different brain mechanisms for recognizing microexpressions and macroexpressions. Hence, the present study aimed to provide evidence for separating microexpressions and macroexpressions by investigating EEG/ERPs and synchronized oscillatory activity.

We used an affective priming paradigm, in which a picture of a facial expression is the prime and an emotional word is the target. Meanwhile, we mainly focused on the ERPs components of N400 and N170. The N400 can be produced not only in instances of semantic mismatch but also in other incongruous meaningful stimuli, such as words and faces. The effects of the N400 can also be observed in response to line drawings, pictures, and faces when primed by single items or sentence contexts, but not in the absence of priming (Kutas and Federmeier, 2011). In a pilot experiment, we found that the N400 could be elicited in the expression – emotional word priming paradigm. This N400 amplitude is more negative for incongruent than for congruent emotional content of face-word pairs. To produce a greater N400 effect, an incongruent condition in the experiment that elicits a greater negative-going wave than does the congruent condition should also be present. Therefore, we employed pictures of facial expressions (happy, fearful, and neutral) as priming stimuli and emotional words (positive and negative) as targets. Consequently, there were three conditions (congruent, incongruent, and control) with respect to the congruence of emotional valence.

Expressions can have different durations. There are expressions with short duration (e.g., less than 200 ms). If we recognize them as the same because of the limited time to process them, then during conditions of short duration for recognizing expressions, there is no congruence or incongruence due to expressions with different short durations being observed as the same by the participants. On the contrary, expressions with different long durations (e.g., greater than 200 ms) will appear to be different to the participants due to the extensive time for processing, which can result in congruence and incongruence. Thus, there will be no effect of congruence at the short duration; however, the effect of congruence at the long duration will be significant. To put it another way, when the presentation of an expression is transient, there is no top-down influences on the recognition of the expression. Consequently, there should be no difference between the congruent and incongruent conditions. Only when the duration of the expression is sufficiently long do the participants engage in top-down processing and recognize the expressions differently, which results in the congruent and incongruent conditions. Therefore, we could expect that there will be a significant relationship between duration and congruence while measuring N400 effects which reflect the top-down influences (Newman and Connolly, 2009). If we could extensively process expressions with short duration (less than 200 ms), meaning there was no difference between expressions with short and long duration, then there would be no relationship between duration and congruence.

For facial processing, one of the most prominent components in the ERPs is the N170 (Rossion and Jacques, 2011), and the face-sensitive N170 is modified by facial expressions of emotion (Batty and Taylor, 2003; Righart and De Gelder, 2008). As noted by Joyce and Rossion (2005), Eimer (2011), the N170 may be a vertex positive potential (VPP), resulting from changing the reference electrodes from the mastoid to the common average reference. The N170/VPP may be a valuable tool for studying the cognitive and neurobiological mechanisms underlying expression recognition. If there is a turning point in accuracy for recognizing expressions with different durations (i.e., there is a duration boundary for microexpressions and macroexpressions; see Shen et al., 2012), then we can expect that the main effect of duration will be significant. Specifically, there should be a significant difference between an expression with duration of less than 200 ms (microexpression) and an expression with duration of longer than 200 ms (macroexpression) while measuring N170/VPP. That is, there should be two groups of N170/VPP, one for microexpressions and one for macroexpressions (which can lead to a conclusion that expressions with short and long durations fall into two different categories).

It is worth noting that the priming paradigm provides an avenue for studying expression perception and recognition, which is appropriate for our aims. First, we wanted to investigate the effect of duration on the ERPs of expression recognition; when the duration of the expression is longer (macroexpression), the valence of the expression will be processed and the later processing of the emotional word will be facilitated or inhibited. Consequently, the N400 will reflect the facilitated or inhibited effect, i.e., there should be a smaller N400 when the valence of the expression and the emotional word are congruent. However, when the duration of expression is shorter (microexpression) the valence of the expression may not be fully processed, and there may be no facilitated or inhibited effect. Therefore, the N400 for the processing of microexpressions should not be affected, regardless of the congruence of the emotional valence. Second, this paradigm offers insights into the time course of the perception and the recognition of microexpressions and macroexpressions while measuring N170/ VPP.

The information regarding oscillatory dynamics from the EEG signal is largely lost by the time-locked averaging of single trials in the traditional ERPs approach. Researching functional correlates of brain oscillations is an important current trend in neuroscience. The traditional spectral analysis cannot fully address the issue of rapidly changing neural oscillations. Time–frequency analysis of an EEG allows researchers to study the changes of the signal spectrum over time, taking into account the power (or amplitude) of the EEG signal at a given frequency as well as changes in the phase or latency (Buzsáki, 2006; Roach and Mathalon, 2008; Güntekin and Başar, 2014). Some recent studies investigated the mechanisms of perception and categorization of emotional stimuli through brain oscillatory activity extracted from EEG signals (Keil, 2013). Oscillatory dynamics of theta, alpha, beta, and gamma bands, and the interplay of these frequencies, relates to the processing of emotional stimuli (Güntekin and Başar, 2014). Furthermore, some EEG studies show that the theta band activities, which are associated with subcortical brain regions and are considered to be the fingerprint of all limbic structures, are involved in affective processes (Knyazev and Slobodskaya, 2003). Meanwhile, theta band activity was observed during emotional stimulus presentation and it was associated with emotion comprehension (Balconi and Pozzoli, 2007). Therefore, this study mainly explores the dynamic oscillatory patterns of theta bands activities in the EEG signal while recognizing microexpressions and macroexpressions.

Previous studies (Esslen et al., 2004; Costa et al., 2014) had found that different emotional conditions had different activation patterns in different brain regions by using the low resolution brain electromagnetic tomography. In the current study, we also employed Standardized Low Resolution Brain Electromagnetic Tomography (sLORETA; Fuchs et al., 2002; Pascual-Marqui, 2002; Jurcak et al., 2007) to identify brain regions involved in recognizing expressions with long and short durations.

MATERIALS AND METHODS

All experimental protocols were approved by the Institutional Review Board of the Institute of Psychology, Chinese Academy of Sciences. The methods were carried out in accordance with the approved guidelines.

Participants

Sixteen paid volunteers (8 female, ages 20 to 25 years, mean age = 22.3; 8 male, ages 22 to 24, mean age = 22.6) with no history of neurological injury or disorder were recruited from local college campuses. They gave written informed consent before participating. All participants had normal or corrected-to-normal vision and were predominantly right-handed (self-reported). Data from four participants containing too many artifacts were excluded from the analysis (including one participant with higher score of SDS, see the Results section), and the final analyses were conducted on twelve participants (7 female, mean age ± SD: 22.4 ± 1.4 years).

Stimuli and Experimental Design

The pictures of faces consisted of 10 different individuals displaying fear (negative), happiness (positive) or a neutral expression; a total of 30 pictures of facial expressions were selected from 10 models taken from the Pictures of Facial Affect (POFA[1]). The emotional words consisted of 50 positive and 50 negative Chinese words selected from Wang and Fu (2011). The picture stimuli were 200 pixels × 300 pixels, and the word stimuli were 100 pixels × 150 pixels.

The stimuli were presented at a viewing distance of approximately 80 cm and displayed at a moderate contrast (black letters on a silver-gray background) in the center of a 17-inch computer screen with a refresh rate of 60 Hz. The experimental design was as follows: 4 durations (40, 120, 200, and 300 ms) × 3 congruencies (Congruence, Incongruence, and Control).

[1] http://www.paulekman.com

Procedure

The participants were seated in a comfortable armchair in a dimly lit, sound damped booth. Emotional faces and words were presented using a priming paradigm. Subjects were asked to remember all of the content displayed on the screen to focus their attention on the task and to ensure the depth of processing of the words and pictures. No other tasks were imposed on the subjects during the ERPs recordings to avoid confounding the EEG for emotion processing with electrophysiological activity associated with motion for response selection and response execution. The experiment was divided into four blocks according to duration, with each lasting approximately 15 min. At the end of each block, the participants were given a test of recognition. After each block, the subjects were allowed to rest for 2 min. After the EEG recordings, each subject was asked to rate their mood using the Chinese version of the Zung Self Rating Anxiety and Depression Scales (SAS), SDS, selected from Wang et al. (1993).

Stimuli appeared one at a time in trials consisting of pictures of faces and emotional words. Four blocks were divided by the duration of exposure to the pictures of faces, which were 40, 120, 200, and 300 ms. Each trial consisted of a succession of stimuli: a fixation (with a duation randomly selected from 300 to 500 ms), a facial expression picture expressing one of the three emotions (with duration of 40, 120, 200, or 300 ms), a blank screen (the range in duration from 100 to 400 ms), one of the positive or negative emotional words (with duration of 1000 ms), and an interval (the range in duration from 1200 to 1500 ms). There were 300 trials per block. The order of presentation of the four blocks was randomized between subjects. The trial order within each block was randomized. At the end of each block, there was a recognition task (the participants had to judge whether some items including pictures and words were presented before), and the accuracy was measured to monitor the degree of cooperation of the participants. A break of approximately 2 min controlled by the participants separated each successive block.

Electrophysiological Recording and Analyses

EEG/ERPs Acquisition and Analyses

Data were acquired from a 32-channel NuAmps Quickcap, 40-channel NuAmps DC amplifier and Scan 4.5 Acquisition Software (Compumedics Neuroscan, Inc., Charlotte, NC, USA). The EEG data were recorded from 30 scalp sites (Fp1, Fp2, F7, F8, F3, F4, FT7, FT8, T3, T4, FC3, FC4, C3, C4, CP3, CP4, TP7, TP8, T5, T6, P3, P4, O1, O2, Fz, FCz, Cz, CPz, Pz, and Oz). The NuAmps (Model 7181) amplifier had a fixed range of ±130 mV sampled with a 22-bit A/D converter, where the least significant bit was 0.063 μV. The impedance of the recording electrodes was monitored for each subject prior to data collection, and the threshold was always kept below 5 KΩ. The amplifier was set at a gain of 19, with a sampling rate of 1000 Hz and with a signal band limited to 70 Hz. In addition, no notch filter was applied. The electro-oculograms (EOG) were measured to exclude them from the EEG recordings. Vertical EOG (VEOG) was recorded by electrodes 2 cm above and below the left eye and in line with the pupil. The horizontal EOG (HEOG) was recorded by electrodes placed 2 cm from the outer canthi of both eyes. The ground electrode was positioned 10 mm anterior to Fz. The right mastoid electrode (M2) was used as the reference for all recordings and all data were offline re-referenced to a common average reference.

The EEG was later reconstructed into discrete, single-trial epochs. For analyzing the N170/VPP of facial expressions with different durations, an EEG epoch length of 400 ms was used, with a 100 ms pre-stimulus baseline and a 300 ms period, following the onset of the emotional faces. EEG epochs that exceeded ±100 μV were excluded, all trials were visually scanned for further artifacts generated by non-cerebral sources, and corrections were made for eye blinks. Participants had no fewer than 90 accepted epochs in any condition. The accepted epochs were recomputed to the average reference offline and were baseline corrected. The ERPs were averaged separately for each experimental condition. For the averaged N170/VPP wave, a mean amplitude measure within a 140–200 ms time window from onset of the facial stimuli of each participant was provided. The mean amplitude of the N170/VPP then was analyzed by a repeated-measures analyses of variance (ANOVA), in which the factors Valence (positive, negative, and neutral) × Duration (40, 120, 200, and 300 ms) to the mean amplitude were compared.

Facial stimuli under the incongruent condition elicited greater centroparietal ERPs negativity than those under the congruent condition. We termed this negative-going waveform as N400. For this ERPs wave, the epoch length of 1000 ms was used, with a 200 ms pre-stimulus baseline and an 800 ms period, following the onset of the emotional words. A mean amplitude measure within a 350–500 ms time window from the emotional word stimulus onset was provided. The time window of the N400 was selected by visually inspecting, and it more closely resembled a conventional time window of N400. An ANOVA was performed on the N400 mean amplitude.

The N400 was typically maximal over the centro-parietal electrode sites. Therefore, electrodes Cz and CPz were selected for further N400 statistical analysis (one-way analysis of variance, ANOVA), which was carried out on the mean N400 amplitude measurements at the midline central (Cz) and parietal (CPz) electrode locations separately, in which the factors Condition (congruent, incongruent, and control) × Duration (40, 120, 200, and 300 ms) were compared to the N400 mean amplitude. A Greenhouse–Geisser correction to *p*-values was used when appropriate to decrease the risk of falsely significant results.

EEG Time–Frequency Analysis

Time–frequency analysis can be used to reveal event-related oscillations properties, which cannot be depicted by ERPs (Roach and Mathalon, 2008). Time–frequency analysis can represent the energy content of the EEG signal time-locked to an event in the joint time–frequency domain, in which a complex number is estimated for each time point in the time-domain signal, yielding both time and frequency domain information. According to the time–frequency decomposition, the Event-Related Spectral Perturbation [ERSP, the mean change in spectral power (in dB) compared to baseline] analysis was performed (see Makeig et al., 2004; Roach and Mathalon, 2008), particularly the ERSP of theta band activities were analyzed based on the analysis in the introduction. The eeglab 13 (Delorme and Makeig, 2004) was employed for the time–frequency analysis.

Source-Localization Analysis

To compare cortical source differences between EEG activities of expressions with a long duration (>200 ms, macroexpressions) and expressions with a short duration (<200 ms, microexpressions), the standardized low resolution brain electromagnetic tomography (sLORETA) software (publicly available free academic software[2]) was used to estimate the underlying source activity by an equivalent distributed linear inverse solution (Pascual-Marqui et al., 1994, 1999, 2002). sLORETA is an improvement over the previously developed tomography LORETA (Pascual-Marqui et al., 1994). LORETA solves the "inverse problem" by finding the smoothest of all solutions with no *a priori* assumptions about the number, location, or orientation of the generators. It is important to emphasize that sLORETA has no localization bias even in the presence of measurement and biological noise (Pascual-Marqui et al., 2002).

In the current implementation of sLORETA, computations were performed in a realistic head model (Fuchs et al., 2002) using the MNI152 template (Mazziotta et al., 2001), with the three-dimensional solution space restricted to cortical gray matter as determined by the probabilistic Talairach atlas (Lancaster et al., 2000). The standard electrode positions on the MNI152 scalp were taken from Jurcak et al. (2007) and Oostenveld and Praamstra (2001). The intracerebral volume was partitioned in 6239 voxels at a 5 mm spatial resolution. To find the underlying neural generator activity that was most likely responsible for the differences in the recorded scalp potentials, sLORETA calculated the current density (A/m^2) at each voxel allocated by a dipolar source.

To find the brain regions that are most likely involved in processing expressions with different durations, we calculated difference waves by subtracting the N170/VPP for 300 ms trials from the N170/VPP for 40 ms trials during a time window of 140–200 ms. Similarly, we calculated difference waves by subtracting the N400 of incongruent trials from the N400 of congruent trials during a time window of 350–500 ms.

RESULTS

In the survey of the Chinese version of the Zung Self Rating Anxiety and Depression Scales [SAS, SDS, cf., Lui et al. (2009), all scores of our participants were below the critical value of 50 for SAS (mean score = 35.9, *SD* = 5.7), and the scores of all but one participant (who scored 58 and was excluded from further analysis) were under the critical value of 53 for the SDS (mean score = 41.5, *SD* = 7.4). The results of the SAS and SDS clearly demonstrated the participants' normal mood state. All the participants reached accuracy of greater than 80% during all the recognition tasks.

N170/VPP

ERPs Data

The face-sensitive potential of VPP was maximal at the central electrode sites. Therefore, electrodes Cz and CPz were selected for statistical analysis. A 2 Channel (Cz and CPz) × 4 durations (40, 120, 200, and 300 ms) × 3 valence (happiness, fear, and neutral) repeated measures analysis of variance (ANOVA) was conducted. The main effect of Channel is significant, $F(1,11) = 5.567$, $p = 0.038$, $\eta_p^2 = 0.336$; the main effects of duration and valence are both significant, $F(3,33) = 4.176$, $p = 0.037$, $\eta_p^2 = 0.275$; $F(2,22) = 10.412$, $p = 0.001$, $\eta_p^2 = 0.486$. In order to better evaluate the effect of duration and valence on the N170/VPP effect, another ANOVA was conducted for Cz and CPz electrodes separately.

For electrode Cz, there was a main effect of duration, $F(3,33) = 5.027$, $p = 0.006$, $\eta_p^2 = 0.314$. There was also a main effect of valence, $F(2,22) = 10.824$, $p = 0.001$, $\eta_p^2 = 0.496$, and a significant interaction was present, $F(6,66) = 2.766$, $p = 0.018$, $\eta_p^2 = 0.201$. Follow-up t-tests indicated that there is no difference between the N170/VPP mean amplitudes of happiness and fear at duration of 40 and 300 ms [$t(11) = -0.166$; $p < 0.871$; $t(11) = -1.006$; $p < 0.336$]. However, the N170/VPP mean amplitudes of happiness was bigger than that of fear at duration of 120 and 200 ms [$t(11) = -3.612$; $p = 0.004$; $t(11) = -4.127$; $p = 0.002$]. Planned comparisons of durations showed that the N170/VPP amplitude was larger for 40 ms than for 200 ms ($p = 0.022$, see **Figure 1A**). Pairwise comparisons of valence revealed that the N170/VPP amplitude was larger for fearful than for happy faces ($p = 0.004$). There was no difference between other pairings. For the electrode CPz, there was a main effect of duration, $F(3,33) = 2.965$, $p = 0.046$, $\eta_p^2 = 0.212$. There was also a main effect of valence, $F(2,22) = 6.628$, $p = 0.006$, $\eta_p^2 = 0.376$; however, there were no significant interactions, $F(6,66) = 1.002$, $p = 0.432$, $\eta_p^2 = 0.083$. **Figure 1** illustrates the grand average waveforms of N170/VPP at the electrodes Cz and CPz (Panel A). The scalp potential 3D maps of mean amplitude at 140–200 ms for the four corresponding levels of duration are depicted in Panel B).

ERSP Data

As shown in Panel C of **Figure 1**, the results of the ERSP showed that the mean post-stimulus spectral power for fleeting facial expressions with durations of 40 and 120 ms were similar (see the solid red box), and facial expressions with durations of 200 and 300 ms had a similar ERSP pattern (see the dashed purple box).

As shown in **Figure 1C**, the amplitude of theta response (4 to 8 Hz, as traditionally employed based on Berger's studies; see Buzsáki, 2006) was higher for expressions with short duration (<200 ms) than for expressions with longer duration (>200 ms). Therefore, data of theta band activities from 100 to 260 ms of CPz were exported for performing a one-way ANOVA with repeated measures. The results showed that there was a main effect of duration, $F(3,33) = 3.238$, $p = 0.035$, $\eta_p^2 = 0.227$. A *post hoc* pairwise comparison of the theta response of expressions with four levels of duration showed that theta band activity of recognition for expressions with a duration of 40 ms was

[2] http://www.uzh.ch/keyinst/loreta.htm

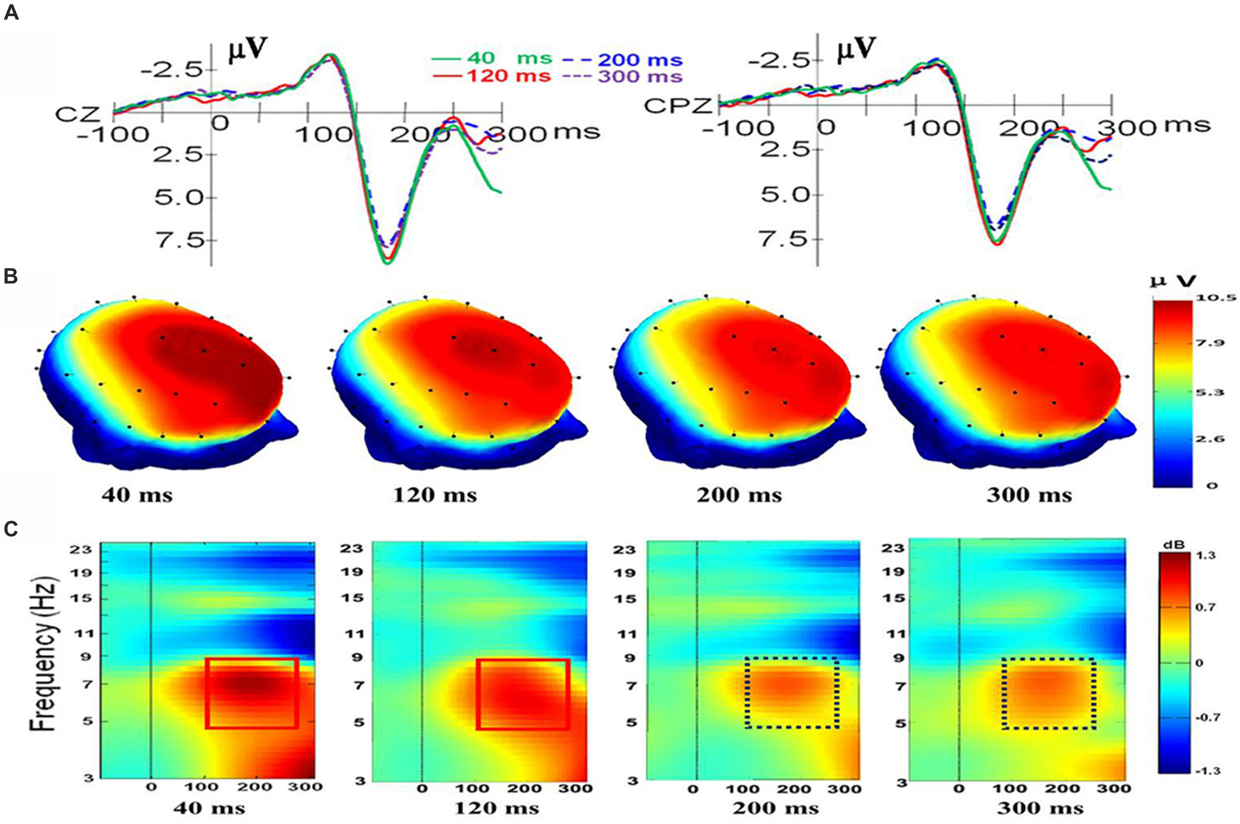

FIGURE 1 | The electroencephalogram (EEG)/event-related potentials (ERPs) results at the Cz and CPz electrode sites. (A) The grand-averaged ERPs waveforms (N170/VPP) elicited by a fleeting facial expression with a duration of 40 (green solid), 120 (red solid), 200 (blue dashed), and 300 ms (purple dashed) at the Cz and CPz electrode sites. **(B)** Scalp potential 3D maps reveal the topography of the N170/VPP for the time window (140–200 ms). **(C)** Event-Related Spectral Perturbation (ERSP) plot showing the mean increases or decreases in spectral power following stimulation. Non-green areas in the time/frequency plane show significant ($p < 0.01$) post-stimulus increases or decreases (see color scale) in log spectral power at the CPz electrode site relative to the mean power in the averaged 1-s pre-stimulus baseline (the interval for the ERSP analysis was −1000–1500 ms).

significantly higher than that of 200 and 300 ms ($p = 0.006$; $p = 0.039$). The comparisons found no significant difference in the theta response for pairs of expressions with durations of 40 and 120 ms or pairs of expressions with durations of 200 and 300 ms ($p = 0.308$; $p = 0.920$).

Source-Localization Data

Based on the scalp-recorded electric potential distribution, sLORETA was used to compute the cortical three-dimensional distribution of the current density of facial expressions with different durations. First, we explored standardized current density maxima for facial expressions with durations of 40, 120, 200, and 300 ms. All durations showed the same activation areas (fusiform gyrus, BA 20). To identify possible differences in the N170/VPP neural activation between the groups with durations of 40 and 300 ms, non-parametric statistical analyses of functional sLORETA images (Statistical non-Parametric Mapping; SnPM, c.f. Nichols and Holmes, 2002) were performed for the paired group while employing a t statistic (on log-transformed data). The results corresponded to maps of t statistics for each voxel, for a corrected $p < 0.05$. **Figure 2** shows sLORETA statistical non-parametric maps comparing the electric neuronal activity of recognizing expressions with durations of 40 and 300 ms at the N170/VPP latency of 140 to 200 ms. The **Figure 2** shows that the most active area of the cortex localized in the left hemisphere, in the Superior Frontal Gyrus (Brodmann area 8).

N400

ERPs Data

An ANOVA on the factors of duration (40, 120, 200, and 300 ms) and congruence (congruent, incongruent, and control) was performed on the mean amplitude (350–500 ms) of the N400 to determine whether the N400 effects were influenced by the different durations.

For the electrode Cz, there was no significant main effect of duration [$F(3,33) = 2.319$, $p = 0.093$, $\eta_p^2 = 0.174$], and the main effect of congruence was significant [$F(2,22) = 4.503$, $p = 0.023$, $\eta_p^2 = 0.290$]. The duration showed no significant interaction with congruence [$F(6,66) = 1.986$, $p = 0.080$, $\eta_p^2 = 0.153$]. For the electrode CPz, there was no significant main effect for duration [$F(3,33) = 2.250$, $p = 0.101$, $\eta_p^2 = 0.170$]. The main

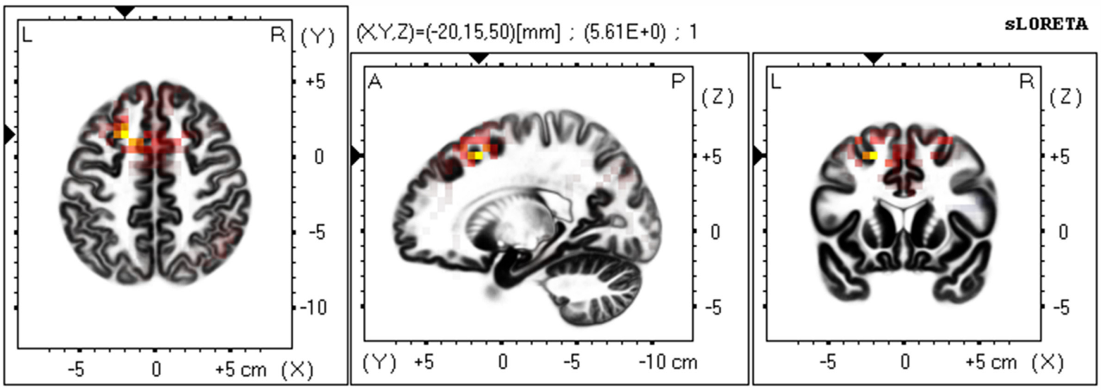

FIGURE 2 | The estimated sources of N170/VPP during a time window of 140–200 ms. sLORETA-based statistical non-parametric maps (SnPM) comparing the standardized current density values between facial expressions with durations of 40 and 300 ms ($n = 12$) at the N170/VPP latency (140–200 ms). Significantly increased activation ($p < 0.05$) at the 40 ms duration compared to the 300 ms duration is shown in red. Each map consists of axial, sagittal, and coronal planes. The maxima are color coded as yellow. L, left; R, right; A, anterior; P, posterior.

effect of congruence was significant [$F(2,22) = 3.731$, $p = 0.040$, $\eta_p^2 = 0.253$]. The effect of interaction of duration and congruence was not significant [$F(6,66) = 1.142$, $p = 0.348$, $\eta_p^2 = 0.094$]. Because there appears to be no effect of duration, **Figure 3** shows the N400 collapsed across all duration levels.

Source-Localization Data

Statistical analysis demonstrated significant differences ($p < 0.05$) between the levels of duration of 40 and 300 ms in the beta 2 (19–21 Hz) and beta 3 (22–30 Hz) frequency bands. In the beta 2 band, 371 voxels showed significant current-source density differences. In the beta 3 band, 964 voxels showed significant differences. A comparison of current density images between the 40 and 300 ms durations for beta 2 and beta 3 is shown in **Figure 4**. Yellow areas correspond to significantly higher activity in the 40 ms condition ($p < 0.05$, t threshold = 1.314).

DISCUSSION

The aim of the current study was to determine if there are different neural mechanisms underlying the recognition of microexpressions and macroexpressions. The results indicate that there are different ERPs and ERSP characteristics for recognizing microexpressions and macroexpressions. The brain regions responsible for the differences might be the inferior temporal gyrus and widespread areas in the frontal lobe. Furthermore, the left hemisphere was more involved in processing microexpressions. These results suggest that different neural mechanisms are responsible for the recognition of microexpressions and macroexpressions.

For expressions, there is a critical factor for recognition that is less well understood: the duration. A microexpression is presented for a short duration, which may result in the recipient barely perceiving it. The most commonly cited description of the duration of a microexpression: *microexpressions (1/25–1/5 of a second)*. Thus, the duration is the core difference between microexpressions and macroexpressions. Moreover, as ten Brinke and Porter (2013, p. 227) noted, a microexpression is "*a brief but complete facial expression.*" Therefore, the key characteristic differentiating microexpressions from macroexpressions is not the completeness of the expression (which may be related to intensity of emotion) but the duration of the expression. Considering that duration is the critical feature of a microexpression, in the current study, we manipulated the durations of expressions and expected that there would be different brain mechanisms for recognizing microexpressions and macroexpressions. The duration boundary may be around 200 ms, which can be used to differentiate a microexpression from a macroexpression (see Shen et al., 2012). The findings of this study show that recognizing expressions with durations of less than 200 ms and expressions with durations of greater than 200 ms are associated with different EEG/ERPs characteristics. Thus, we further confirmed that the boundary of the duration of expressions for differentiating microexpressions and macroexpressions is around 200 ms.

The present study manipulated the duration of facial expressions and examined the influence of duration on expression recognition by exploring the N170/VPP, the N400 effect and related EEG indicators. For the N170/VPP, there is a main effect of duration that clearly indicates the effects of duration on processing facial expressions with different durations. As shown in **Figure 1A**, there are two groups of ERPs, one for expressions with durations of greater than 200 ms and one for expressions with durations of less than 200 ms, suggesting that a duration boundary of 200 ms can differentiate microexpressions and macroexpressions. As for the interaction of duration and valence at electrode Cz, we should be cautious to draw any inference because there is no interaction at electrode CPz. The interaction of duration and valence should be elucidated further in the future.

As shown in **Figure 1A**, there is an enhanced N170/VPP in response to expressions with a short duration (<200 ms) compared to expressions with a long duration (>200 ms). On

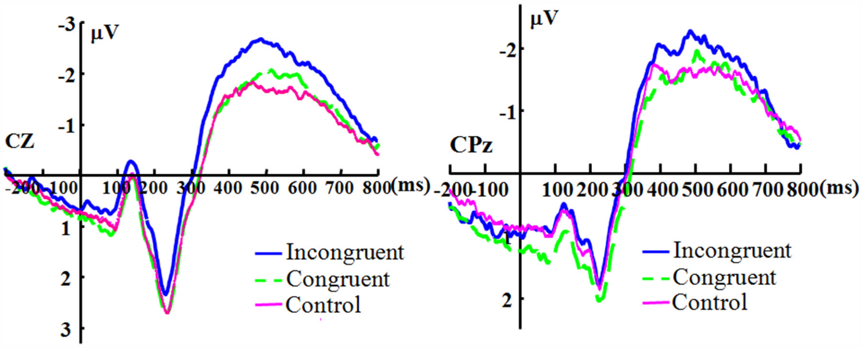

FIGURE 3 | The grand-averaged ERPs of the N400. The grand-averaged ERPs waveforms of the N400 under the conditions of incongruence (blue solid), congruence (green dashed), and control (red) at the Cz and CPz electrode sites.

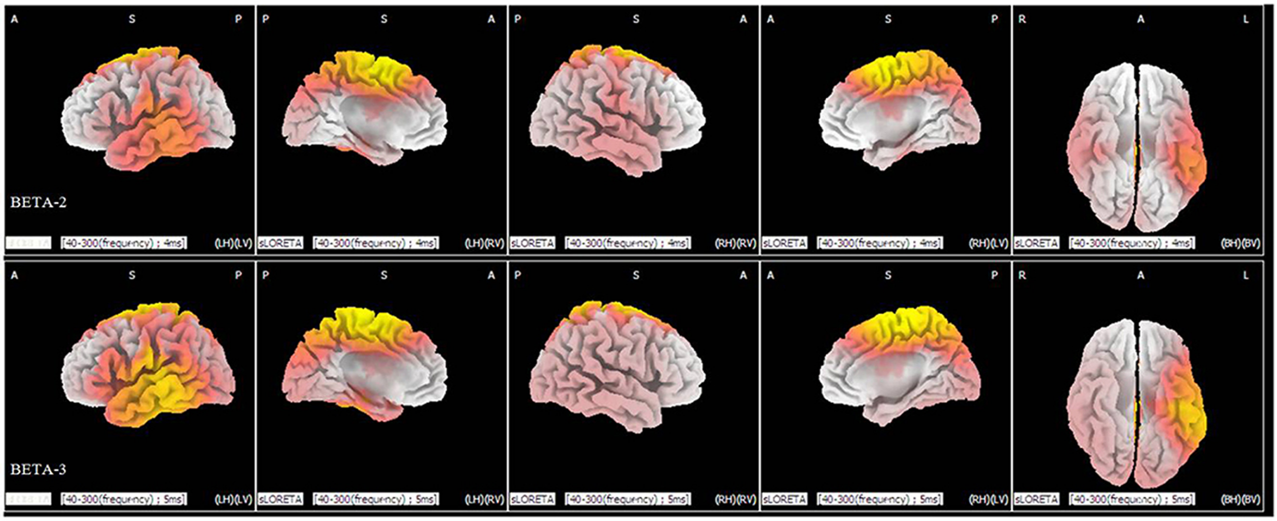
A S P
P S A
P S A
A S P
R A L
BETA-2
[40-300(frequency) ; 4ms] (LH)(LV)
sLORETA [40-300(frequency) ; 4ms] (LH)(RV)
sLORETA [40-300(frequency) ; 4ms] (RH)(RV)
sLORETA [40-300(frequency) ; 4ms] (RH)(LV)
sLORETA [40-300(frequency) ; 4ms] (BH)(BV)
A S P
P S A
P S A
A S P
R A L
BETA-3
[40-300(frequency) ; 5ms] (LH)(LV)
sLORETA [40-300(frequency) ; 5ms] (LH)(RV)
sLORETA [40-300(frequency) ; 5ms] (RH)(RV)
sLORETA [40-300(frequency) ; 5ms] (RH)(LV)
sLORETA [40-300(frequency) ; 5ms] (BH)(BV)

FIGURE 4 | sLORETA differences in two frequency bands (beta 2 and beta 3) between the 40 and 300 ms duration conditions (collapsed across all three congruence conditions). In the beta 2 **(upper** panel) and beta 3 frequency bands **(lower** panel), activity was significantly higher for 40 ms than for 300 ms in widespread areas, including the medial frontal gyrus and superior frontal gyrus. Images depicting statistical parametric maps observed from different perspectives are based on voxel by-voxel log of *F*-ratio values of differences between the two groups for the beta 2 and beta 3 bands. Structural anatomy is shown in gray scale (A, anterior; P, posterior; S, superior; I, inferior; LH, left hemisphere; RH, right hemisphere; BH, both hemispheres; LV, left view; RV, right view; BV, bottom view). Yellow indicates increases for 40 ms compared to 300 ms ($t_{0.05} = 1.314$, $t_{0.01} = 1.510$, one tail), which are mainly in the medial frontal gyrus and the superior frontal gyrus.

the one hand, the results may be due in part to attention (as a mediator variable). Attention to faces and facial expressions can modulate the N170 amplitude (Eimer, 2000, 2011; Eimer and Holmes, 2007). In the current study, recognizing expressions with short durations (e.g., 40 ms) may need significantly more attention resources (because short-duration expressions are somewhat difficult to perceive) than do expressions with long durations (e.g., 300 ms), which may result in a higher amplitude of N170/VPP for recognizing expressions with short durations. On the other hand, we automatically mimic the exposed facial expressions while recognizing them (Dimberg et al., 2000; Tamietto et al., 2009), if the exposing duration is short (say less than 200 ms, it is the case in microexpression recognition), then there is not much time to mimic the transient expression with short duration. Therefore, the mimicry of microexpression has to consult the memory to reach recognition, which may result in a stronger processing in the brain than the recognition of macroexpression, because recognizing macroexpression (with duration of greater than 200 ms) can only rely on the perceptual features of expressions.

As shown in **Figure 1**, sharp contrasts in scalp potential maps (**Figure 1B**) and ERSP (**Figure 1C**) are present between microexpressions (durations of less than 200 ms) and macroexpressions (durations of greater than 200 ms). The microexpressions elicited stronger power changes in theta band activities than did macroexpressions (see the comparison of the box of a solid line and the box of a dashed line in **Figure 1C**), which might also be interpreted as relating to the larger attention demands that are imposed on recognizing fleeting microexpressions.

In the current study, the ERSP results of N170/VPP showed that the amplitude of theta response was higher for microexpressions (with durations of less than 200 ms) than for macroexpressions (with durations of greater than 200 ms), which suggests that the theta response is also modulated by the duration of emotional expressions. Meanwhile, cognitive load may be related to the theta oscillatory activity (Bates et al., 2009). The reduced theta oscillatory activity for expressions with longer durations may be partly explained by cognitive load. For expressions with longer durations, there should be a lower cognitive load and there should be higher cognitive load for expressions with short durations. Meanwhile, the brain oscillations in the theta band are involved in active maintenance of memory representations (Jensen and Tesche, 2002). For expressions with shorter duration, one should make much more efforts to maintain the representations for further processing. For expressions with longer durations, one can check it anytime; therefore, the load for holding the representation is low. As shown in **Figure 2**, the neural generators that respond to the difference between recognizing expressions with durations of 40 and of 300 ms are located in the frontal lobe while measuring the N170/VPP, which is consistent some previous work that showed the frontal theta power increased with the cognitive load (Scheeringa et al., 2009). There are distinct EEG mu responses while viewing positively and negatively valenced emotional faces (see Moore et al., 2012), therefore, besides the beta rhythm, we should use mu response to further investigate the recognition of microexpression and macroexpression in ther future.

The statistical results of the N400 effects show no effects of duration and only a marginal significant interaction between duration and congruence at Cz, which does not support the prediction regarding the N400 effects (there should be a significant interaction). The effect of congruence is significant in the N400, which can be observed in **Figure 3**. The results suggest that even under the condition of short duration of expression, the participants could engage in top-down processing and the meaning of the expression was processed regardless of the duration (long or short in the current study), which implies that the fleeting emotional expressions (even with a duration of 40 ms) can be rapidly identified at a conceptual level (Potter, 2012). The results are consistent with some previous studies (Murphy and Zajonc, 1993; Milders et al., 2008).

As shown in **Figure 4**, there are significant differences in the profiles of the beta 2 and beta 3 powers between the 40 ms duration condition and the 300 ms duration condition, which suggests a strong involvement of beta-band synchronization in the processing of duration of an expression. Beta rhythm has been observed experimentally under the conditions of extensive recruitment of excitatory neurons (Whittington et al., 2000), suggesting there are more excitatory neurons for processing a facial expression with a duration of 40 ms than there are for an expression with a duration of 300 ms. Meanwhile, from the results of the sLORETA in **Figure 4**, we can observe that there is an increase in the power of beta activities. The locations are mainly in the frontal lobe and temporal lobe and involve more left than right hemispheric voxels (a similar neural activities pattern that involve more left than right hemispheric voxels can be seen in **Figure 2**). The results suggest there is left-hemisphere dominance for recognizing microexpressions. The lateralization of emotion has long been studied (Indersmitten and Gur, 2003) and many studies show evidence supporting right-hemispheric dominance for emotion processing (Schwartz et al., 1975; Hauser, 1993). There is, however, some debate regarding right-hemispheric dominance (De Winter et al., 2015). The current results show that the left hemisphere might respond during the processing of fleeting (<200 ms) expressions, regardless of valence. The effect of duration on the hemispheric dominance for emotional expressions processing should be further investigated and some objective indexes such as weighted lateralization index (see De Winter et al., 2015) should be provided.

It should be noted that the differences between the recognition of microexpressions and macroexpressions is not the same as the differences in recognizing supraliminal and subliminal facial expression (Balconi and Lucchiari, 2005). According to Shen et al. (2012), the accuracy of recognition for expressions with durations of 40 ms is above 40%, which is higher than the chance level (1/6), which means that the recognizing microexpression is conscious. Even the expressions cannot be perceived consciously, we still can unconsciously "resonate" the facial expressions we have seen during emotion communication (Dimberg et al., 2000; Tamietto et al., 2009), which may facilitate the recognition of microexpression and macroexpression.

In summary, we wanted to determine the exact differences in neural substrates for recognizing microexpression and macroexpression in the current study. The EEG/ERPs results revealed a distinct amplitude of the N170/VPP and oscillatory neuronal dynamics in response to microexpressions (with durations of less than 200 ms) and macroexpressions (with durations of greater than 200 ms). These results suggest that the EEG/ERPs characteristics are different between the recognition of microexpressions and macroexpressions.

Our understanding of how we perceive and recognize microexpressions and macroexpressions will be further advanced by studying the EEG/ERPs, their oscillatory neuronal dynamics, and their association with the processes of recognition. Based on this understanding of microexpression recognition, we can further explore the association between microexpressions and deception. Although controversial, microexpressions are closely related to deception and are used as a vital behavioral clue for lie detection (Frank and Svetieva, 2015). According to Weinberger (2010), few published peer-review studies address microexpressions for political reasons. Linking microexpressions to deception is "a leap of gargantuan dimensions" (for a review, see Weinberger, 2010). Many more studies are needed to understand the mechanisms underlying recognition of microexpression and its association with deception.

In the future, dynamic facial expressions with greater ecological validity should be employed. The brain mechanisms involved in recognizing a number of fleeting social emotions, including shame, guilt, and remorse (ten Brinke et al., 2012), and the fundamental properties of microexpressions recognition (Svetieva, 2014) should be explored. The influence of some factors, for instance, contextual cues (Van den Stock and de Gelder, 2014) which co-occur with facial microexpresion and macroexpression, age (Zhao et al., 2016), empathy (Svetieva and Frank, 2016), on how we recognize microexpresion and macroexpression and the underlying brain mechanisms should also be investigated.

AUTHOR CONTRIBUTIONS

XF and XS conceived the experiments. QW and KZ conducted the experiments. XS analyzed the results and wrote the main manuscript text and prepared all figures and tables. All authors reviewed the manuscript.

FUNDING

This project was partially supported by the National Natural Science Foundation of China (No. 61375009, 31460251) and the MOE (Ministry of Education in China) Youth Fund Project of Humanities and Social Sciences (Project No. 13YJC190018).

ACKNOWLEDGMENT

The authors would like to thank Bo Wang for his comments on the initial manuscript.

REFERENCES

Balconi, M., and Lucchiari, C. (2005). In the face of emotions: event-related potentials in supraliminal and subliminal facial expression recognition. *Genet. Soc. Gen. Psychol. Monogr.* 131, 41–69. doi: 10.3200/MONO.131.1.41-69

Balconi, M., and Pozzoli, U. (2007). Event-related oscillations (EROs) and event-related potentials (ERPs) comparison in facial expression recognition. *J. Neuropsychol.* 1, 283–294. doi: 10.1348/174866407X184789

Bates, A. T., Kiehl, K. A., Laurens, K. R., and Liddle, P. F. (2009). Low-frequency EEG oscillations associated with information processing in schizophrenia. *Schizophr. Res.* 115, 222–230. doi: 10.1016/j.schres.2009.09.036

Batty, M., and Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Cogn. Brain Res.* 17, 613–620. doi: 10.1016/S0926-6410(03)00174-5

Bhushan, B. (2015). "Study of facial micro-expressions in psychology," in *Understanding Facial Expressions in Communication*, eds M. K. Mandal and A. Awasthi (New Delhi: Springer), 265–286.

Buzsáki, G. (2006). *Rhythms of the Brain*. New York, NY: Oxford University Press.

Costa, T., Cauda, F., Crini, M., Tatu, M.-K., Celeghin, A., de Gelder, B., et al. (2014). Temporal and spatial neural dynamics in the perception of basic emotions from complex scenes. *Soc. Cogn. Affect. Neur.* 9, 1690–1703. doi: 10.1093/scan/nst164

De Winter, F.-L., Zhu, Q., Van den Stock, J., Nelissen, K., Peeters, R., de Gelder, B., et al. (2015). Lateralization for dynamic facial expressions in human superior temporal sulcus. *Neuroimage* 106, 340–352. doi: 10.1016/j.neuroimage.2014.11.020

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Dimberg, U., Thunberg, M., and Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychol. Sci.* 11, 86–89. doi: 10.1111/1467-9280.00221

Eimer, M. (2000). Attentional modulations of event-related brain potentials sensitive to faces. *Cogn. Neuropsychol.* 17, 103–116. doi: 10.1080/026432900380517

Eimer, M. (2011). "The face-sensitive N170 component of the event-related brain potential," in *Oxford handbook of Face Perception*, eds G. Rhodes, A. Calder, M. Johnson, and J. V. Haxby (New York, NY: Oxford University Press), 329–344.

Eimer, M., and Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia* 45, 15–31. doi: 10.1016/j.neuropsychologia.2006.04.022

Ekman, P. (1971). "Universals and cultural differences in facial expressions of emotion," in *Nebraska Symposium on Motivation*, ed. J. Cole (Lincoln: University of Nebraska Press), 207–283.

Ekman, P. (1992). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. New York, NY: Norton.

Ekman, P. (2003). *Emotions revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*. New York, NY: Times Books.

Ekman, P. (2009). "Lie Catching and Microexpressions," in *The Philosophy of Deception*, ed. C. Martin (New York, NY: Oxford University Press), 118–133.

Ekman, P., and Friesen, W. V. (1969). Nonverbal Leakage and Clues to Deception. *Psychiatry* 32, 88–106.

Ekman, P., Rolls, E., Perrett, D., and Ellis, H. (1992). Facial expressions of emotion: an old controversy and new findings [and discussion]. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 335, 63–69. doi: 10.1098/rstb.1992.0008

Esslen, M., Pascual-Marqui, R., Hell, D., Kochi, K., and Lehmann, D. (2004). Brain areas and time course of emotional processing. *Neuroimage* 21, 1189–1203. doi: 10.1016/j.neuroimage.2003.10.001

Frank, M. G., and Svetieva, E. (2015). "Microexpressions and deception," in *Understanding Facial Expressions in Communication*, eds M. K. Mandal and A. Awasthi (New Delhi: Springer), 227–242.

Fuchs, M., Kastner, J., Wagner, M., Hawes, S., and Ebersole, J. S. (2002). A standardized boundary element method volume conductor model. *Clin. Neurophysiol.* 113, 702–712. doi: 10.1016/S1388-2457(02)00030-5

Güntekin, B., and Başar, E. (2014). A review of brain oscillations in perception of faces and emotional pictures. *Neuropsychologia* 58, 33–51. doi: 10.1016/j.neuropsychologia.2014.03.014

Hauser, M. D. (1993). Right hemisphere dominance for the production of facial expression in monkeys. *Science* 261, 475–477. doi: 10.1126/science.8332914

Indersmitten, T., and Gur, R. C. (2003). Emotion processing in chimeric faces: hemispheric asymmetries in expression and recognition of emotions. *J. Neurosci.* 23, 3820–3825.

Jensen, O., and Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *Eur. J. Neurosci.* 15, 1395–1399. doi: 10.1046/j.1460-9568.2002.01975.x

Joyce, C., and Rossion, B. (2005). The face-sensitive N170 and VPP components manifest the same brain processes: the effect of reference electrode site. *Clin. Neurophysiol.* 116, 2613–2631. doi: 10.1016/j.clinph.2005.07.005

Jurcak, V., Tsuzuki, D., and Dan, I. (2007). 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage* 34, 1600–1611. doi: 10.1016/j.neuroimage.2006.09.024

Keil, A. (2013). "Electro-and magnetoencephalography in the study of emotion," in *The Cambridge Handbook of Human Affective Neuroscience*, eds J. Armony and P. Vuilleumier (New York, NY: Cambridge University Press), 107–132.

Knyazev, G. G., and Slobodskaya, H. R. (2003). Personality trait of behavioral inhibition is associated with oscillatory systems reciprocal relationships. *Int. J. Psychophysiol.* 48, 247–261. doi: 10.1016/S0167-8760(03)00072-2

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., et al. (2000). Automated Talairach atlas labels for functional brain mapping. *Hum. Brain Mapp.* 10, 120–131. doi: 10.1002/1097-0193(200007)10:3<120::AID-HBM30>3.0.CO;2-8

Lui, S., Huang, X., Chen, L., Tang, H., Zhang, T., Li, X., et al. (2009). High-field MRI reveals an acute impact on brain function in survivors of the magnitude 8.0 earthquake in China. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15412–15417. doi: 10.1073/pnas.0812751106

Makeig, S., Debener, S., Onton, J., and Delorme, A. (2004). Mining event-related brain dynamics. *Trends Cogn. Sci.* 8, 204–210. doi: 10.1016/j.tics.2004.03.008

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., et al. (2001). A probabilistic atlas and reference system for the human brain: international consortium for brain mapping (ICBM). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356, 1293–1322. doi: 10.1098/rstb.2001.0915

Metzinger, T. (2006). Exposing Lies. *Sci. Am. Mind* 17, 32–37. doi: 10.1038/scientificamericanmind1006-32

Milders, M., Sahraie, A., and Logan, S. (2008). Minimum presentation time for masked facial expression discrimination. *Cogn. Emot.* 22, 63–82. doi: 10.1080/02699930701273849

Moore, A., Gorodnitsky, I., and Pineda, J. (2012). EEG mu component responses to viewing emotional faces. *Behav. Brain Res.* 226, 309–316. doi: 10.1016/j.bbr.2011.07.048

Murphy, S. T., and Zajonc, R. B. (1993). Affect, cognition, and awareness: affective priming with optimal and suboptimal stimulus exposures. *J. Pers. Soc. Psychol.* 64, 723–739. doi: 10.1037/0022-3514.64.5.723

Newman, R., and Connolly, J. (2009). Electrophysiological markers of pre-lexical speech processing: evidence for bottom–up and top–down effects on spoken word processing. *Biol. Psychol.* 80, 114–121. doi: 10.1016/j.biopsycho.2008.04.008

Nichols, T. E., and Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp.* 15, 1–25. doi: 10.1002/hbm.1058

Niedenthal, P. M., and Brauer, M. (2012). Social functionality of human emotion. *Annu. Rev. Psychol.* 63, 259–285. doi: 10.1146/annurev.psych.121208.131605

Oostenveld, R., and Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiol.* 112, 713–719. doi: 10.1016/S1388-2457(00)00527-7

Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol.* 24(Suppl. D), 5–12.

Pascual-Marqui, R. D., Esslen, M., Kochi, K., and Lehmann, D. (2002). Functional imaging with low resolution brain electromagnetic tomography (LORETA): review, new comparisons, and new validation. *Jpn. J. Clin. Neurophysiol.* 30, 81–94.

Pascual-Marqui, R. D., Lehmann, D., Koenig, T., Kochi, K., Merlo, M. C. G., Hell, D., et al. (1999). Low resolution brain electromagnetic tomography (LORETA) functional imaging in acute, neuroleptic-naive, first-episode, productive schizophrenia. *Psychiatry Res.* 90, 169–179. doi: 10.1016/S0925-4927(99)00013-X

Pascual-Marqui, R. D., Michel, C. M., and Lehmann, D. (1994). Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int. J. Psychophysiol.* 18, 49–65. doi: 10.1016/0167-8760(84)90014-X

Porter, S., and ten Brinke, L. (2008). Reading between the lies: identifying concealed and falsified emotions in universal facial expressions. *Psychol. Sci.* 19, 508–514. doi: 10.1111/j.1467-9280.2008.02116.x

Potter, M. C. (2012). Conceptual short term memory in perception and thought. *Front. Psychol.* 3:113. doi: 10.3389/fpsyg.2012.00113

Righart, R., and De Gelder, B. (2008). Rapid influence of emotional scenes on encoding of facial expressions: an ERP study. *Soc. Cogn. Affect. Neur.* 3, 270–278. doi: 10.1093/scan/nsn021

Roach, B. J., and Mathalon, D. H. (2008). Event-related EEG time-frequency analysis: an overview of measures and an analysis of early gamma band phase locking in schizophrenia. *Schizophr. Bull.* 34, 907–926. doi: 10.1093/schbul/sbn093

Rossion, B., and Jacques, C. (2011). "The N170: understanding the time-course of face perception in the human brain," in *The Oxford Handbook of ERP Components*, eds S. Luck and E. Kapenhman (New York, NY: Oxford University Press), 115–142.

Scheeringa, R., Petersson, K. M., Oostenveld, R., Norris, D. G., Hagoort, P., and Bastiaansen, M. C. (2009). Trial-by-trial coupling between EEG and BOLD identifies networks related to alpha and theta EEG power increases during working memory maintenance. *Neuroimage* 44, 1224–1238. doi: 10.1016/j.neuroimage.2008.08.041

Schubert, S. (2006). A look tells all. *Sci. Am. Mind* 17, 26–31. doi: 10.1038/scientificamericanmind1006-26

Schwartz, G. E., Davidson, R. J., and Maer, F. (1975). Right hemisphere lateralization for emotion in the human brain: interactions with cognition. *Science* 190, 286–288. doi: 10.1126/science.1179210

Shen, X., Wu, Q., and Fu, X. (2012). Effects of the duration of expressions on the recognition of microexpressions. *J. Zhejiang Univ. Sci. B* 13, 221–230. doi: 10.1631/jzus.B1100063

Svetieva, E. (2014). *Seeing the Unseen: Explicit and Implicit Communication Effects of Naturally Occuring Emotion Microexpressions*. Ph.D. dissertation, State University of New York at Buffalo, New York, NY.

Svetieva, E., and Frank, M. G. (2016). Empathy, emotion dysregulation, and enhanced microexpression recognition ability. *Motiv. Emot.* 40, 309–320. doi: 10.1007/s11031-015-9528-4

Tamietto, M., Castelli, L., Vighetti, S., Perozzo, P., Geminiani, G., Weiskrantz, L., et al. (2009). Unseen facial and bodily expressions trigger fast emotional reactions. *Proc. Natl. Acad. Sci. U.S.A.* 106, 17661–17666. doi: 10.1073/pnas.0908994106

ten Brinke, L., MacDonald, S., Porter, S., and O'Connor, B. (2012). Crocodile tears: facial, verbal and body language behaviours associated with genuine and fabricated remorse. *Law Hum. Behav.* 36, 51–59. doi: 10.1037/h0093950

ten Brinke, L., and Porter, S. (2013). "Discovering deceit: applying laboratory and field research in the search for truthful and deceptive behavior," in *Applied Issues in Investigative Interviewing, Eyewitness Memory, and Credibility Assessment*, eds B. S. Cooper, D. Griesel, and M. Ternes (New York, NY: Springer), 221–237.

Van den Stock, J., and de Gelder, B. (2014). Face identity matching is influenced by emotions conveyed by face and body. *Front. Hum. Neurosci.* 8:53. doi: 10.3389/fnhum.2014.00053

Wang, B., and Fu, X. (2011). Time course of effects of emotion on item memory and source memory for Chinese words. *Neurobiol. Learn.* 95, 415–424. doi: 10.1016/j.nlm.2011.02.001

Wang, X., Wang, X., and Ma, H. (eds) (1993). *Rating Scales for Mental Health*. Beijing: Publisher of Chinese Mental Health Journal.

Weinberger, S. (2010). Airport security: intent to deceive? *Nature* 465, 412–415. doi: 10.1038/465412a

Whittington, M., Faulkner, H., Doheny, H., and Traub, R. (2000). Neuronal fast oscillations as a target site for psychoactive drugs. *Pharmacol. Ther.* 86, 171–190. doi: 10.1016/S0163-7258(00)00038-3

Zhao, M. F., Zimmer, H. D., Shen, X., Chen, W., and Fu, X. (2016). Exploring the cognitive processes causing the age-related categorization deficit in the recognition of facial expressions. *Exp. Aging Res.* 42, 348–364. doi: 10.1080/0361073X.2016.1191854

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ORIGINAL RESEARCH
published: 10 May 2017
doi: 10.3389/fpsyg.2017.00761

Neural Responses to Rapid Facial Expressions of Fear and Surprise

*Ke Zhao[1,2,3], Jia Zhao[4], Ming Zhang[5], Qian Cui[1,3] and Xiaolan Fu[1,3]**

[1] State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China, [2] Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China, [3] Department of Psychology, University of Chinese Academy of Sciences, Beijing, China, [4] Key Laboratory of Cognition and Personality (Ministry of Education) and Faculty of Psychology, Southwest University, Chongqing, China, [5] Department of Psychology, Dalian Medical University, Dalian, China

Facial expression recognition is mediated by a distributed neural system in humans that involves multiple, bilateral regions. There are six basic facial expressions that may be recognized in humans (fear, sadness, surprise, happiness, anger, and disgust); however, fearful faces and surprised faces are easily confused in rapid presentation. The functional organization of the facial expression recognition system embodies a distinction between these two emotions, which is investigated in the present study. A core system that includes the right parahippocampal gyrus (BA 30), fusiform gyrus, and amygdala mediates the visual recognition of fear and surprise. We found that fearful faces evoked greater activity in the left precuneus, middle temporal gyrus (MTG), middle frontal gyrus, and right lingual gyrus, whereas surprised faces were associated with greater activity in the right postcentral gyrus and left posterior insula. These findings indicate the importance of common and separate mechanisms of the neural activation that underlies the recognition of fearful and surprised faces.

Keywords: fearful face, surprised face, amygdala, recognition

OPEN ACCESS

Edited by:
Michael Noll-Hussong,
University of Ulm, Germany

Reviewed by:
Caroline F. Zink,
Lieber Institute for Brain Development, USA
Daniela Altavilla,
Sapienza University of Rome, Italy

****Correspondence:***
Xiaolan Fu
fuxl@psych.ac.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *23 November 2016*
Accepted: *26 April 2017*
Published: *10 May 2017*

Citation:
Zhao K, Zhao J, Zhang M, Cui Q and Fu X (2017) Neural Responses to Rapid Facial Expressions of Fear and Surprise. Front. Psychol. 8:761. doi: 10.3389/fpsyg.2017.00761

INTRODUCTION

Different emotions are associated with specific facial expressions, and the recognition of these facial expressions is important for social communication (Haxby et al., 2002). Among the six basic facial expressions (fear, sadness, surprise, happiness, anger, and disgust), fear and surprise are easily confused because surprised and fearful faces are "wide-eyed, information gathering" facial expressions (Kim et al., 2003, 2004; Zhao et al., 2013). A fearful expression involves open eyes and mouth and conveys shock in response to a frightening event, which signals a potential threat. A surprised expression also involves wide eyes and an open mouth, which indicate unexpectedness and novelty (Schroeder et al., 2004; Duan et al., 2010). According to Ekman's (1993) terminology, surprise is expressed by specific combinations involving two, three, or four action units, including the raised inner and outer brow, the raised upper eyelid, and the open mouth. Fear patterns also involve these action units; however, two specific action units, namely, the brow lower and the lip stretcher, might be part of fear patterns but not of surprise patterns (Ekman, 1993).

The recognition of facial expression is mediated by a distributed neural system (Haxby et al., 2000; Adolphs, 2002). This process is associated with increased activation in numerous visual areas (fusiform gyrus and lingual gyrus), temporal areas (middle/superior temporal gyrus and MTG), prefrontal areas (medial frontal gyrus and middle frontal gyrus), and limbic areas (amygdala and parahippocampal gyrus).

The discrimination of fear and surprise may be reflected in the brain activity patterns that underlie facial expression recognition. A fear expression indicates a potential threat, whereas surprise conveys a sense of novelty or unexpectedness (Adolphs et al., 1995; Schroeder et al., 2004). Fear has been described as negatively valenced surprise (Vrticka et al., 2014). Although no studies have directly investigated the different neural mechanisms that underlie these two faces, some brain regions have been found to be specialized for different emotional functions. The parahippocampal gyrus has been found to exhibit greater activation for surprised faces than fearful faces because surprised faces are consciously or unconsciously perceived due to their novelty (Schroeder et al., 2004; Duan et al., 2010). Correspondingly, the conscious and unconscious perception of faces with fearful expressions has been found to be associated with a significant amygdala response, which suggests a role of vigilance and the close monitoring of environmental cues (Morris et al., 1996; Whalen et al., 1998). However, other studies provide evidence that the human amygdala is also responsive to surprised facial expressions (Kim et al., 2003; Kim et al., 2004). A recent study revealed that poorer classification accuracy among all emotion categories was observed in the amygdala and hippocampus (Saarimaki et al., 2016).

As mentioned above, the specific brain regions that are most sensitive to fear or surprise remain unknown. To investigate the specific neural substrates, we directly contrasted the neural responses to fearful faces and surprised faces. In addition, previous studies have reported extremely high accuracies in the recognition of different emotions; however, the presentation times in these studies are long (Duan et al., 2010; Saarimaki et al., 2016). In a previous study, we found that performance in recognizing fearful and surprised faces was lower when the presentation time of the target face was short (100–500 ms) (Zhao et al., 2013). The present study used event-related functional magnetic resonance imaging (fMRI) to identify the neural substrates that mediate the perception of rapid surprised and fearful faces in healthy volunteers. By comparing the different patterns of neural activity in response to these faces, we identified similarities and differences between the mechanisms that underlie the recognition of fearful and surprised facial expressions.

MATERIALS AND METHODS

Subjects

Fifteen healthy subjects (8 males) aged 20.5 ± 1.24 years were recruited for the experiment. All of the subjects were right-handed, free of neurological or psychiatric diseases, and had normal or corrected-to-normal vision. The subjects were paid for their participation. The experimental procedures were approved by the IRB of the Faculty of Psychology, Southwest University, and informed written consent was obtained from all of the subjects.

Stimulation and Experimental Design

The present study investigated the perception of surprised and fearful faces. The target stimuli included images of two types of facial expressions (fear and surprise) posed by 43 individual human models from the NimStim database (Tottenham et al., 2009). Eighty-six images were selected from the database and trimmed to 192 × 220 pixels. The protocol was based on Ekman and Friesen's Brief Affect Recognition Test (Ekman and Friesen, 1974). In each trial, a black fixation cross was initially presented in the center of the silver–gray background for 200 ms, followed by a facial expression image presented in the center of the screen for 100, 300, or 500 ms. The subjects were instructed to identify the facial expression by using the right thumb to press a key ("1" or "2"). After the participants selected an answer, an inter-trial interval (ISI) was randomly inserted between the trials (**Figure 1**). The entire trial lasted 6 s, and the ISI did not include the fixation presentation, face presentation, and response time. We also included four blank intervals of 6 s duration among the trials.

Data Acquisition and Analysis

Functional magnetic resonance imaging data were acquired using a Siemens 3.0 Tesla Trio scanner with a standard head coil at the Key Laboratory of Cognition and Personality (Ministry of Education) at Southwest University (China). The functional scanning used a whole-brain gradient-echo, echo-planar-imaging sequence, and the repetition time was 2000 ms (30 ms echo time, 32 slices, 3.44 mm × 3.44 mm in-plane resolution, 1 mm slice gap, voxel size 3.4 × 3.4 × 4, field of view 220 mm × 220 mm, matrix 64 × 64, and flip angle = 90°).

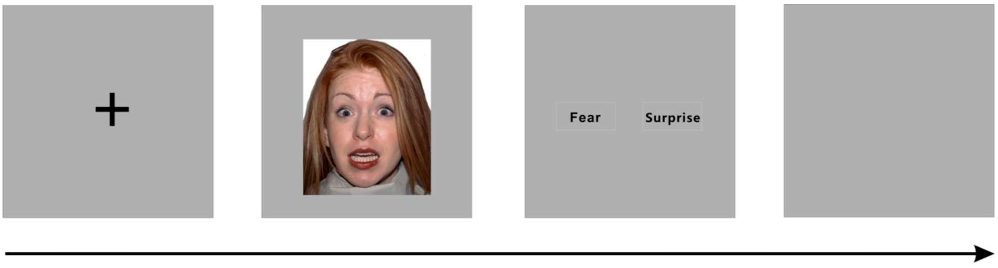

FIGURE 1 | Illustration of a single trial of facial expression recognition.

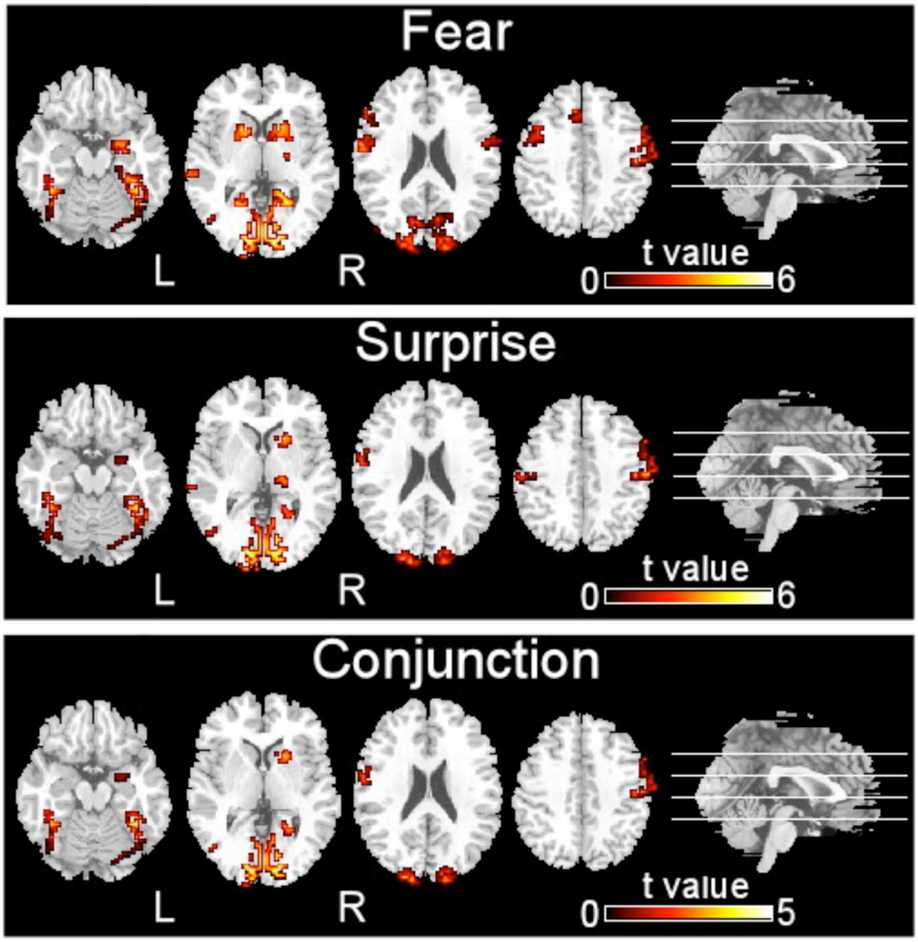

FIGURE 2 | Brain regions activated by two types of facial expressions, fear and surprise (p < 0.001, corrected with Monte Carlo simulations).

Complete fMRI data were acquired for 15 subjects and included in the following analysis. The data were preprocessed and analyzed using Statistical Parametric Mapping software SPM8 (Wellcome Trust Center for Neuroimaging, London, UK). The first five volumes for each subject were discarded to allow for signal equilibration. The images were slice-time corrected, motion corrected, normalized to the Montreal Neurological Institute (MNI) space at 3 mm × 3 mm × 3 mm, and spatially smoothed using a Gaussian kernel of 8 mm full width at half maximum (FWHM) (Ashburner and Friston, 2005). Then, two types of individual events (time-locked to the photographs) were modeled by a canonical hemodynamic response with two conditions: facial expressions of fear and surprise. A general linear model (GLM) was applied to the data to estimate the parameters of event-related activity corresponding to correct trials for each voxel in the volume under two conditions. Incorrect trials of both fearful faces and surprised faces were modeled separately in the GLM and discarded in the following analyses. Finally, statistical parametric maps with t-values were generated for each condition and each subject after first-level analysis (Calhoun et al., 2004).

A second-level random effects approach was applied to the group-level statistical analyses, which estimated the error variance of the interested conditions across subjects. During the second-level analysis, t-tests and conjunction analysis were applied to the two condition to identify the brain activations under each condition and the common activations of the two conditions, respectively. To examine the brain regions that are particularly involved in the perception of a specific emotional expression, the two emotional conditions were directly compared using paired t-tests (surprise vs. fear, fear vs. surprise). Multiple comparisons were applied to the inferences from the statistical parametric maps for the threshold corrected across gray matter in whole brain with Monte Carlo simulations (the cluster connection radius was 5 mm, and the number of Monte Carlo simulations was set to 1000) (Forman et al., 1995). The mask we used in the multiple correction with Monte Carlo simulations was extracted from WFU_PickAtlas software (gray matter in tissue type) (Maldjian et al., 2003) and then resampled to 3 × 3 × 3 volume as the gray matter mask (the number of voxels in the mask was 19956).

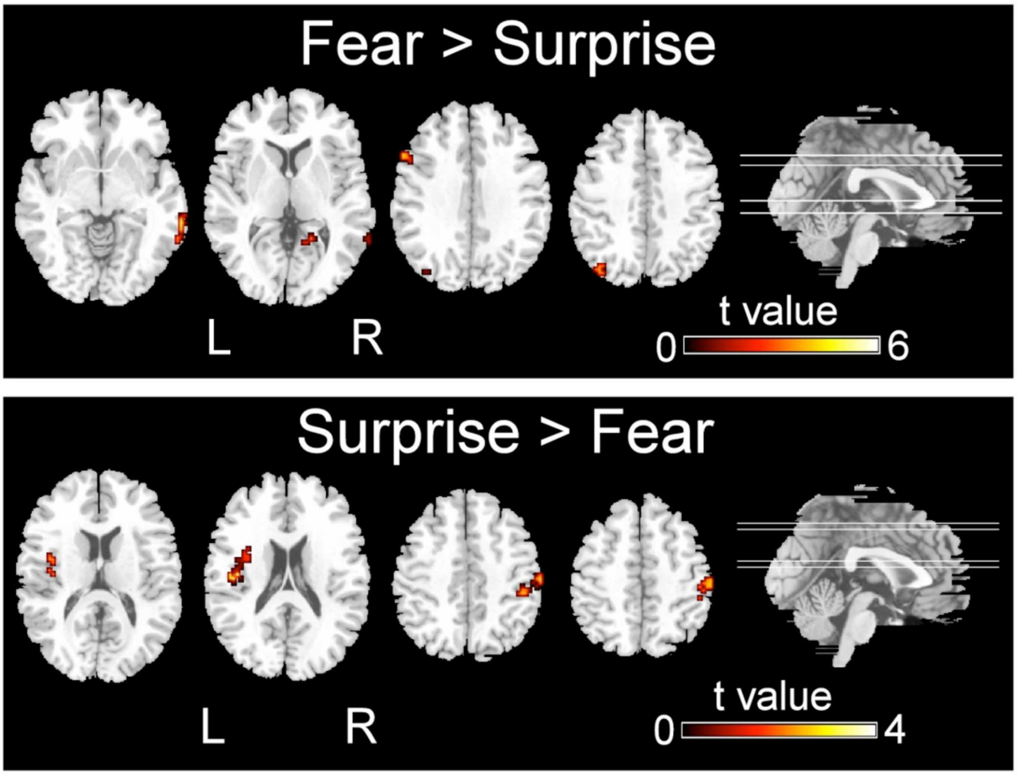

FIGURE 3 | Significant differences in the activation of brain regions during the recognition of fearful versus surprised faces ($p < 0.001$, corrected with Monte Carlo simulations).

In addition, a correlation analysis was utilized to assess the associations between the subject's sensitivity and brain activation under the two experimental conditions. The correlations between the sensitivity index (d') and the brain activations of each subject for each condition were calculated. The common areas that were significantly correlated with the recognition score under the two face stimuli were extracted as regions of interest (ROIs) using the MarsBar toolbox[1]. Then, the brain activities in the constructed ROIs were analyzed.

RESULTS

There was no difference in recognition accuracy scores between fearful faces (0.78 ± 0.08) and surprised faces (0.77 ± 0.11; $t = 0.52, p = 0.61$). We initially determined the brain regions that exhibited increased activation when the subjects watched the two types of facial expressions (**Figures 2**, **3**). To illustrate the detailed activation information, the MNI coordinates of the peak T-values and voxel numbers for all significant clusters were extracted and are displayed in **Tables 1–5**.

The brain regions that exhibited increased activation in response to fearful faces included the left postcentral gyrus, left middle temporal gyrus, left cuneus, left putamen, left inferior occipital gyrus, left precentral gyrus, left supplementary motor area, right precentral gyrus, right inferior occipital gyrus, right parahippocampal gyrus, and right amygdala ($p < 0.001$, corrected with Monte Carlo simulations; **Figure 2** and **Table 1**). Compared to fearful faces, surprised faces were associated with increased activation of the left postcentral gyrus, left middle occipital gyrus, left supplementary motor area, right lentiform nucleus, right calcarine, right postcentral gyrus, right precentral gyrus, right inferior occipital gyrus, right parahippocampal gyrus, and right amygdala ($p < 0.001$, corrected; **Figure 2** and **Table 2**).

The conjunction analysis revealed that the brain regions that exhibited increased activation in response to both the surprised and fearful faces included the left postcentral gyrus, left middle occipital gyrus, left fusiform, left inferior occipital gyrus, left cuneus, left supplementary motor area, right postcentral gyrus, right inferior occipital gyrus, right calcarine, right putamen, right parahippocampal gyrus, and right amygdala (**Figure 2**).

Regarding the differences in the perceptual processing of fearful faces versus surprised faces, the significant clusters included the left precuneus, left middle frontal gyrus, right MTG and right lingual gyrus for the contrast between the fear and surprise conditions ($p < 0.001$, corrected; **Figure 3** and **Table 4**). For the contrast between the surprise and fear conditions, differences were located at two clusters, including the left posterior insula and right postcentral gyrus ($p < 0.001$, corrected; **Figure 3** and **Table 5**).

[1] http://marsbar.sourceforge.net

TABLE 1 | Neural activity in response to facial expression of fear.

Brain region	MNI co-ordinates *x, y, z*	Volume (voxels)
L middle temporal gyrus	–51, –76,10	52
L putamen	–18,11,4	43
L inferior occipital gyrus	–42, –79, –11	26
L postcentral gyrus	–60, –7,16	117
L cuneus	–15, –91,1	480
L precentral gyrus	–42,8,31	132
L supplementary motor area	–6,14,52	79
R precentral gyrus	57, –10,49	332
R inferior occipital gyrus	39, –79, –8	30
R amygdala	27, –7, –14	203
R parahippocampal gyrus	18, –49, –5	539

L = left, R = Right. Significant at corrected p < 0.001.

TABLE 2 | Neural activity in response to facial expression of surprise.

Brain region	MNI co-ordinates *x, y, z*	volume (voxels)
L middle occipital gyrus	–15, –103,10	360
L supplementary motor area	–3,11,55	38
L postcentral gyrus	–57, –16,31	114
R lentiform nucleus	21,14,4	64
R inferior occipital gyrus	39, –79, –8	29
R calcarine	3, –82,1	440
R postcentral gyrus	57, –16,52	241
R parahippocampal gyrus /amygdala	27, –7, –14	20
R precentral gyrus	54, –7,10	19

L = left, R = Right. Significant at corrected p < 0.001.

TABLE 3 | Conjunction of neural activity for facial expressions of fear and surprise.

Brain region	MNI co-ordinates *x, y, z*	volume (voxels)
L middle occipital gyrus	–15, –103,7	97
L fusiform	–42, –43, –23	38
L inferior occipital gyrus	–42, –79, –11	26
L postcentral gyrus	–60, –13,28	64
L cuneus	–15, –91,1	162
L supplementary motor area	–3,11,55	37
R postcentral gyrus	57, –16,52	199
R inferior occipital gyrus	39, –79, –8	29
R calcarine	3, –82,1	416
R putamen	21,14,4	62
R parahippocampal gyrus (amygdala)	27, –7, –14	20

L = left, R = Right. Significnat at corrected p < 0.001.

Correlation analyses were employed to examine the relationship between sensitivity of discrimination between two faces (a score calculated as the *Z* score for a correct response minus the *Z* score for a false alarm) and brain activity (**Figure 4**). The activity of the right postcentral area was significantly correlated with this sensitivity index under the fearful face condition ($r = 0.52$, $p < 0.05$) and under the surprised face condition ($r = 0.61$, $p < 0.05$).

TABLE 4 | Neural activity showing more activation for fear than for surprise.

Brain region	MNI co-ordinates x, y, z	volume (voxels)
L precuneus	–39, –73,37	14
L middle frontal gyrus	–57,17,34	10
R middle temporal gyrus	63, –40, –14	36
R lingual gyrus	18, –49,1	12

L = left, R = Right. Significant at corrected p < 0.001.

TABLE 5 | Neural activity showing more activation for surprise than for fear.

Brain region	MNI co-ordinates *x, y, z*	volume (voxels)
L insula	–45, –19,19	29
R postcentral gyrus	42, –34,34	47

L = left, R = Right. Significant at corrected p < 0.001.

DISCUSSION

The current findings indicated similarities and differences in the neural mechanisms that underlie the recognition of fearful and surprised faces. In the present study, brain regions within the temporal and occipital cortices, such as the left fusiform gyrus, were activated during the perception of fearful and surprised faces, which may indicate these brain regions are involved in the general perceptual recognition of facial expressions (Haxby et al., 2000; Winston et al., 2004). Regions of the occipital and temporal visual cortices play a critical role in the perceptual processing of socially and emotionally relevant visual stimuli (Haxby et al., 2000, 2002). Increased activation of these areas may represent top-down modulatory effects on the visual processing stream, which reflect attentional enhancement as a result of emotional

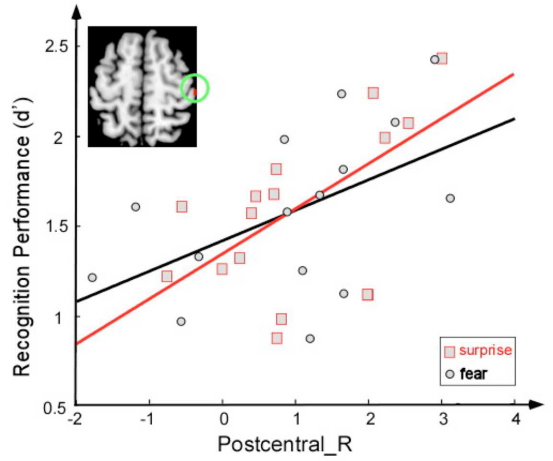

FIGURE 4 | Correlation between the sensitivity index (*d*′) and activation magnitude (*T*-value) under two conditions: fearful face stimuli (black) and surprised face stimuli (red).

significance (Vuilleumier and Schwartz, 2001; Pessoa et al., 2002). In addition, fearful faces appear sufficient to evoke increased amygdala activation. Our results indicate that the amygdala (particularly in the right hemisphere) is responsive to surprised faces and are consistent with a previous study reporting that the right amygdala was activated in response to both fearful and surprised faces (Kim et al., 2003). The right parahippocampal gyrus was similarly activated during the recognition of fearful and surprised faces. The amygdala and hippocampus are strongly interconnected and receive inputs from extrastriate visual areas in the occipital and temporal cortices (Amaral and Insausti, 1992; Morris et al., 1999). Our findings indicated that the amygdala and parahippocampal gyrus form an important part of the emotion network but are unable to distinguish between fearful and surprised faces. This result is consistent with a previous study that found that although limbic regions, including the amygdala, hippocampus, and thalamus, appear to form an important part of the emotion network, the limbic components of the network revealed poorer classification accuracy than did the cortical components (Vrticka et al., 2014).

Our results indicate that fearful faces induced more activation than did surprised faces in the frontal and temporal lobes. The middle frontal gyrus was activated during fearful face recognition. Previous research has indicated that this brain region is implicated in contingency awareness in human aversive conditioning (Knight et al., 2004; Carter et al., 2006). The 'attentional network' has been extensively researched and is thought to involve fronto-parietal regions, including the middle frontal gyrus (MFG) (Pessoa, 2009). Thus, the activity of this region may reflect the attention being paid to fearful faces. Neurons in the human MTG respond to socially important aspects of faces such as expression, orientation, and eye-gaze direction (Perrett et al., 1985; Hasselmo et al., 1989). In a study by Morris et al. (1998), the right MTG received a greater contribution from the amygdala during the processing of fearful expressions (Morris et al., 1998). Depth EEG results have indicated that the amygdala is activated along with the MTG (Krolak-Salmon et al., 2004). A previous study identified the activation of this region during the recognition of fear versus disgust (Phillips et al., 1998). In other work, functional activation specifically associated with a fearful face prime was found in the activated bilateral middle temporal gyrus (Fan et al., 2011). In addition, anomia for facial emotions has been reported in patients with lesions in the right middle temporal gyrus (Rapcsak et al., 1993; Cornwell et al., 2008). The activation of this brain region might be due to the reception and correct labeling of potential threat information from fearful faces.

The facial expression of surprise has a distinct character and might be universally recognized. Psychological theories suggest that surprise is an adaptive mechanism to restructure and extend cognitive concepts following the analysis of an unexpected event (Schutzwohl, 1998); moreover, it provides important indicators of emotion with respect to unexpectedness and novelty (Schroeder et al., 2004). In the present study, surprised faces induced greater activation in the postcentral cortices than did fearful faces, which suggests that additional activity in this region was required to correctly recognize surprised faces. The sensitivity of recognition between two faces was positively correlated with the activation of this area for both the fearful face and surprised face conditions. One interpretation of these findings is that viewing facial expressions of emotion triggers an emotional response in the perceiver that mirrors the emotion presented in the stimulus (Pitcher et al., 2008; Wood et al., 2016). Moreover, the representation of this emotional response in the somatosensory cortices may provide information regarding the emotion. In particular, the somatosensory, motor, and premotor cortices have been associated with emotion recognition in research with lesion patients (Adolphs et al., 2000) and research using transcranial magnetic stimulation (TMS) (Pourtois et al., 2004; Pitcher et al., 2008). Regarding the posterior insula, previous studies have suggested that the left and right insula preferentially encode positive and negative affect, respectively (Craig, 2009). Left insular activation has been identified in subjects experiencing joy (Takahashi et al., 2008). Damage to this area may impair gustatory information processing (Calder et al., 2001). Thus, the greater activation of this brain region in the surprise condition might be attributed to the surprised face being experienced as more positive than the fearful face. Fear was described as negatively valenced surprise in a recent study (Vrticka et al., 2014).

CONCLUSION

The present study used fMRI to explore the activation of different brain regions in response to fearful and surprised faces. Our results indicate that the limbic system, including the amygdala and parahippocampal gyrus, is responsible for both of these faces. The fearful faces elicited greater activation in some frontal regions and the right middle temporal gyrus, whereas the insula and postcentral cortices were largely activated in the recognition of surprised faces. These results suggest that fear leads to greater activation of the attention and memory systems, whereas surprise results in greater activation of the emotion experience system.

ETHICS STATEMENT

The experimental procedures were approved by the local ethics committee in Southwest University (China).

AUTHOR CONTRIBUTIONS

KZ, JZ, XF contributed in designing the experiment, analyzing the data, and writing the manuscript. MZ contributed in collecting the data and analyzing the data, and QC contributed in writing the manuscript.

ACKNOWLEDGMENT

This research was supported by grant from the National Natural Science Foundation of China (61375009, 31400876).

REFERENCES

Adolphs, R. (2002). Neural systems for recognizing emotion. *Curr. Opin. Neurobiol.* 12, 169–177. doi: 10.1016/S0959-4388(02)00301-X

Adolphs, R., Damasio, H., Tranel, D., Cooper, G., and Damasio, A. R. (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *J. Neurosci.* 20, 2683–2690.

Adolphs, R., Tranel, D., Damasio, H., and Damasio, A. R. (1995). Fear and the human amygdala. *J. Neurosci.* 15, 5879–5891.

Amaral, D. G., and Insausti, R. (1992). Retrograde transport of D-[3H]-aspartate injected into the monkey amygdaloid complex. *Exp. Brain Res.* 88, 375–388. doi: 10.1007/BF02259113

Ashburner, J., and Friston, K. J. (2005). Unified segmentation. *Neuroimage* 26, 839–851. doi: 10.1016/j.neuroimage.2005.02.018

Calder, A. J., Lawrence, A. D., and Young, A. W. (2001). Neuropsychology of fear and loathing. *Nat. Rev. Neurosci.* 2, 352–363. doi: 10.1038/35072584

Calhoun, V. D., Stevens, M. C., Pearlson, G. D., and Kiehl, K. A. (2004). fMRI analysis with the general linear model: removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. *Neuroimage* 22, 252–257. doi: 10.1016/j.neuroimage.2003.12.029

Carter, R. M., O'Doherty, J. P., Seymour, B., Koch, C., and Dolan, R. J. (2006). Contingency awareness in human aversive conditioning involves the middle frontal gyrus. *Neuroimage* 29, 1007–1012. doi: 10.1016/j.neuroimage.2005.09.011

Cornwell, B. R., Carver, F. W., Coppola, R., Johnson, L., Alvarez, R., and Grillon, C. (2008). Evoked amygdala responses to negative faces revealed by adaptive MEG beamformers. *Brain Res.* 1244, 103–112. doi: 10.1016/j.brainres.2008.09.068

Craig, A. D. (2009). How do you feel–now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70. doi: 10.1038/nrn2555

Duan, X., Dai, Q., Gong, Q., and Chen, H. (2010). Neural mechanism of unconscious perception of surprised facial expression. *Neuroimage* 52, 401–407. doi: 10.1016/j.neuroimage.2010.04.021

Ekman, P. (1993). Facial expression and emotion. *Am. Psychol.* 48, 384–392. doi: 10.1037/0003-066X.48.4.384

Ekman, P., and Friesen, W. V. (1974). Detecting deception from body or face. *J. Pers. Soc. Psychol.* 29, 288–298. doi: 10.1037/h0036006

Fan, J., Cu, X. S., Liu, X., Guise, K. G., Park, Y., Martin, L., et al. (2011). Involvement of the anterior cingulate and frontoinsular cortices in rapid processing of salient facial emotional information. *Neuroimage* 54, 2539–2546. doi: 10.1016/j.neuroimage.2010.10.007

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., and Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn. Reson. Med.* 33, 636–647. doi: 10.1002/mrm.1910330508

Hasselmo, M. E., Rolls, E. T., and Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behav. Brain Res.* 32, 203–218. doi: 10.1016/S0166-4328(89)80054-3

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci.* 4, 223–233. doi: 10.1016/S1364-6613(00)01482-0

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biol. Psychiatry* 51, 59–67. doi: 10.1016/S0006-3223(01)01330-0

Kim, H., Somerville, L. H., Johnstone, T., Alexander, A. L., and Whalen, P. J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. *Neuroreport* 14, 2317–2322. doi: 10.1097/01.wnr.0000101520.44335.20

Kim, H., Somerville, L. H., Johnstone, T., Polis, S., Alexander, A. L., Shin, L. M., et al. (2004). Contextual modulation of amygdala responsivity to surprised faces. *J. Cogn. Neurosci.* 16, 1730–1745. doi: 10.1162/0898929042947865

Knight, D. C., Cheng, D. T., Smith, C. N., Stein, E. A., and Helmstetter, F. J. (2004). Neural substrates mediating human delay and trace fear conditioning. *J. Neurosci.* 24, 218–228. doi: 10.1523/JNEUROSCI.0433-03.2004

Krolak-Salmon, P., Henaff, M. A., Vighetto, A., Bertrand, O., and Mauguiere, F. (2004). Early amygdala reaction to fear spreading in occipital, temporal, and frontal cortex: a depth electrode ERP study in human. *Neuron* 42, 665–676. doi: 10.1016/S0896-6273(04)00264-8

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., and Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage* 19, 1233–1239. doi: 10.1016/S1053-8119(03)00169-1

Morris, J. S., Friston, K. J., Buchel, C., Frith, C. D., Young, A. W., Calder, A. J., et al. (1998). A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain* 121(Pt 1), 47–57. doi: 10.1093/brain/121.1.47

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., et al. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature* 383, 812–815. doi: 10.1038/383812a0

Morris, J. S., Ohman, A., and Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proc. Natl. Acad. Sci. U.S.A.* 96, 1680–1685. doi: 10.1073/pnas.96.4.1680

Perrett, D. I., Smith, P. A., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., et al. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proc. R. Soc. Lond. B Biol. Sci.* 223, 293–317. doi: 10.1098/rspb.1985.0003

Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends Cogn. Sci.* 13, 160–166. doi: 10.1016/j.tics.2009.01.006

Pessoa, L., Kastner, S., and Ungerleider, L. G. (2002). Attentional control of the processing of neural and emotional stimuli. *Brain Res. Cogn. Brain Res.* 15, 31–45. doi: 10.1016/S0926-6410(02)00214-8

Phillips, M. L., Young, A. W., Scott, S. K., Calder, A. J., Andrew, C., Giampietro, V., et al. (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proc. Biol. Sci.* 265, 1809–1817. doi: 10.1098/rspb.1998.0506

Pitcher, D., Garrido, L., Walsh, V., and Duchaine, B. C. (2008). Transcranial magnetic stimulation disrupts the perception and embodiment of facial expressions. *J. Neurosci.* 28, 8929–8933. doi: 10.1523/JNEUROSCI.1450-08.2008

Pourtois, G., Sander, D., Andres, M., Grandjean, D., Reveret, L., Olivier, E., et al. (2004). Dissociable roles of the human somatosensory and superior temporal cortices for processing social face signals. *Eur. J. Neurosci.* 20, 3507–3515. doi: 10.1111/j.1460-9568.2004.03794.x

Rapcsak, S. Z., Comer, J. F., and Rubens, A. B. (1993). Anomia for facial expressions: neuropsychological mechanisms and anatomical correlates. *Brain Lang.* 45, 233–252. doi: 10.1006/brln.1993.1044

Saarimaki, H., Gotsopoulos, A., Jaaskelainen, I. P., Lampinen, J., Vuilleumier, P., Hari, R., et al. (2016). Discrete neural signatures of basic emotions. *Cereb. Cortex* 26, 2563–2573. doi: 10.1093/cercor/bhv086

Schroeder, U., Hennenlotter, A., Erhard, P., Haslinger, B., Stahl, R., Lange, K. W., et al. (2004). Functional neuroanatomy of perceiving surprised faces. *Hum. Brain Mapp.* 23, 181–187. doi: 10.1002/hbm.20057

Schutzwohl, A. (1998). Surprise and schema strength. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 1182–1199. doi: 10.1037/0278-7393.24.5.1182

Takahashi, H., Matsuura, M., Koeda, M., Yahata, N., Suhara, T., Kato, M., et al. (2008). Brain activations during judgments of positive self-conscious emotion and positive basic emotion: pride and joy. *Cereb. Cortex* 18, 898–903. doi: 10.1093/cercor/bhm120

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Res.* 168, 242–249. doi: 10.1016/j.psychres.2008.05.006

Vrticka, P., Lordier, L., Bediou, B., and Sander, D. (2014). Human amygdala response to dynamic facial expressions of positive and negative surprise. *Emotion* 14, 161–169. doi: 10.1037/a0034619

Vuilleumier, P., and Schwartz, S. (2001). Emotional facial expressions capture attention. *Neurology* 56, 153–158. doi: 10.1212/WNL.56.2.153

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., and Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *J. Neurosci.* 18, 411–418.

Winston, J. S., Henson, R. N., Fine-Goulden, M. R., and Dolan, R. J. (2004). fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *J. Neurophysiol.* 92, 1830–1839. doi: 10.1152/jn.00155.2004

Wood, A., Rychlowska, M., Korb, S., and Niedenthal, P. (2016). Fashioning the face: sensorimotor simulation contributes to facial expression recognition. *Trends Cogn. Sci.* 20, 227–240. doi: 10.1016/j.tics.2015.12.010

Zhao, K., Yan, W. J., Chen, Y. H., Zuo, X. N., and Fu, X. (2013). Amygdala volume predicts inter-individual differences in fearful face recognition. *PLoS ONE* 8:e74096. doi: 10.1371/journal.pone.0074096

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

METHODS
published: 13 October 2017
doi: 10.3389/fpsyg.2017.01745

Dual Temporal Scale Convolutional Neural Network for Micro-Expression Recognition

Min Peng[1,2†], *Chongyang Wang*[1,2†], *Tong Chen*[1,2,3*], *Guangyuan Liu*[1,2] *and Xiaolan Fu*[3]

[1] *Chongqing Key Laboratory of Non-linear Circuit and Intelligent Information Processing, Southwest University, Chongqing, China,* [2] *School of Electronic and Information Engineering, Southwest University, Chongqing, China,* [3] *Institute of Psychology, University of Chinese Academy of Sciences, Beijing, China*

Facial micro-expression is a brief involuntary facial movement and can reveal the genuine emotion that people try to conceal. Traditional methods of spontaneous micro-expression recognition rely excessively on sophisticated hand-crafted feature design and the recognition rate is not high enough for its practical application. In this paper, we proposed a Dual Temporal Scale Convolutional Neural Network (DTSCNN) for spontaneous micro-expressions recognition. The DTSCNN is a two-stream network. Different of stream of DTSCNN is used to adapt to different frame rate of micro-expression video clips. Each stream of DSTCNN consists of independent shallow network for avoiding the overfitting problem. Meanwhile, we fed the networks with optical-flow sequences to ensure that the shallow networks can further acquire higher-level features. Experimental results on spontaneous micro-expression databases (CASME I/II) showed that our method can achieve a recognition rate almost 10% higher than what some state-of-the-art method can achieve.

Keywords: micro-expression recognition, deep learning, optical flow, convolutional neural network, feature fusion

OPEN ACCESS

Edited by:
Ping Hu,
Renmin University of China, China

Reviewed by:
Tommaso Costa,
Università degli Studi di Torino, Italy
Li Liu,
Chongqing University, China

****Correspondence:***
Tong Chen
c_tong@swu.edu.cn

[†] *These authors have contributed equally to this work.*

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *14 April 2017*
Accepted: *20 September 2017*
Published: *13 October 2017*

Citation:
Peng M, Wang C, Chen T, Liu G and Fu X (2017) Dual Temporal Scale Convolutional Neural Network for Micro-Expression Recognition. Front. Psychol. 8:1745. doi: 10.3389/fpsyg.2017.01745

INTRODUCTION

Facial expression plays an important role in people's daily communication and emotion expression. Typically, a full facial expression last from 1/2 to 4 s (Ekman, 2003b) and can be easily identified by humans. Over the past few decades, many researchers have made their efforts to help computer better understand facial expressions and the form of emotional communications among humans (Fasel and Juergen, 2003; Zhang and Tjondronegoro, 2011; Li X. et al., 2013; Li Y. et al., 2013). However, psychological studies (Porter and Ten Brinke, 2008; Ekman, 2009) indicate that the recognition of human emotion based on facial expressions may be misleading. In other words, someone may try to hide their emotion by exerting an opposite facial expression.

As a special facial expression, micro-expression is defined as a rapid facial movement that is not subject to people's consciousness and can reveal the genuine emotion (Ekman, 2003a). Micro-expression was first discovered by Haggard and Isaacs (1966), they found the Micro-expression is related to self-defense mechanism and can reveal depressed emotions. In 1969, Ekman and Friesen also observed a specific kind of micro-expression when they were analyzing a video from a depressive patient who attempted to tell a lie to cover his suicidal intent. In that video, the patient was optimistic by observing his facial expression, but when the video was played in a slower speed and inspected frame by frame, Ekman et al. saw an intense expression of extreme anguish just within two frames as the patient was answering a question from the doctor. The short expression lasted <1/12 s. From then on, understanding and recognizing micro-expression became a popular research topic (Russell et al., 2006; Endres and Laidlaw, 2009; Pfister et al., 2011).

For its authenticity and objectivity, micro-expression recognition possesses great value in diverse fields, such as, affect monitoring (Porter and Ten Brinke, 2008), criminal detection (Russell et al., 2006), and homeland security (Weinberger, 2010). However, due to its characteristics, micro-expression recognition is very challenging. Firstly, micro-expressions are fleeting and imperceptible, which typically last <1/2 s and can be easily neglected by human eyes (Yan et al., 2013a). Secondly, its intensity is very subtle and localized (Porter and Ten Brinke, 2008), i.e., micro-expression is a tiny movement confined to a small area of the face region. In 2009, Frank et al. found that only highly trained individuals are able to distinguish various micro-expressions, but the recognition accuracy is just 47%.

Related Research Works

For the reason of the difficulty for human to notice or recognize micro-expressions, in recent years, automatic facial micro-expressions recognition has attracted increasing attentions in both the field of pattern recognition and computer vision (Polikovsky et al., 2009; Pfister et al., 2011). Polikovsky et al. (2009) presented a 3D-Gradient orientation histogram descriptor to represent the motion information in facial micro-expressions. Shreve et al. (2011) proposed a spatio-temporal strain method for automatic micro-expression spotting in long-term videos. Wu et al. (2011) designed an automatic micro-expression recognition system by using Gabor feature and GentleSVM classifier.

Thanks to Pfister et al. (2011), Li X. et al. (2013), Yan et al. (2013b, 2014), three spontaneous micro-expressions database (SMIC, CASMEI, and CASMEII) were built in well designed and strictly controlled laboratory environment and publicly introduced to the community. A brief summary of these three databases are given in **Table 1**. Based on the spontaneous database, many methods for micro-expression recognition have been proposed. Pfister et al. (2011) performed the first successful attempt in spontaneous facial micro-expression recognition. By Combining the Local Binary Pattern on Three Orthogonal Planes (LBP-TOP) descriptor and Random Forest (RF) classifier, the best accuracy of 78.9% on the SMIC database was obtained. Considering the redundant information in LBP-TOP features, Wang et al. (2014) proposed a LBP-Six Intersection Points (LBP-SIP) method and the experiment on CASMEII database shows that the LBP-SIP is more accurate and computational efficient than LBP-TOP. Huang et al. (2016) considered more information such as, sign, magnitude and orientation and proposed Spatiotemporal Completed Local Quantization Patterns (STCLQP) for facial micro-expression analysis. Compared with the LBP-TOP and LBP-SIP methods, STCLQP achieves a substantial improvement on recognition rate tested on the three public spontaneous micro-expression databases. Aside from concentrating on Spatiotemporal Local Texture Descriptors (SLTD) based methods, a more comprehensive research is done by Liu et al. (2015). In their work, a simple but efficient method called Main Directional Mean Optical-flow (MDMO) was employed, which utilized optical flow estimation technique to compute the subtle movement of facial regions of interest (ROIs) that were spotted based on the Facial Action Coding System (FACS). For 36 ROIs, the length of a MDMO feature vector is just 72. Besides, they also proposed an optical-flow-driven method to align all frames of a micro-expression video clip. To address the problem of constant head movements in typical micro-expression applications, Xu et al. (2017) presented Facial Dynamics Map (FDM) to characterize micro-expression. Based on Facial Landmark Location, "Coarse Alignment and Face Cropping" were conducted on the raw micro-expression clips, then a pixel-level alignment method was applied before FDM feature extraction. By classifying more categories and taking a different measuring method of recognition rate, the recognition accuracy on three databases (SMIC, CASMEI, and CASMEII) are 71.43, 42.02, and 41.96%, respectively.

TABLE 1 | Three main spontaneous micro-expression database.

	Index			
	Clips number	**Camera speed**	**Frame size**	**AU coding/Labeling**
SMIC	164	100 fps	640 × 480	No/By Emotion
CASME I	195	60 fps	Part A: 1280 × 720 Part B: 640 × 480	Yes/By Emotion
CASME II	247	200 fps	640 × 480	Yes/By Emotion

The aforementioned works make solid contribution in automatic micro-expression recognition and inspire the community. However, there is still space to improve the methods. Firstly, the methods rely excessively on hand-crafted features and the process of feature selection depends heavily on the experience of researchers, which makes it difficult for psychologist lack of such experience to use the methods. Secondly, the recognition rate of the methods is not high enough for practical applications. Therefore, a more effective method that can generate high-level feature automatically for micro-expressions recognition is desired.

Related Research Works

Convolutional Neural Networks (CNNs) (LeCun et al., 1998), as an effective deep learning model, has recently made unprecedented progress in many fields such as, computer vision (Szegedy et al., 2015), speech recognition (Abdel-Hamid et al., 2012), and natural language processing (Sutskever et al., 2014). Some popular CNN models like LeNet-5 (LeCun et al., 1998), AlexNet (Krizhevsky et al., 2012), GoogLeNet (Szegedy et al., 2015), and VGG-Net (Simonyan and Zisserman, 2014a) are well tested and widely used by many researchers. In spite of the difference in network structure, these popular deep networks are all shown their powerful ability for understanding the property of raw data. Except for 2D information processing, Karpathy et al. (2014) extended the connectivity of CNN to time domain and introduced a video descriptor to learn the spatio-temporal information. In the experiment on UCF101 Action Recognition dataset that contains 1 million YouTube videos belonging to 487 classes, the best recognition rate reached 63.9%.

In those successful works of CNN, large dataset is needed to train the network. However, the micro-expression database that we can use so far is much smaller than traditional database fed to CNN. A serious overfitting problem would occur if we directly apply CNN on the existing micro-expression database. In this paper, The proposed Dual Temporal Scale Convolutional Neural Network (DTSCNN) addressed the overfitting problem from three aspects: (i) the feature extraction was done on the micro-expression clips by using two shallow network separately; (ii) data augmentation and higher drop-out ratio were used in each network; (iii) CASMEI and CASMEII database were used together to train the network.

Meanwhile, the shallow network of DTSCNN has the risk of only learning low-level features. To ensure the proposed architecture can obtain high-level features, the data fed to the network was not raw data but the optical-flow, which is higher level feature than raw data and has been proved to be effective in micro-expression recognition (Liu et al., 2015).

The proposed DTSCNN is a two-stream convolutional network, each stream is a simplified network that uses 3D convolution kernel and pooling cell (Tran et al., 2015) to automatically represent the property of subtle facial movements. Because the frame rates of the video clips in CASMEI and CASMEII were 60 and 200 fps, respectively. One stream of the DTSCNN took 64 fps input ($64 = 2^6$ adapts to CASMEI), and the other stream took 128 fps input ($128 = 2^7$ adapts to CASMEII). Neither do we need the sophisticated frame alignment method nor the complicated feature design. The DTSCNN takes optical-flow sequences in different temporal scales as the input and outputs their higher level features. Experimental results on CASME I/II database demonstrate that our proposed method gave higher recognition rate than some state-of-the-art recognition methods, such as, STCLQP (Huang et al., 2016), MDMO (Liu et al., 2015), and FDM (Xu et al., 2017).

The following sections are organized as: section Convolutional Neural Networks gives a brief introduction of Deep learning (DL), and Convolutional Neural Network (CNN) principle; section Micro-Expression Recognition describes our proposed DTSCNN; section Experiments Results and Analysis presents and discusses the experiment design and results; section Conclusion gives the conclusion.

CONVOLUTIONAL NEURAL NETWORKS

In this section, we give a brief introduction of Deep learning (DL) and the Convolutional Neural Network (CNN) principle, which lays a foundation for proposing DTSCNN in section Micro-Expression Recognition.

Deep Learning

Deep learning is evolved from the research on neural networks. Typically, it is composed of multiple processing layers and has powerful abilities to learn representations of data using multiple levels of abstraction. Currently, many deep network structures have been put forward. Such as, Deep Belief Network (Hinton et al., 2006), Stacked Auto Encoders (Vincent et al., 2010), Convolutional Neural Network (LeCun et al., 1998), and Recurrent Neural Network (Mikolov et al., 2011). For the dramatically great success of CNN in visual object recognition and detection, in this paper, we mainly discuss the CNN for micro-expression recognition.

Convolutional Neural Network (CNN)

CNN is a biologically-inspired model and firstly proposed by LeCun et al. (1998). Shown in **Figure 1** is a general structure of a CNN.

In **Figure 1**, the input layer receives normalized images with identical size. A set of units in a small neighborhood (local receptive field) in the input layer will be processed by a convolution kernel to form the unit in a feature map of the subsequent convolutional layer. One pixel in the feature map can be calculated by using

$$C_k = f(x * W + b) \tag{1}$$

where C_k is the value of the k-th pixel in the feature map, x is the pixel value vector of the units in the local receptive field corresponding to C_k, W, and b are the coefficient vector and bias, respectively, determined by the feature map, while f is the activation function (sigmoid, tanh, ReLU, etc.). Since studies in

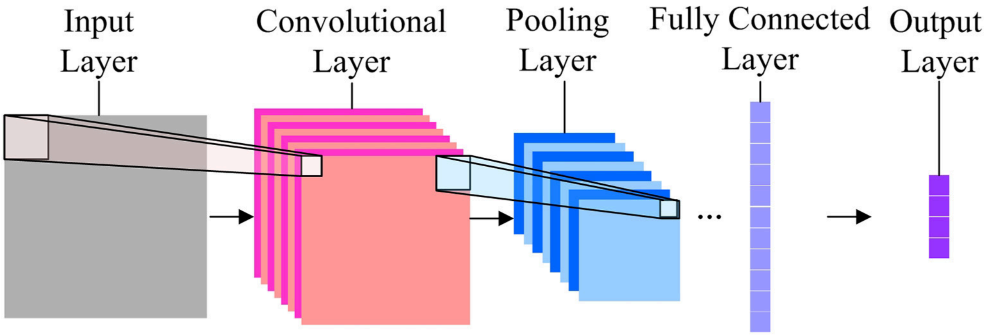

FIGURE 1 | The General structure of a CNN.

Nair and Hinton (2010) have suggested that ReLU function is superior to sigmoid function, in our work, the ReLU function has been employed. For the input$t$, ReLU function can be expressed as

$$f(t) = \max(0, t) \tag{2}$$

Each feature map has only one convolutional kernel, i.e., for all x in the input plane, the W and b are the same. This design of CNN can largely save calculation time and make specific feature stand out in a feature map. There is normally more than one feature map in a convolutional layer, so that multiple features are included in the layer.

To make the feature invariant to the geometrical shift and distortion, the convolutional layer is followed by a pooling layer which can subsample the feature maps. For the k-th unit in a feature map in the pooling layer, its value can be calculated by using

$$P_k = f(\beta * down(C) + \alpha) \tag{3}$$

where P_k is the value of the k-th unit in feature map in the pooling layer, C is the value vector in the feature map of the convolutional layer, β and α are the coefficient and bias, respectively, and $down\,(\bullet)$ is the subsampling function. Max pooling function is used for subsampling, in that case, $down\,(C)$can be written as

$$down(C) = \max\left\{C_{s,l}\,\middle|\, s \le m, l \le m, s, l \in z^{+}\right\} \tag{4}$$

where $C_{s,l}$ is the pixel value of the unit C in the feature map, $m$is the subsampling size.

The first convolutional layer and pooling layer would acquire low-level information of the image, while the stack of them would enable high-level feature extraction.

The output layer is connected to its formal layer with Softmax Regression. For the output vector F from the upper layer, the probability of classifying into class c is:

$$p\left(y^{(F)} = c|F;\theta\right) = \frac{e^{\theta_c^T F}}{\sum_{n=1}^{N} e^{\theta_n^T F}}\ 1 \le c \le N \tag{5}$$

where $y^{(F)}$ is the group identity of input F, θ is weight vector between output layer and previous layer, N is the number of the groups. The loss function is defined as:

$$J(\theta) = -\sum_{c=1}^{N} 1\left\{y^{(F)} = c\right\}\log\frac{e^{\theta_c^T F}}{\sum_{n=1}^{N} e^{\theta_n^T F}} 1 \le c \le N \tag{6}$$

Where, $1\{\bullet\}$ is the eigenfunction, when $\{\bullet\}$ is true, it will return 1. Practically, in CNN training, we would compute the sum of loss function from multiple inputs, and update the weight of network using stochastic gradient descent (Wilson and Martinez, 2003).

MICRO-EXPRESSION RECOGNITION

Pre-Processing

At the stage of data pre-processing, two techniques are contained: face alignment and normalization. In face alignment, we take the method presented in Yan et al. (2014). In their method, 68 facial landmarks are detected in the first frame in each micro-expression video clips using Active Shape Model (ASM) (Cootes et al., 1995). Then the first frame of each sequence is normalized according to the alignment template, the subsequent frames in each clips are all aligned to the first frame by using Local Weighted Mean (LWM) transformation (Goshtasby, 1988). In normalization, we normalize the aligned micro-expression samples both in spatial and temporal domain. For spatial domain normalization, all images are cropped within face region to 96 × 112 pixels, which is in the average size of the original face region in the database. For temporal normalization, we employ the linear interpolation method to obtain a sufficient number of frames. The linear interpolation method is widely used and proved to be effective in frame normalization (Liu et al., 2015; Xu et al., 2017). As mentioned in the early section, the training set that we used contains two subsets, where video clips are normalized to 65 frames and 129 frames, respectively, to compensate for frame differences of those two databases.

Optical Flow Estimation

Optic-flow technique can detect the motion information between adjacent frames. In analyzing visual motion information, optical flow is typically served as a high level feature in machine learning area. Recently, some large-scale video classification works with CNNS (Simonyan and Zisserman, 2014b; Tran et al., 2015) has also suggested that optical flow sequences are more efficient to use than the original image sequences.

In a video clip, suppose that $I\left(x, y, t\right)$ is the value at point $\left(x, y, t\right)$. After a lapse of δt to the next frame, the pixel moved to $\left(x + \delta x, y + \delta y, t + \delta t\right)$ with its intensity I $\left(x + \delta x, y + \delta y, t + \delta t\right)$. Based on invariance of brightness during small period, we have

$$I\left(x, y, t\right) = I\left(x + \delta x, y + \delta y, t + \delta t\right) \tag{7}$$

where $\delta x = u\delta t, \delta y = v\delta t$, with $u\left(x, y\right)$ and $v\left(x, y\right)$ to be the horizontal component and vertical component that need to be estimated in the optical flow field.

If we assume that the pixel value in an image is a continuous function of its position and time, according to the Taylor series expansion, the right part of the function (7) can be written as:

$$\begin{aligned} I\left(x + \delta x, y + \delta y, t + \delta t\right) &= I\left(x, y, t\right) + \delta x\frac{\partial I}{\partial x} + \delta y\frac{\partial I}{\partial y} \\ &\quad + \delta t\frac{\partial I}{\partial t} + \varepsilon \end{aligned} \tag{8}$$

Where ε is the two order or above unbiased estimator of time δt. When δt tends to be infinitesimal, we can let both sides of formula (8) to be divided by time δt and the Equation (7), then the optical flow equation is obtained as follows:

$$\frac{\delta x}{\delta t}\frac{\partial I}{\partial x} + \frac{\delta y}{\delta t}\frac{\partial I}{\partial y} + \frac{\partial I}{\partial t} = 0 \tag{9}$$

that is,

$$u\frac{\partial I}{\partial x}+v\frac{\partial I}{\partial y}+\frac{\partial I}{\partial t}=0 \tag{10}$$

For video clips of micro-expression, computing the tiny movement of facial region accurately is crucial before recognition. In Liu's work (Liu, 2009), he made subtle movements in the video more obvious by computing its optical flow estimation, which is also suitable for us in order to recognize the micro-expression information. Further, the matrix of u, v fields can be transformed to image by using Munsell Color System (Gargi et al., 2000). **Figure 2** shows two pre-processed samples and their optical field estimations from CASME I/II. To human eyes, it is hard to notice the facial change in those clips. However, in optical flow fields, we could demonstrate the subtle movement in different colors.

DTSCNN

DTSCNN is a two streams network with 3D convolution and pooling units. Unlike typical convolution or pooling cell in convolutional neural network, the 3D convolution and pooling in DTSCNN have a kernel in size of k × k × l, where k is spatial size, l is temporal depth. The micro-expression clip that we refer to in DSTCNN has a size of d × w × h × c, where w, h, and c are width, height and number of channels of every single frame, respectively, and d is the number of frames.

In an input layer of a typical convolutional neural network, every single image is treated as an object to be identified. Nevertheless, in video classification, each video clip is used as a bag of words and fed into the network. In our work, we calculated the optical flow estimation in size of 64 × 96 × 112 × 3 and 128 × 96 × 112 × 3 for each micro-expression video clip in CASME I/II dataset.

For continuous-time visual information processing, temporal information is as important as spatial information. However, how to probe the spatio-temporal information sufficiently and effectively is critical to video identification task. In Karpathy's work (Karpathy et al., 2014), three connectivity patterns of convolution neural network in video identification task were presented. **Figure 3** shows these three kinds of fusion model.

In **Figure 3**, The Late Fusion model is similar to parallel convolutional neural network and each single-frame network shares parameters in a fixed frame distance. The Early Fusion model design is based on single-frame networks and only utilizes 3D convolution with a size of k × k × l in the first layer to extract the spatio-temporal information. The Slow Fusion model is a more comprehensive combination, which utilizes the 3D convolution and pooling technique throughout the network while learning more elaborate information from both spatial and temporal domains. Although this would progressively generate higher-level information, it is slow and memory-consuming.

For micro-expressions recognition, considering that the micro-expression is continuous and is not contained in a specific frame or few adjacent frames, the Late Fusion and Early Fusion may be inadequate. In addition, Karpathy's (Karpathy et al., 2014) and Du's (Tran et al., 2015) experiments show that Slow Fusion model can give better performance than Late and Early fusion

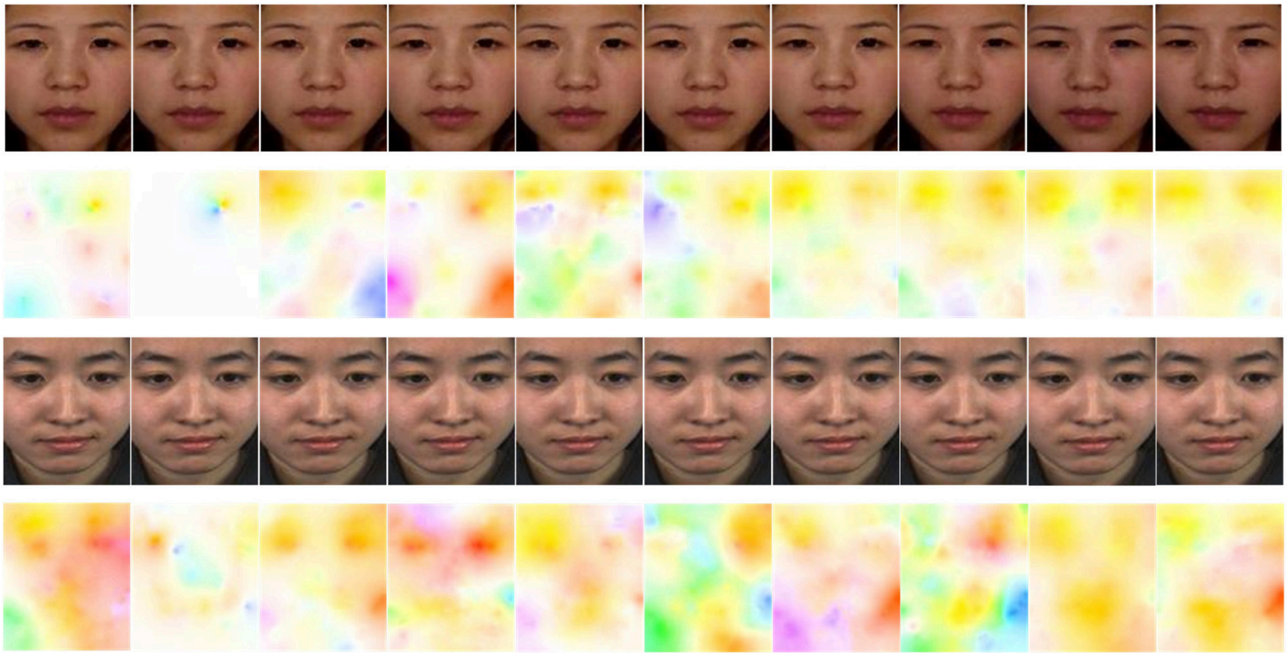

FIGURE 2 | Two samples are pre-processed and estimated optical fields. The first row and third row micro-expression sequence are from CASMEI (subject 01, EP12_3, frames: 45–54) and CASMEII (subject 17, EP05_02, frames: 53–62) database, respectively. The former one expressed a surprise emotion and the latter expressed a positive emotion (happiness). Row 2 and 4 is the optical flow sequence that computed from their above line, respectively.

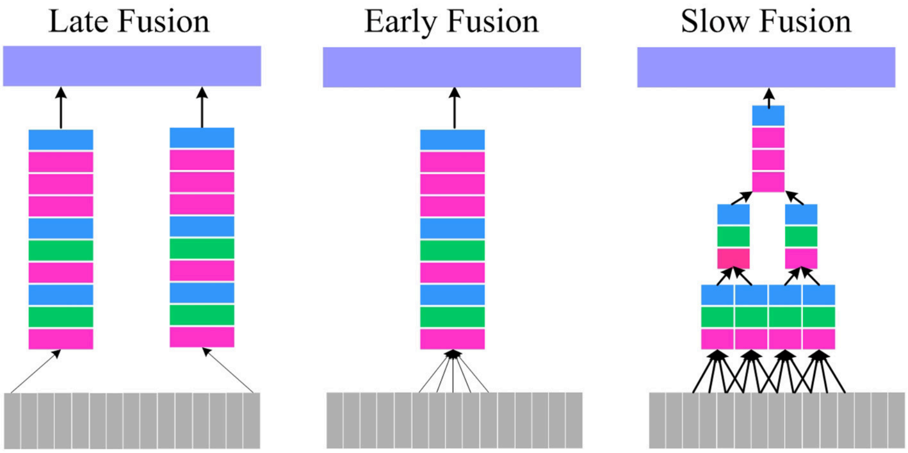

FIGURE 3 | Fusion model. **Left**: Late Fusion. **Middle**: Early Fusion. **Right**: Slow Fusion. The red, green, blue, and purple box represent convolutional, pooling, normalization and Fully-Connected (FC) layers, respectively.

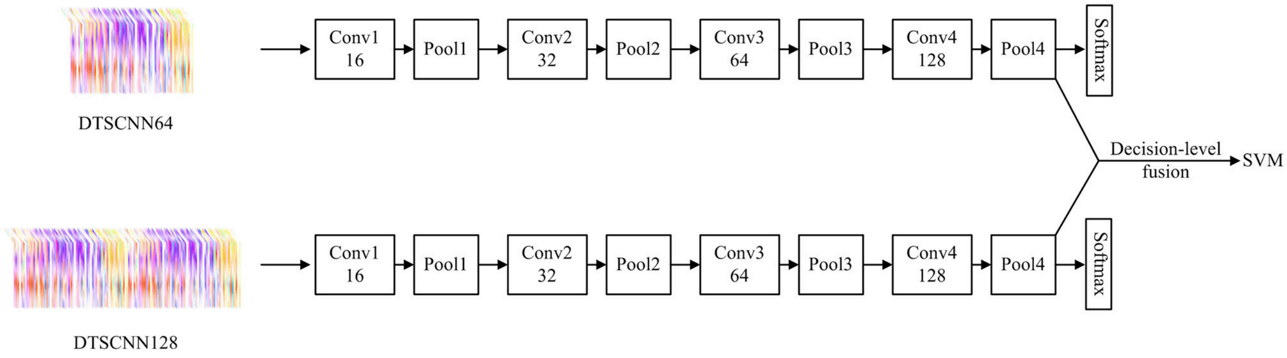

FIGURE 4 | Dual Temporal Scale Convolutional Neural Network (DTSCNN).

model. Especially, Du et al proposed the C3D (Tran et al., 2015) and proved that a 3 × 3 × 3 convolution kernel used in every layer would give the best performance. Therefore, in our work, we combine the Slow Fusion model and C3D implement for micro-expression recognition. Specially, a DTSCNN is proposed and the architecture is shown in **Figure 4**.

In **Figure 4**, we can see that DTSCNN is a two-stream convolutional network consisting of DTSCNN64 and DTSCNN128. Each stream is compact with only 5 layers (4 convolutional layers and 1 fully-connected layer, the number of filters for the four convolution layers is 16, 32, 64, and 128, respectively. The detail of the kernel parameter setting of the network is given in **Table 2**. In the first convolution layer (3 × 3 × 8 conv or 3 × 3 × 16 conv), a big spatial and temporal stride is set to omit redundant information in that initial level and save memories. The setting of second and third layer (3 × 3 × 3 conv) follow Du's (Tran et al., 2015) conclusions. The reasons of the fourth layer utilizing 3 × 3 × 4 convolutional filter is that a 3 × 3 × 3 convolution filter may create more temporally indefinite factors when it operates previous layers that with 4-frames length. The last layer is an output layer since keeping an extra FC layer consumes time and memory. Using a two-stream architecture can not only overcome the frame rate difference between CASMEI and CASMEII but also the overfitting problem due to small data size. Also, taking the optical-flow data as input can help the simplified network to learn high-level feature. When learning is finished, a linear SVM classifier is used to take features from the final layer of each stream. The result of the SVM classifier is used for decision-level fusion to give the overall recognition rate.

The DTSCNN64 and DTSCNN128 are designed to take micro-expression video clips in size of 64 × 96 × 112 × 3 and 128 × 96 × 112 × 3, respectively. The DSTCNN64 is used to adapt to the frame rate of 60fps of CASMEI, and the DSTCNN128 to CASMEII. This design is important in real application. Because there is no agreed standard frame rate so far for recoding the micro-expressions, i.e., the micro-expression video could be recorded in various frame rate. The design of different streams of the network can adapt to different frame rates, which may make the whole network robust to the frame rate of the input data. The prediction falls into four different classes. **Figure 5** shows the detail of DTSCNN64.

EXPERIMENTS RESULTS AND ANALYSIS

Database and Experiment Setting

In CASMEI database (Yan et al., 2013b), there are 189 spontaneous micro-expression video clips collected from 19 subjects. Each clip was filmed by a 60-fps camera with a size of 640 × 480 pixels. The data can be classified into 8 classes. Compared with CASMEI, CASMEII (Yan et al., 2014) is more like an updated version. Namely, it contains 255 spontaneous micro-expression video clips from 26 subjects and includes emotion belonging to seven classes. Especially, it was recorded by camera with a speed of 200 fps and the face region occupies a larger proportion in the image. In our experiment, we selected data from CASMEI and CASMEII to form the experiment dataset CASME I/II. Following the recommended strategy (Yan et al., 2013b, 2014), we categorized CASME I/II into four classes: Negative, Others, Positive and Surprise. The specific emotion that each class contains and the number of clips in CASME I/II are shown in **Table 3**.

TABLE 2 | Parameters and size of the kernel in DTSCNN.

Layer	DTSCNN64 (Kernel parameter settings)	DTSCNN128 (Kernel parameter settings)
Input	–	–
Conv1	3 × 3 × 8, Sp:1,Ss:2,Tp:2,Ts:4	3 × 3 × 16, Sp:1,Ss:2,Tp:4,Ts:8
pool1	2 × 2 × 1, Ss:2,Ts:1	2 × 2 × 1,Ss:2,Ts:1
Conv2	3 × 3 × 3,Sp:1,Ss:1,Tp:1,Ts:1	3 × 3 × 3,Sp:1,Ss:1,Tp:1,Ts:1
pool2	2 × 2 × 2,Ss:2,Ts:2	2 × 2 × 2,Ss:2,Ts:2
Conv3	3 × 3 × 3,Sp:1,Ss:1,Tp:1,Ts:1	3 × 3 × 3,Sp:1,Ss:1,Tp:1,Ts:1
Pool3	2 × 2 × 2,Ss:2,Ts:2	2 × 2 × 2,Ss:2,Ts:2
Conv4	3 × 3 × 4,Sp:1,Ss:1,Tp:0,Ts:1	3 × 3 × 4,Sp:1,Ss:1,Tp:0,Ts:1
pool4	2 × 2 × 1,Ss:2,Ts:1	2 × 2 × 1,Ss:2,Ts:1
Classify	–	–

Sp, Ss, Tp, and Ts denote spatial padding, spatial stride, temporal padding and temporal stride respectively.

Currently, many methods have been tried on the spontaneous micro-expressions database. In this paper, we compare DTSCNN with three state-of-the-art methods, i.e., STCLQP (Huang et al., 2016), MDMO (Liu et al., 2015), and FDM (Xu et al., 2017). The 3-fold cross-validation was used for all the methods evaluated on CASME I/II dataset.

However, from **Table 3** we can see that the size of training data in each fold of cross-validation is relatively small for DTSCNN. Another problem that may affect the classification task is imbalanced classification data (He and Garcia, 2009). In cross-validation, there exists some imbalance in our training set. To address the issue, Liu et al. (2015) applied polynomial SVM to evaluate the accuracy of the testing phase, Huang et al. (2016) and Xu et al. (2017) used F_1 score as an important index to measure the identification performance.

In our work, we utilized a sampling method as a data augmentation strategy to solve both imbalanced learning and small-sample problems in each fold of the cross-validation. The sampling method is illustrated by using a flow chart in **Figure 6** and conducted as following steps.

(1) A slice of images with 2 pixels in width or height is cut from every frame in the CASMEI/II. By cutting at different places, i.e., up, down, left, right, center and upper left, upper right, lower left and lower right part of the frame, nine new frames can be created.

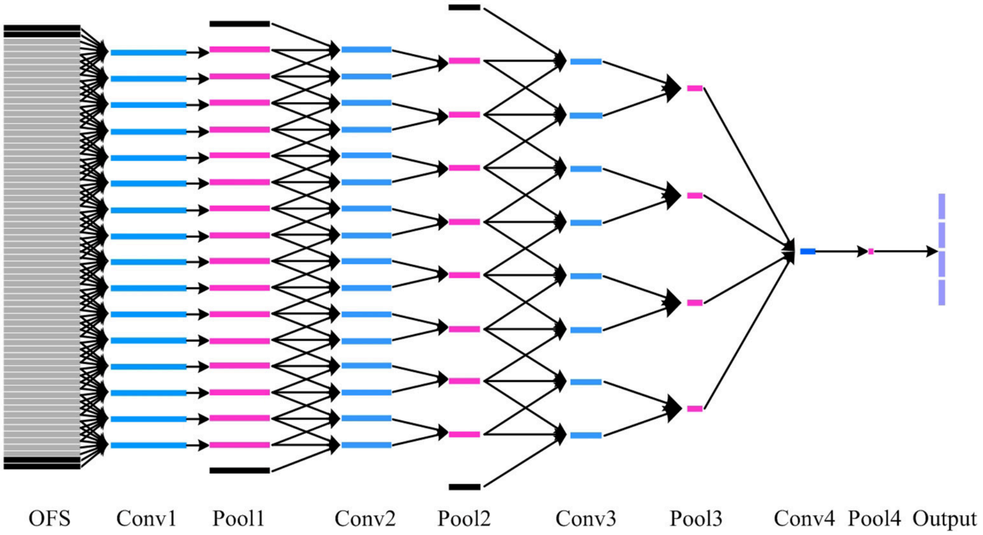

FIGURE 5 | The details of how 3D convolutional kernel in DTSCNN64 process a video clip. The blue, pink and purple box represents convolutional, pooling, and classifying layers, respectively. OFS denote optical flow sequence. The black box indicates the padding image in temporal domain. Each gray box represents an optical flow image. Besides, all parallel blue boxes in the same layer share parameters.

(2) The created nine frames are spatially normalized to 96 × 112 pixels.
(3) Repeat step 1 and 2 till all frames in the CASMEI/II are processed.

TABLE 3 | Specific emotions in each category and clip numbers in experiment database.

CASMEI/II	CASMEI	CASMEII
Negative (124)	Disgust (44), sadness (6), fear (2)	Disgust (63), sadness (7), fear (2)
Others (234)	Tense (69), repression (38), contempt (9)	Repression (27), Others (99)
Positive (41)	Happiness (9)	Happiness (32)
Surprise (45)	Surprise (20)	Surprise (25)

(4) For one class of emotion, the index of video clip is randomly selected, suppose the j-th video clip is selected.
(5) In the j-th video clip, replace every original frame with the frame randomly selected from its corresponding nine spatially normalized frames to create a new video clip.
(6) The created video clip are normalized to 65 frames if it is in CASMEI or 129 frames if it is in CASMEII by using linear interpolation method.
(7) Repeat step 4 to 6 until 500 video clips are created in this class of emotion.
(8) Repeat step 4–7 until every class of four classes has 500 video clips.

Finally, for each training set, we have 20,000 clips in total (4 × 500), the number of video clips in each test set remained unchanged.

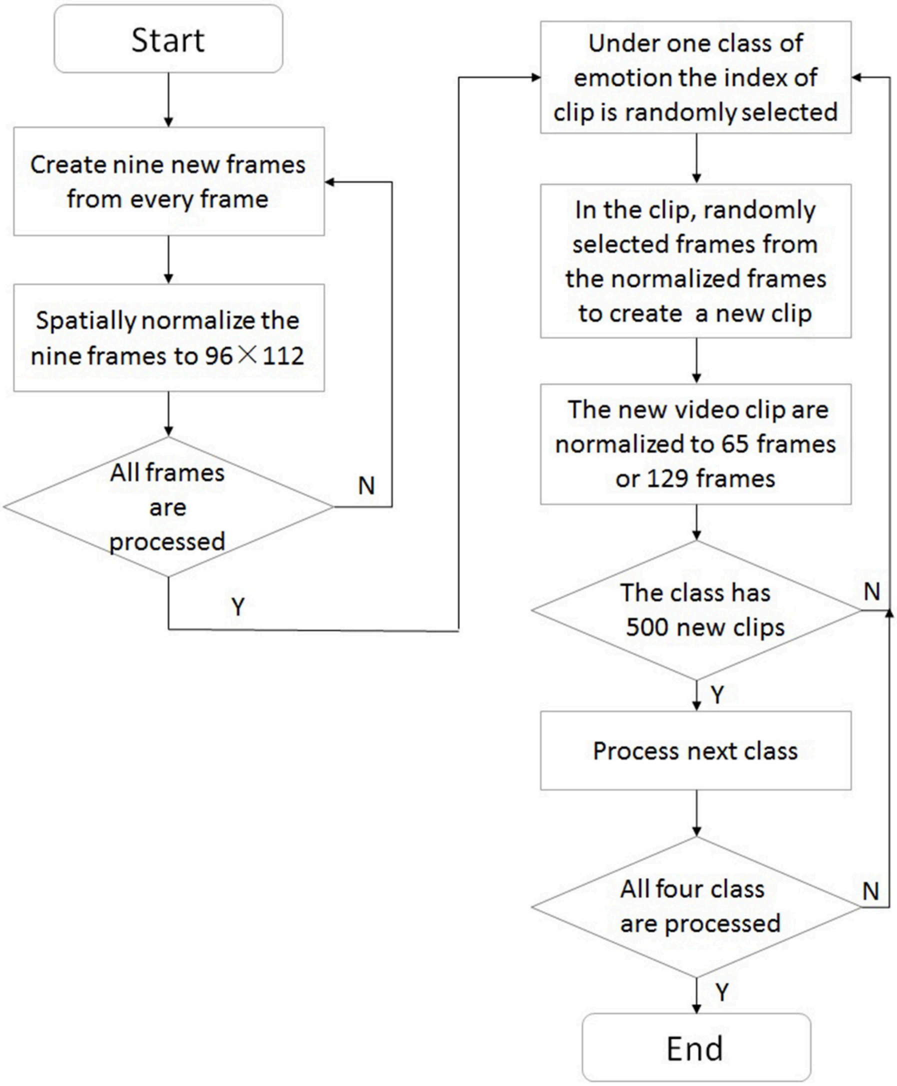
Start
Create nine new frames from every frame
Spatially normalize the nine frames to 96×112
All frames are processed
N
Y
Under one class of emotion the index of clip is randomly selected
In the clip, randomly selected frames from the normalized frames to create a new clip
The new video clip are normalized to 65 frames or 129 frames
The class has 500 new clips
N
Y
Process next class
All four class are processed
N
Y
End

FIGURE 6 | Flow chart of the sampling method.

Experimental Results on CASME I/II Database

Table 4 shows the micro-expression recognition accuracy of DTSCNN compared with the three methods. The experimental details of DTSCNN, STCLQP, MDMO, and FDM are as follows.

DTSCNN

The detail of the parameter setting of DTSCNN is given in **Table 2**. In the training phase, each stream employed batch gradient descent with a momentum of 0.9 (Wilson and Martinez, 2003), the dropout ratio of the FC layer was set to 0.5 and the minimum batch size was set to 10. The initial learning rate was set to 0.0001 and would get divided by 10 after every 10 epochs. Each stream of DTSCNN was trained separately, and the output feature of the Pool4 layer from each stream was fused for classification using linear SVM which would take decision-level fusion method to obtain final classification results. For each stream, if a sample x is classified into class C1 and class C2 with possibility of P1 and P2, respectively, then a decision-level fusion result C would be computed as follows

$$C = \begin{cases} C1 \ \ if.P_1 \geq P_2 \\ C2 \ \ if.P_1 < P_2 \end{cases} \tag{11}$$

STCLQP

Firstly, each video clip in CASME I/II database was given the same treatment as mentioned in section Pre-processing. Then, STCLQP feature extraction method that presented in Huang et al. (2016) was applied to each of the preprocessed sample. Finally, we used a SVM classifier with polynomial kernel (Schölkopf and Smola, 2002) to perform the classification. The parameters of SVM were refined using grid searching (Hsu et al., 2003).

MDMO

Prior to analyze the MDMO feature extraction methods, some processing steps which slightly different with the original paper (Liu et al., 2015) must be clarified. In particular, each raw video clip in CASMEI and CASME II database was given the same normalization treatment as mentioned in section Pre-processing. The following 36-ROIs detection of the first frame in each video clip was completed by using ASM model (Cootes et al., 1995) and FACS (Ekman and Friesen, 1977). Then, the optical flow sequence for each normalized samples were computed in the way mentioned in section Optical Flow Estimation. Finally, for each video clip, we obtained two optical-flow sequences in size of 64 × 480 × 640 × 3 and 128 × 480 × 640 × 3. In feature extraction and classification stage, the MDMO method was applied to each optical flow sequence to extract features from 36 ROIs, while a SVM with Gaussian kernel (Schölkopf and Smola, 2002) served as the classifier to evaluate the feature extraction performance.

TABLE 4 | The micro-expression recognition results (%) on CASMEI/II dataset with different methods, the fusion in bracket denotes the result is computed after done the decision-level fusion using Equation 10.

Methods	Fold1	Fold2	Fold3	Average
DTSCNN64 TIM64	65.45	65.45	65.45	65.45
DTSCNN128 TIM128	65.45	66.36	65.45	65.75
DTSCNN (fusion)	67.27	67.27	65.45	**66.67**
STCLQP TIM64	56.36	55.45	52.73	54.85
STCLQP TIM128	57.27	53.64	53.64	54.85
STCLQP (fusion)	58.18	56.36	54.55	**56.36**
MDMO TIM64	54.54	52.73	52.73	53.33
MDMO TIM128	54.54	54.54	53.63	54.24
MDMO (fusion)	57.27	55.45	53.63	**55.45**
FDM TIM64	53.64	53.64	54.55	53.94
FDM TIM128	53.64	53.64	55.45	54.24
FDM (fusion)	57.27	57.27	56.36	**56.97**

FDM

For each video clip in CASME I/II, firstly, we took the pretreatment method mentioned in section Pre-processing. Then, the optical flow computation and FDM feature extraction step (Xu et al., 2017) were conducted. Finally, we used a Linear SVM classifier to evaluate the accuracy of the feature.

As shown in **Table 4**, an average accuracy of 66.67% is achieved by DTSCNN, which is higher than every single stream network (DTSCNN64: 65.47%, DTSCNN128: 65.75%), and outperforms all the traditional feature extraction based method (STCLQP: 56.36%, MDMO: 52.12%, FDM: 56.97%). Particularly, the recognition accuracy of DTSCNN is almost 10 percent higher than STCLQP, MDMO, and FDM.

Figure 7 shows the average confusion matrices of the four methods. Apparently, the prediction of traditional methods would prefer the class with larger number of samples. For example, all three methods predict "Negative" as "Others" with chance of more than 55%, because "Other" class has larger training set. However, DTSCNN is robust to this imbalance-data effect, it can still predict "Negative" as "Negative" with chance of 50.54%. The good performance of the DTSCNN may be due to the sampling method employed by DTSCNN to address the problem of imbalanced data.

Among traditional methods, FDM is more robust to imbalanced-data effect. In predicting "Surprise," only FDM can predict it with higher chance (60.61%), STCLQP and MDMO predict it as "Others" with chance of 54.55 and 45.45%, respectively.

DTSCNN has almost the highest rate of correct prediction according the confusion matrices (except in predicting "Others"). Especially in recognizing "Negative," the DTSCNN has a correct prediction rate of 50.54%, which is more than 20% higher than those of STCLQP, MDMO, and FDM (26.88, 21.51, and 23.66%, respectively).

The main reason for the low recognition rate of "Positive" for all methods is due to very limited training samples (only 31). Nevertheless, the proposed DTSCNN method still archives the highest recognition accuracy rate of 13.33%.

To sum up DTSCNN can not only effectively learn features from imbalanced data, but also interpret the subtle movement in facial micro-expression clips internally and give an outstanding performance for quandary classification with negative, others, positive, and surprise.

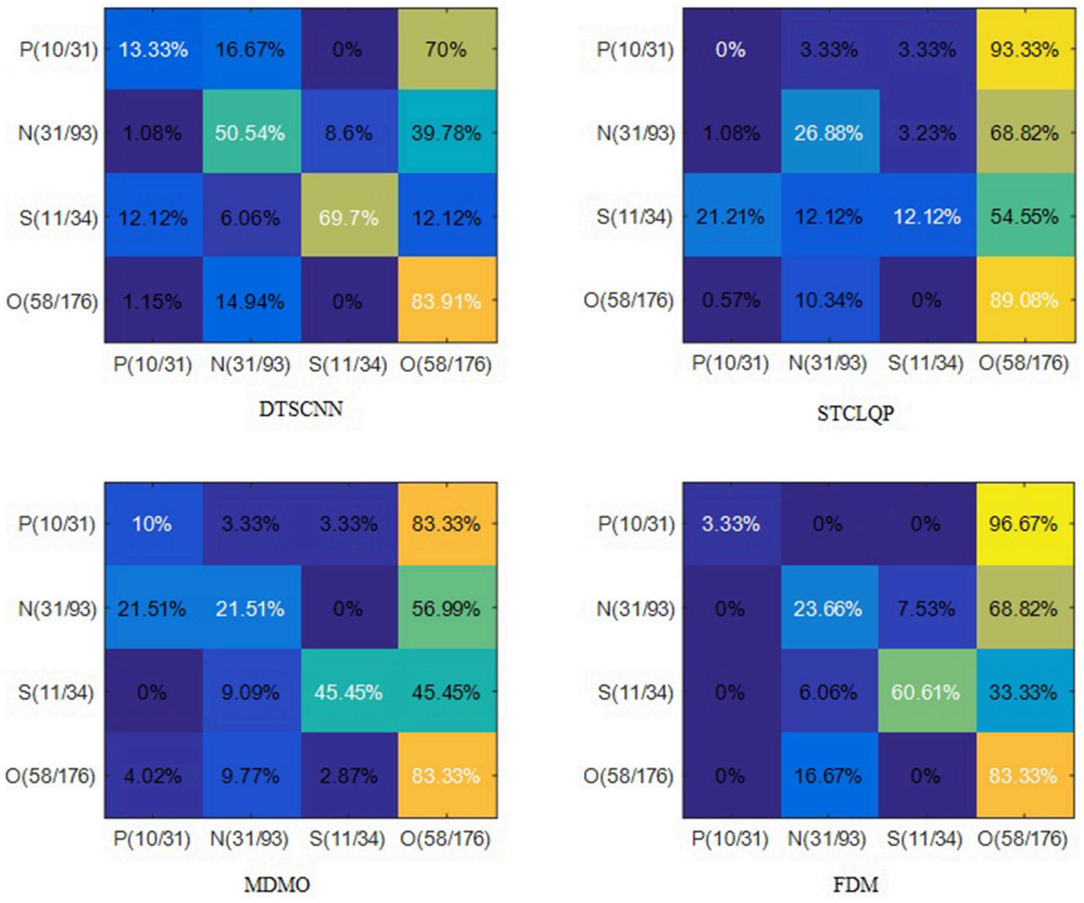

FIGURE 7 | Confusion matrices on CASMEI/II dataset. The N, O, P, and S denote the classes of Negative, Others, Positive, and Surprise, respectively. The number in packets indicates the number of samples in testing set vs. the training set.

CONCLUSION

In this paper, we proposed a DTSCNN architecture to recognize spontaneous micro-expression. The DTSCNN is a simplified design and end-to-end trainable two-stream network. Specifically, each convolution and pooling cell is a 3D structure that employs the Slow Fusion model mechanism to process micro-expression sequence internally, while the two-stream architecture is designed to take sequences normalized to 64 frames and 128 frames separately so that more discriminative features can be learned from data in different temporal length.

In pretreatment, unlike traditional methods that take complicated processing to obtain better recognition performance, we took much simpler method. The first step was to align clips to their first frame. The second was to calculate the optical flow estimation from the aligned and normalized samples.

In the experiment, we tested the DTSCNN on CASME I/II dataset. Unlike the traditional hand-crafted feature based method, which is labor-expensing and time-consuming, the DTSCNN can automatically learn features from simply pre-processed samples and complete the classification for recognition. Experimental results demonstrated that the proposed method can achieve highest recognition rate among STCLQP, MDMO, and FDM. This also suggests that our proposed DTSCNN could be a promising method for micro-expression applications.

AUTHOR CONTRIBUTIONS

MP and CW performed the data analysis, TC conceived the research, all authors wrote and read the article.

ACKNOWLEDGMENTS

This work was partially funded by the National Natural Science Foundation of China (Grant No. 61301297), the National Natural Science Foundation of China (NSFC), and the German Research Foundation (DFG) in project Cross Modal Learning, NSFC 6162113608/DFG TRR-169, and the Southwest University Doctoral Foundation (No. SWU115093).

REFERENCES

Abdel-Hamid, O., Mohamed, A. R., Jiang, H., and Penn, G. (2012). "Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Kyoto), 4277–4280.

Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active shape models-their training and application. *Comput. Vis. Image Underst*. 61, 38–59. doi: 10.1006/cviu.1995.1004

Ekman, P. (2003a). Deception, and facial expression. *Ann. N.Y. Acad. Sci.* 1000, 205–221. doi: 10.1196/annals.1280.010

Ekman, P. (2003b). *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotion Life*. New York, NY: Times Books, Henry Holt and Company.

Ekman, P. (2009). "Lie catching and microexpressions," in *The Philosophy of Deception*, ed C. W. Martin (Oxford: Oxford University Press), 118–133. doi: 10.1093/acprof:oso/9780195327939.003.0008

Ekman, P., and Friesen, W. V. (1977). *Facial Action Coding System*. Menlo Park, CA: Consulting Psychologists Press.

Ekman, P., and Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry* 32, 88–106. doi: 10.1080/00332747.1969.11023575

Endres, J., and Laidlaw, A. (2009). Micro-expression recognition training in medical students: a pilot study. *BMC Med. Educ.* 9:47. doi: 10.1186/1472-6920-9-47

Fasel, B., and Juergen, L. (2003). Automatic facial expression analysis: a survey. *Pattern Recognit*. 36, 259–275. doi: 10.1016/S0031-3203(02)00052-3

Frank, M. G., Herbasz, M., Sinuk, K., Keller, A., and Nolan, C. (2009). "I see how you feel: training laypeople and professionals to recognize fleeting emotions," in *The Annual Meeting of the International Communication Association* (New York, NY: Sheraton New York).

Gargi, U., Kasturi, R., and Strayer, S. H. (2000). Performance characterization of video-shot-change detection methods. Circuits and Systems for Video Technology, *IEEE Trans Circ. Syst. video Technol.* 10, 1–13. doi: 10.1109/76.825852

Goshtasby, A. (1988). Image registration by local approximation methods. *Image Vis. Comput.* 64, 255–261. doi: 10.1016/0262-8856(88)90016-9

Haggard, E. A., and Isaacs, K. S. (1966). "Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy," in *Methods of Research in Psychotherapy. The Century Psychology Series* (Boston, MA: Springer), 154–165. doi: 10.1007/978-1-4684-6045-2_14

He, H., and Garcia, E. A. (2009). Learning from imbalanced data. Knowledge and data engineering. *IEEE Trans. Knowl. Data Eng.* 21, 1263–1284. doi: 10.1109/TKDE.2008.239

Hinton, G. E., Osindero, S., and Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural Comput.* 18, 1527–1554. doi: 10.1162/neco.2006.18.7.1527

Hsu, C. W., Chang, C. C., and Lin, C. J. (2003). *A Practical Guide to Support Vector Classification*. Department of Computer Science, National Taiwan University, Taipei, Taiwan, 1–16.

Huang, X., Zhao, G., Hong, X., Zheng, W., and Pietikäinen, M. (2016). Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns. *Neurocomputing* 175, 564–578. doi: 10.1016/j.neucom.2015.10.096

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., and Fei-Fei, L. (2014). "Large-scale video classification with convolutional neural networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (Columbus, OH), 1725–1732.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* (Lake Tahoe, NV).

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi: 10.1109/5.726791

Li, X., Pfister, T., Huang, X., Zhao, G., and Pietikäinen, M. (2013). "A spontaneous micro-expression database: Inducement, collection and baseline," in *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (Shanghai).

Li, Y., Wang, S., Zhao, Y., and Ji, Q. (2013). Simultaneous facial feature tracking and facial expression recognition. *IEEE Trans. Image Process.* 22, 2559–2573. doi: 10.1109/TIP.2013.2253477

Liu, C. (2009). *Beyond Pixels: Exploring New Representations and Applications for Motion Analysis*. Dissertion, Massachusetts Institute of Technology.

Liu, Y. J., Zhang, J. K., Yan, W. J., Wang, S. J., Zhao, G., and Fu, X. (2015). A Main directional mean optical flow feature for spontaneous micro-expression recognition. *IEEE Trans. Affect. Comput.* 7, 299–310. doi: 10.1109/TAFFC.2015.2485205

Mikolov, T., Kombrink, S., Burget, L., Cernocky, J., and Khudanpur, S. (2011). "Extensions of recurrent neural network language model," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Prague, CZ), 5528–5531.

Nair, V., and Hinton, G. E. (2010). "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (Haifa), 807–814.

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

Polikovsky, S., Kameda, Y., and Ohta, Y. (2009). "Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor," *3rd International Conference on in Crime Detection and Prevention (ICDP 2009)* (London).

Porter, S., and Ten Brinke, L. (2008). Reading between the lies identifying concealed and falsified emotions in universal facial expressions. *Psychol. Sci.* 19, 508–514. doi: 10.1111/j.1467-9280.2008.02116.x

Russell, T. A., Chu, E., and Phillips, M. L. (2006). A pilot study to investigate the effectiveness of emotion recognition remediation in schizophrenia using the micro-expression training tool. *Br. J. Clin. Psychol.* 45, 579–583. doi: 10.1348/014466505X90866

Schölkopf, B., and Smola, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press.

Shreve, M., Godavarthy, S., Goldgof, D., and Sarkar, S. (2011). "Macro-and micro-expression spotting in long videos using spatio-temporal strain," in *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops* (Santa Barbara, CA), 51–56.

Simonyan, K., and Zisserman, A. (2014a).Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv*.1409.1556.

Simonyan, K., and Zisserman, A. (2014b). "Two-stream convolutional networks for action recognition in videos," in *Advances in Neural Information Processing Systems* (Montreal, QC), 568–576.

Sutskever, I., Vinyals, O., and Le, Q. V. (2014). "Sequence to sequence learning with neural networks," in *Advances in Neural Information Processing Systems* (Montreal, QC), 3104–3112.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Boston, MA: IEEE), 1–9.

Tran, D., Bourdev, L., Fergus, R., Torresani, L., and Paluri, M. (2015). "Learning spatiotemporal features with 3d convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision* (Santiago), 4489–4497.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., and Manzagol, P. A. (2010). Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *J. Machine Learn. Res.* 11, 3371–3408. Available online at: http://www.jmlr.org/papers/v11/vincent10a.html

Wang, Y., See, J., Phan, R. C. W., and Oh, Y. H. (2014). *Lbp with Six Intersection Points: Reducing Redundant Information in lbp-top for Micro-Expression Recognition. Computer Vision-ACCV 2014* (Singapore: Springer International Publishing), 525–537.

Weinberger, S. (2010). Airport security: intent to deceive? *Nature* 412–415. doi: 10.1038/465412a

Wilson, D. R., and Martinez, T. R. (2003). The general inefficiency of batch training for gradient descent learning. *Neural Netw*. 16, 1429–1451. doi: 10.1016/S0893-6080(03)00138-2

Wu, Q., Shen, X., and Fu, X. (2011). *The Machine Knows What you are Hiding: an Automatic Micro-Expression Recognition System. Affective Computing and Intelligent Interaction.* Memphis, TN; Berlin; Heidelberg: Springer.

Xu, F., Zhang, J., and Wang, J. (2017). Microexpression identification and categorization using a facial dynamics map. *IEEE Trans. Affect. Comput.* 8, 254–267. doi: 10.1109/TAFFC.2016.2518162

Yan, W. J., Li, X., Wang, S. J., Zhao, G., Liu, Y. J., Chen, Y. H., et al. (2014). CASME II: An improved spontaneous micro-expression database and the baseline evaluation. *PLoS ONE* 9:e86041. doi: 10.1371/journal.pone.0086041

Yan, W. J., Wu, Q., Liang, J., Chen, Y. H., and Fu, X. (2013a). How fast are the leaked facial expressions: the duration of micro-expressions. *J. Nonverbal Behav.* 37, 217–230. doi: 10.1007/s10919-013-0159-8

Yan, W. J., Wu, Q., Liu, Y. J., Wang, S. J., and Fu, X. (2013b). "CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces," *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (Shanghai).

Zhang, L., and Tjondronegoro, D. (2011). Facial expression recognition using facial movement features. *IEEE Trans. Affect. Comput.* 2, 219–229. doi: 10.1109/T-AFFC.2011.13

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Psychology

REVIEW
published: 10 July 2018
doi: 10.3389/fpsyg.2018.01128

A Survey of Automatic Facial Micro-Expression Analysis: Databases, Methods, and Challenges

Yee-Hui Oh[1†], *John See*[2*†], *Anh Cat Le Ngo*[3], *Raphael C. -W. Phan*[1,4] *and Vishnu M. Baskaran*[5]

[1] Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia, [2] Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia, [3] School of Psychology, University of Nottingham, Nottingham, United Kingdom, [4] Research Institute for Digital Security, Multimedia University, Cyberjaya, Malaysia, [5] School of Information Technology, Monash University Malaysia, Bandar Sunway, Malaysia

Over the last few years, automatic facial micro-expression analysis has garnered increasing attention from experts across different disciplines because of its potential applications in various fields such as clinical diagnosis, forensic investigation and security systems. Advances in computer algorithms and video acquisition technology have rendered machine analysis of facial micro-expressions possible today, in contrast to decades ago when it was primarily the domain of psychiatrists where analysis was largely manual. Indeed, although the study of facial micro-expressions is a well-established field in psychology, it is still relatively new from the computational perspective with many interesting problems. In this survey, we present a comprehensive review of state-of-the-art databases and methods for micro-expressions spotting and recognition. Individual stages involved in the automation of these tasks are also described and reviewed at length. In addition, we also deliberate on the challenges and future directions in this growing field of automatic facial micro-expression analysis.

Keywords: facial micro-expressions, subtle emotions, survey, spotting, recognition, databases, spontaneous, expressions

OPEN ACCESS

Edited by:
Wenfeng Chen,
Renmin University of China, China

Reviewed by:
Xunbing Shen,
Jiangxi University of Traditional Chinese Medicine, China
Xiaobai Li,
University of Oulu, Finland
Tong Chen,
Southwest University, China

****Correspondence:***
John See
johnsee@mmu.edu.my

† These authors have contributed equally to this work.

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *02 December 2017*
Accepted: *12 June 2018*
Published: *10 July 2018*

Citation:
Oh Y-H, See J, Le Ngo AC, Phan RC-W and Baskaran VM (2018) A Survey of Automatic Facial Micro-Expression Analysis: Databases, Methods, and Challenges. Front. Psychol. 9:1128. doi: 10.3389/fpsyg.2018.01128

1. INTRODUCTION

In 1969, Ekman and Friesen (1969) spotted a quick full-face emotional expression in a filmed interview which revealed a strong negative feeling a psychiatric patient was trying to hide from her psychiatrist in order to convince that she was no longer suicidal. When the interview video was played in slow motion, it was found that the patient was showing a very brief sad face that lasted only for two frames (1/12s) followed by a longer-duration false smile. This type of facial expressions is called micro-expressions (MEs) and they were actually first discovered by Haggard and Isaacs (1966) 3 years before the event happened. In their study, Haggard and Isaacs discovered these micromomentary expressions while scanning motion picture films of psychotherapy hours, searching for indications of non-verbal communication between patient and therapist.

MEs are very brief, subtle, and involuntary facial expressions which normally occur when a person either deliberately or unconsciously conceals his or her genuine emotions (Ekman and Friesen, 1969; Ekman, 2009b). Compared to ordinary facial expressions or macro-expressions, MEs usually last for a very short duration which is between 1/25 and 1/5 of a second (Ekman, 2009b). Recent research by Yan et al. (2013a) suggest that the generally accepted upper limit duration of a micro-expression is within 0.5s. Besides short duration, MEs also have other significant characteristics such as low intensity and fragmental facial action units where only part of the action units of full-stretched facial expressions are presented (Porter and Ten Brinke, 2008; Yan et al., 2013a). Due to these three characteristics of the MEs, it is difficult for human beings to perceive micro-expressions with the naked eye.

In spite of these challenges, new psychological studies of MEs and computational methods to spot and recognize MEs have been gaining more attention lately because of its potential applications in many fields, i.e., clinical diagnosis, business negotiation, forensic investigation, and security systems (Ekman, 2009a; Frank et al., 2009a; Weinberger, 2010). One of the very first efforts to improve the human ability at recognizing MEs was conducted by Ekman where he developed the Micro-Expression Training Tool (METT) to train people to recognize seven categories of MEs (Ekman, 2002). However, it was found in Frank et al. (2009b) that the performance of detecting MEs by undergraduate students only reached at most 40% with the help of METT while unaided U.S. coast guards performed not more than 50% at best. Thus, an automatic ME recognition system is in great need in order to help detect MEs such as those exhibited in lies and dangerous behaviors, especially with the modern advancements in computational power and parallel multi-core functionalities. These have enabled researchers to perform video processing operations that used to be infeasible decades ago, increasing the capability of computer-based understanding of videos in solving different real-life vision problems. Correspondingly, in recent years researchers have moved beyond psychology to using computer vision and video processing techniques to automate the task of recognizing MEs.

Although normal facial expression recognition is now considered a well-established and popular research topic with many good algorithms developed (Zeng et al., 2009; Bettadapura, 2012; Sariyanidi et al., 2015) with accuracies exceeding 90%, in contrast the automatic recognition of MEs from videos is still a relatively new research field with many challenges. One of the challenges faced by this field is spotting the ME of a person accurately from a video sequence. As a ME is subtle and short, spotting of MEs is not an easy task. Furthermore, spotting of MEs becomes harder if the video clip consists of spontaneous facial expressions and unrelated facial movements, i.e., eye-blinking, opening and closing of mouth, etc. On the other hand, other challenges of ME recognition include inadequate features for recognizing MEs due to its low change in intensity and lack of complete, spontaneous and dynamic ME databases.

In the past few years, there have been some noteworthy advances in the field of automatic ME spotting and recognition. However, there is currently no comprehensive review to chart the emergence of this field and summarize the development of techniques introduced to solve these tasks. In this survey paper, we first discuss the existing ME corpora. In our perspective, automatic ME analysis involves two major tasks, namely, ME spotting and ME recognition. ME spotting focuses on finding the occurrence of MEs in a video sequence while ME recognition involves assigning an emotion class label to an ME sequence. For both tasks, we look into the range of methods that have been proposed and applied to various stages of these tasks. Lastly, we discuss the challenges in ME recognition and suggest some potential future directions.

2. MICRO-EXPRESSION DATABASES

The prerequisite of developing any automatic ME recognition system is having enough labeled affective data. As ME research in computer vision has only gained attention in the past few years, the number of publicly available spontaneous ME databases is still relatively low. **Table 1** gives the summary of all available ME databases to date, including both posed and spontaneous ME databases. The key difference between posed and spontaneous MEs is in the relevance between expressed facial movement and underlying emotional state. For posed MEs, facial expressions are deliberately shown and irrelevant to the present emotion of senders, therefore not really helpful for the recognition of real subtle emotions. Meanwhile, spontaneous MEs are the unmodulated facial expressions that are congruent with an underlying emotional state (Hess and Kleck, 1990). Due to the nature of the posed and spontaneous MEs, the techniques for inducing facial expressions (for purpose of constructing a database) are contrasting. For the case of posed MEs, subjects are usually asked to relive an emotional experience (or even watching example videos containing MEs prior to the recording session) and perform the expression as well as possible. However, eliciting spontaneous MEs is more challenging as the subjects have to be involved emotionally. Usually, emotionally evocative video episodes are used to induce the genuine emotional state of subjects, and the subjects have to attempt to suppress their true emotions or risk getting penalized.

According to Ekman and Friesen (1969) and Ekman (2009a), MEs are involuntary which could not be created intentionally. Thus, posed MEs usually do not exhibit the characteristics (i.e., the appearance and timing) of spontaneously occurring MEs (Porter and Ten Brinke, 2008; Yan et al., 2013a). The early USD-HD (Shreve et al., 2011) and Polikovsky's (Polikovsky et al., 2009) databases consist of posed MEs rather than spontaneous ones; hence they do not present likely scenarios encountered in real life. In addition, the occurrence duration of their micro-expressions (i.e., 2/3 s) exceeds the generally acceptable duration of MEs (i.e., 1/2 s). To have a more ecological validity, research interest then shifted to spontaneous ME databases. Several groups have developed a few spontaneous MEs databases to aid researchers in the development of automatic ME spotting and recognition algorithms. To elicit MEs spontaneously, participants are induced by watching emotional video clips to experience a high arousal, aided by an incentive (or penalty) to motivate the disguise of emotions. However, due to the challenging process of eliciting these spontaneous MEs, the number of samples collected for these ME databases is still limited.

Table 1 summarizes the known ME databases in the literature, which were elicited through both posed and spontaneous means. The YorkDDT (Warren et al., 2009) is the smallest and oldest database, with spontaneous MEs that also include other irrelevant head and face movements. The Silesian Deception database

TABLE 1 | Micro-expression databases.

Databases	Subset	Subjects	Samples	Frames per sec	Type*	FACS coded	Emotion classes	Expression	Frame annotations
USF-HD		–	100	30	P	No	6	Macro/micro	–
Polikovsky's		10	42	200	P	No	6	Micro	–
YorkDDT		9	18	25	S	No	2	Micro	–
Silesian deception†		101	101	100	S	No	–	Macro/micro	Eye closures, gaze aversion, micro-tensions
SMIC-sub		6	77	100	S	No	3	Micro	–
SMIC	HS	16	164	100	S	No	3	Micro	–
	VIS	8	71	25	S	No	3		
	NIR	8	71	25	S	No	3		
	E-HS	16	157	100	S	No	3	Micro	Onset,offset
	E-VIS	8	71	25	S	No	3		
	E-NIR	8	71	25	S	No	3		
CASME		19	195	60	S	Yes	7	Micro	Onset,offset,apex
CASME II		26	247	200	S	Yes	5	Micro	Onset,offset,apex
CAS(ME)^2	Part A	22	87	30	S	Yes	4	Macro/Micro	Onset,offset,apex
	Part B	22	57	30	S	Yes	4		
SAMM		32	159	200	S	Yes	7‡	Macro/micro	Onset,offset,apex
MEVIEW		16	31	25	S	Yes	5§	macro/micro	onset,offset

*P/S, Posed/Spontaneous.

†Not all samples contain micro-expressions and only a total of 183 occurrences of "micro-tensions" were annotated. No emotion classes were available.

‡Seven objective classes are also provided (Davison et al., 2017).

§Set of emotions are atypical (contempt, surprise, fear, anger, happy), likely in the context of environment. Some sample clips involve person speaking, or only have AUs marked with no emotions observed.

(Radlak et al., 2015) was created for the purpose of recognizing deception through facial cues. This database is annotated with eye closures, gaze aversion, and micro-expression, or "micro-tensions," a phrase used by the authors to indicate the occurrence of rapid facial muscle contraction as opposed to having an emotion category. This dataset is not commonly used in spotting and recognition literature as it does not involve expressions *per se*; its inception primarily for the purpose of automatic deception recognition.

The SMIC-sub (Pfister et al., 2011) database presents a better set of spontaneous ME samples in terms of frame rate and database size. Nevertheless, it was further extended to the SMIC database (Li et al., 2013) with the inclusion of more ME samples and multiple recordings using different cameras types: high speed (HS), normal visual (VIS), and near-infrared (NIS). However, the SMIC-sub and SMIC databases do not provide Action Unit (AU) (i.e., facial components that are defined by FACS to taxonomize facial expressions) labels and the emotion classes were only based on participants' self-reports. Sample frames from SMIC are shown in **Figure 1**.

The CASME dataset (Yan et al., 2013b) provides a more comprehensive spontaneous ME database with a larger amount of MEs as compared to SMIC. However, some videos are extremely short, i.e., <0.2 s, hence poses some difficulty for ME spotting. Besides, CASME samples were captured only at 60 *fps*. An improved version of it, known as CASME II was established to address these inadequacies. The CASME II database (Yan et al., 2014a) is the largest and most widely used database to date (247 videos, sample frames in **Figure 2**) with samples recorded using high frame-rate cameras (200 *fps*).

To facilitate the development of algorithms for ME spotting, extended versions of SMIC (SMIC-E-HS, SMIC-E-VIS, SMIC-E-NIR), CAS(ME)^2 (Qu et al., 2017), and SAMM (Davison et al., 2016a) databases were developed. In SMIC-E databases, long video clips that contain some additional non-micro frames before and after the labeled micro frames were included as well. The CAS(ME)^2 database (with samples given in **Figure 3**) is separated into two parts: Part A contains both spontaneous macro-expressions and MEs in long videos; and Part B includes cropped expression samples with frame from onset to offset. However, CAS(ME)^2 is recorded using a low frame-rate (25 *fps*) camera due to the need to capture both macro- and micro-expressions.

In the SAMM database (with samples shown in **Figure 4**), all micro-movements are treated objectively, without inferring the emotional context after each experimental stimulus. Emotion classes are then labeled by trained experts later. In addition, about 200 neutral frames are included before and after the occurrence of the micro-movement, which makes spotting feasible. The SAMM is arguably the most culturally diverse database among all of them. In short, the SMIC, CASME II, CAS(ME)^2, and SAMM are considered the state-of-the-art databases for ME spotting and recognition that should be widely adopted for research.

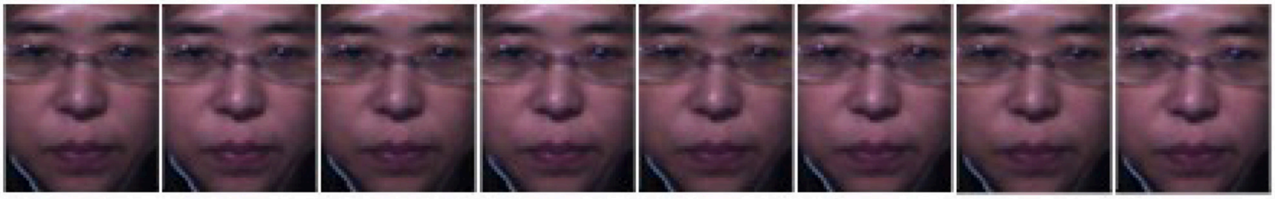

FIGURE 1 | Sample frames from a "Surprise" sequence (Subject 1) in SMIC. Images reproduced from the database with permission from Li et al. (2013).

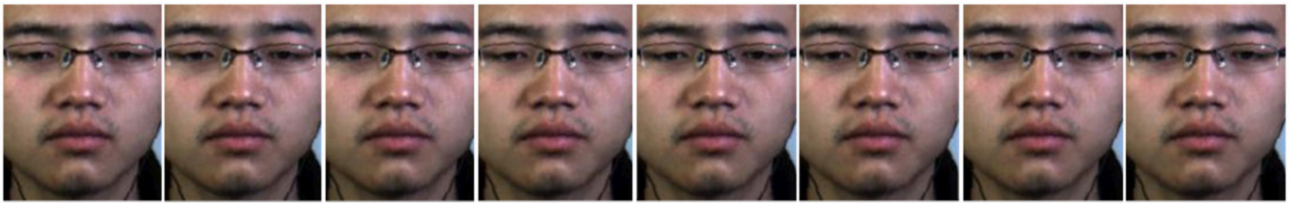

FIGURE 2 | Sample frames from a "Happiness" sequence (Subject 6) in CASME II. Images reproduced from the database with permission from Yan et al. (2014a).

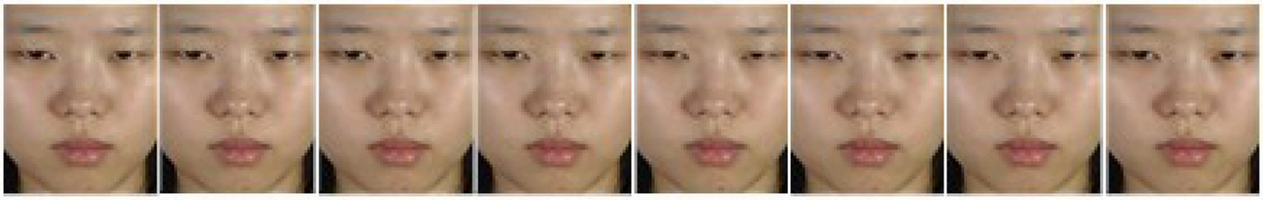

FIGURE 3 | Sample frames from a "Disgust" sequence (Subject 15) in CAS(ME)^2. Images reproduced from the database (©Xiaolan Fu) with permission from Qu et al. (2017).

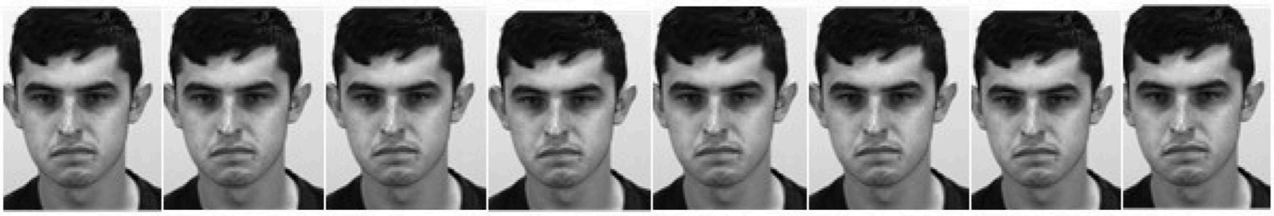

FIGURE 4 | Sample frames from a sequence (Subject 6) in SAMM that contains micro-movements. Images reproduced from the database with permission from Davison et al. (2016a).

The need for data acquired from more unconstrained "in-the-wild" situations have compelled further efforts to provide more naturalistic high-stake scenarios. The MEVIEW dataset (Husak et al., 2017) was constructed by collecting mostly poker game videos downloaded from YouTube with a close-up of the player's face (samples frames in **Figure 5**). Poker games are highly competitive with players often try to conceal or fake their true emotions, which facilitates likely occurrences of MEs. With the camera view switching often, the entire shot with a single face in video (averaging 3s in duration) was taken. An METT-trained annotator labeled the onset and offset frames of the ME with FACS coding and emotion types. A total of 31 videos with 16 individuals were collected.

3. SPOTTING OF FACIAL MICRO-EXPRESSIONS

Automatic ME analysis involves two tasks: ME spotting and ME recognition. Facial ME spotting refers to the problem of automatically detecting the temporal interval of a micro-movement in a sequence of video frames; and ME recognition is the classification task to identify the ME involved in the video samples. In a complete facial ME recognition system, accurately and precisely identifying frames containing facial micro-movements (which contribute to facial MEs) in a video is a prerequisite for high-level facial analysis (i.e., facial ME recognition). Thus, the automatic facial expression spotting frameworks are developed to automatically search the temporal dynamics of MEs in streaming videos. Temporal dynamics refer to the motions of facial MEs that involve onset(start), apex(peak), offset(end), and neutral phases. **Figure 6** shows a sample sequence depicting these phases. According to the work by Valstar and Pantic (2012), the onset phase is the moment where muscles are contracting and appearance of facial changes grows stronger; the apex phase is the moment where the expression peaks (the most obvious); and the offset phase is the instance where the muscles are relaxing and the face returns to its neutral appearance (little or no activation of facial muscles). Typically a facial motion shifts through the sequence of neutral-onset-apex-offset-neutral, but other combinations such as multiple apices are also possible.

In general, a facial ME spotting framework consists of a few stages: the pre-processing, feature description, and lastly the detection of the facial micro-expressions. The details of each of the stages will be further discussed in the following sections.

3.1. Pre-processing

In facial ME spotting, the general pre-processing steps include facial landmark detection, facial landmark tracking, face registration, face masking, and face region retrieval. **Table 2** shows a summary of existing pre-processing techniques that are applied in facial ME spotting.

3.1.1. Facial Landmark Detection and Tracking

Facial landmark detection is the first most important step in the spotting framework to locate the facial points on the facial images. In the field of MEs, two ways of locating the facial points are applied: the manual method and automatic facial landmark detection method. In an early work on facial micro-movement spotting (Polikovsky et al., 2009), facial landmarks are manually selected only at the first frame, and fixed in the consecutive frames as they assumed that the examined frontal faces are located relatively in the same location. In their later work (Polikovsky and Kameda, 2013), a tracking algorithm is applied to track the facial points that had been manually detected at the first frame throughout the whole sequence. To prevent the hassle of manually detecting the facial points, majority of the recent works (Davison et al., 2015, 2016a,b; Liong et al., 2015, 2016b,c; Wang et al., 2016a; Xia et al., 2016) opt to apply automatic facial landmark detection. Instead of running the detection for the whole sequence of facial images, the facial points are only detected at the first frame and fixed in the consecutive frames with the assumption that these points will only change minimally due to the subtleness of MEs.

To the best of our knowledge, the facial landmark detection techniques that are commonly employed for facial ME spotting are promoted Active Shape Model (ASM) (Milborrow and Nicolls, 2014), Discriminative Response Maps Fitting (DRMF) (Asthana et al., 2013), Subspace Constrained Mean-Shifts (SCMS) (Saragih et al., 2009), Face++ automatic facial point detector (Megvii, 2013), and Constraint Local Model (CLM) (Cristinacce and Cootes, 2006). In fact, the promoted ASM, DRMF, and CLM are the notable examples of part based

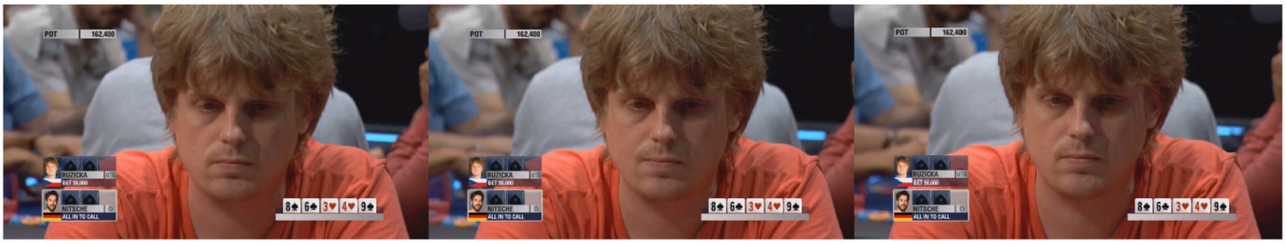

FIGURE 5 | Sample frames from a "Contempt" sequence in MEVIEW that contains micro-movements marked with AU L12. Images reproduced from the database (Husak et al., 2017) under Fair Use.

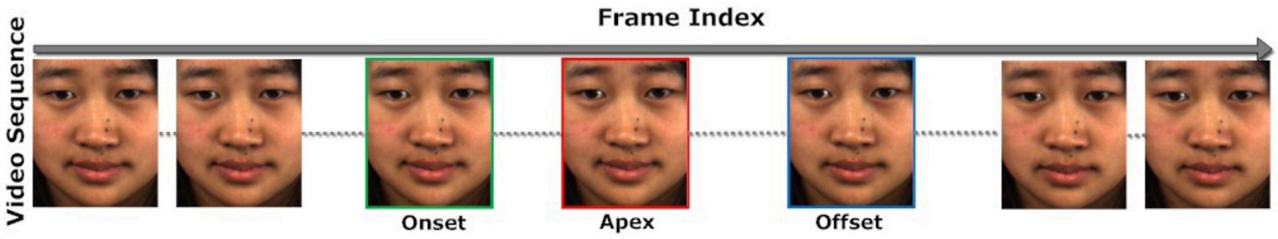

FIGURE 6 | A video sequence depicting the order in which onset, apex and offset frames occur. Sample frames are from a "Happiness" sequence (Subject 2) in CASME II. Images reproduced from the database with permission from Yan et al. (2014a).

TABLE 2 | A survey of pre-processing techniques applied in facial micro-expression spotting.

Work	Landmark detection	Landmark tracking	Face registration	Masking	Face regions
Polikovsky et al., 2009	Manual	–	–	–	12 ROIs
Shreve et al., 2009	–	–	–	–	3 ROIs
Wu et al., 2011	–	–	–	–	Whole face
Shreve et al., 2011	–	–	Face alignment	Eyes, nose and mouth	8 ROIs
Polikovsky and Kameda, 2013	Manual	APF	–	–	12 ROIs
Shreve et al., 2014	SCMS	–	–	Eyes and mouth	4 Parts
Moilanen et al., 2014	Manual	KLT	Face alignment	–	6 × 6 blocks
Davison et al., 2015	Face++	–	Affine transform	–	5 × 5 blocks
Patel et al., 2015	DRMF	OF	–	–	49 ROIs
Liong et al., 2015	DRMF	–		–	3 ROIs
Wang et al., 2016a	DRMF	–	Non-reflective similarity transformation	–	6 × 6 blocks
Liong et al., 2016c	DRMF	–	–	Eyes	3 ROIs
Xia et al., 2016	ASM	–	Procrutes analysis	–	Whole face
Liong et al., 2016b	DRMF	–	–	–	3 ROIs
Davison et al., 2016a	Face++	–	Affine transform	–	4 × 4, 5 × 5 blocks
Davison et al., 2016b	Face++	–	2D-DFT and Piecewise affine warping	Binary masking	26 ROIs
Yan and Chen, 2017	CLM	–	–	–	16 ROIs
Li et al., 2017	Manual	KLT	–	–	6 × 6 blocks
Ma et al., 2017	CLNF from OpenFace	KLT	–	–	5 ROIs
Qu et al., 2017	ASM	–	LWM	–	Various block sizes
Duque et al., 2018	AAM	KLT	–	–	5 ROIs

facial deformable models. Facial deformable models can be roughly separated into two main categories: holistic (generative) models and part based (discriminative) models. The former applies holistic texture-based facial representation for the generic face fitting scenario; and the latter uses the local image patches around the landmark points for the face fitting scenario. Although the holistic-based approaches are able to achieve impressive registration quality, these representations unfaithfully locate facial landmarks in unseen images, when target individuals are not included in the training set. As a result, part based models which circumvent several drawbacks of holistic-based methods, are more frequently employed in locating facial landmarks in recent years (Asthana et al., 2013). The promoted ASM, DRMF, and CLM are from part based deformable models, however their mechanisms are different. The ASM applies shape constraints and searches locally for each feature point's best location; whereas DRMF learns the variation in appearance on a set of template regions surrounding individual features and updates the shape model accordingly; as for CLM, it learns a model of shape and texture variation from a template (similar to active appearance models), but the texture is sampled in patches around individual feature points. In short, the DRMF is computationally lighter than its counterparts.

Part based approaches mainly rely on optimization strategies to approximate the responses map through simple parametric representations. However, some ambiguities still result due to the landmark's small support region and imperfect detectors. In order to address these ambiguities, SCMS which employs Kernel Density Estimator (KDE) to form a non-parametric representation of response maps was proposed. To maximize over the KDE, the mean-shift algorithm was applied. Despite the progress in automatic facial landmark detection, these approaches are still not considerably robust toward "in-the-wild" scenarios, where large out-of-plane tilting and occlusion might exist. The Face++ automatic facial point detector was developed by Megvii (2013) to address such challenges. It employs a coarse-to-fine pipeline with neural network and sequential regression, and it claims to be robust against influences such as partial occlusions and improper head pose up to 90° tilt angle. The efficacy of the method (Zhang et al., 2014) has been tested on the 300-W dataset (Sagonas et al., 2013) (which focuses on facial landmark detection in real-world facial images captured in-the-wild), yielding the highest accuracy among the several recent state-of-the-arts including DRMF.

In ME spotting research, very few works applied tracking to the landmark points. This could be due to the sufficiency of landmark detection algorithms used (since MEs movements are very minute) or that general assumptions have been made to fix the location of the detected landmarks points. The two tracking algorithms that were reportedly used in a few facial ME spotting works (Polikovsky and Kameda, 2013; Moilanen et al., 2014; Li et al., 2017) are Auxiliary Particle Filtering (APF) (Pitt and Shephard, 1999) and Kanade-Lucas-Tomasi (KTL) algorithm (Tomasi and Kanade, 1991).

3.1.2. Face Registration

Image registration is the process of aligning two images—the reference and sensed images, geometrically. In the facial ME spotting pipeline, registration techniques are applied onto the faces to remove large head translations and rotations that might affect the spotting task. Generally, registration techniques can be separated into two major categories: area-based and feature-based approaches. In each of the approaches, either global mapping functions or local mapping functions are applied to transform the the sensed image to be as close as the reference image.

For area-based (a.k.a. template matching or correlation-like) methods, windows of predefined size or even entire images are utilized for the correspondence estimation during the registration. This approach bypasses the need for landmark points, albeit some restriction to only shift and small rotations between the images (Zitova and Flusser, 2003). In the work by Davison et al. (2016b), a 2D-Discrete Fourier Transform (2D-DFT) was used to achieve face registration. This method calculates the cross-correlation of the sensed and reference images before finding the peak, which in turn is used to find the translation between the sensed and reference images. Then, the process of warping to a new image is performed by piece-wise affine (PWA) warping.

For feature-based approach to face registration, salient structures which include region features, line features and point features are exploited to find the pairwise correspondence between the sensed and reference images. Thus, feature-based approach are usually applied when the local structures are more significant than the information carried by the image intensities. In some ME works (Shreve et al., 2011; Moilanen et al., 2014; Li et al., 2017), the centroid of the two detected eyes are selected as the distinctive point (also called control points) and exploited for face registration by using affine transform or non-reflective similarity transform. The consequence of such simplicity entails their inability to handle deformations locally. A number of works (Li et al., 2017; Xu et al., 2017) employed Local Weighted Mean (LWM) (Goshtasby, 1988) which seeks to find a 2-D transformation matrix using 68 facial landmark points of a model face (typically from the first frame). In another work by Xia et al. (2016), Procrustes analysis is applied to align the detected landmark points in frames. It determines a linear transformation (such as translation, reflection, orthogonal rotation, and scaling) of the points in sensed images to best conform them to points in the reference image. Procrustes analysis has several advantages: low complexity for easy implementation and it is practical for similar object alignment (Ross, 2004). However, it requires a one-to-one landmark correspondence and the convergence of means is not guaranteed.

Instead of using mapping functions to map the sensed images to the reference images, a few studies (Shreve et al., 2011; Moilanen et al., 2014; Li et al., 2017) correct the mis-alignment by rotating the faces according to the angle between the pair of lines that join the centroids of the two detected eyes to the horizontal line. In this mechanism, errors can creep in if the face contours of the sensed and reference face images are not consistent with one another, or that the subject's face is not entirely symmetrical to begin with.

Due to the diversity of face images with various types of degradations to be registered, it is challenging to fix a standard method that is applicable to all conditions. Thus, the choice of registration method should correspond to the assumed geometric deformation of the sensed face image.

3.1.3. Masking

In the facial ME spotting task, a masking step can be applied onto the face images to remove noise caused by undesired facial motions that might affect the performance of the spotting task. In the work by Shreve et al. (2011), a static mask ("T"-shaped) was applied on the face images to remove the middle part of the face that includes the eyes, nose, and mouth regions. Eye regions were removed to avoid the noise caused by eye cascades and blinking (which is not considered a facial micro-expression); the nose region is masked as it is typically rigid, which might not reveal much significant information even with it; and mouth region is excluded since opening and closing of the mouth introduces undesired large motion. It is arguable if too much meaningful information may have been removed from the face area in the masking steps introduced in Shreve et al. (2011) and Shreve et al. (2014), as the two most expressive facial parts (in the context of MEs) are actually located near the corner of the eyebrow and mouth areas. Hence, some control is required to prevent excluding too much meaningful information. Typically, specific landmark points around these two areas are used as reference or boundary points in the masking process.

In the work by Liong et al. (2016c), the eye regions are masked to reduce false spotting of the apex frame from long videos. They observed that eye blinking motion is significantly more intense than that of micro-expression motion, thus masking is necessary. To overcome potential inaccurate landmark detection, a 15-pixel margin was added to extend the masked region. Meanwhile, Davison et al. (2016b) applied a binary mask to obtain 26 FACS-based facial regions that include the eyebrows, forehead, cheeks, corners around eyes, mouth, regions around mouth, and etc. The regions are useful for the spotting task as each of these regions contain a single or a group of AUs, which will be triggered when the ME occurs. It is also worth mentioning that a majority of works in the literature still do not include a masking pre-processing step.

3.1.4. Face Region Retrieval

From psychological findings on concealed emotions (Porter and Ten Brinke, 2008), it was revealed that facial micro-expression analysis should be done on the upper and lower halves of the face separately instead of considering the entire face. This finding substantiated an earlier work (Rothwell et al., 2006), whereby ME recognition was also performed on the segmented upper and lower parts of the face. Duan et al. (2016) later showed that the eye region is much more salient than the whole face or mouth region for recognizing micro-expressions, in particular happy and disgust expressions. Prior knowledge from these works encourage splitting of the face into important regions for automatic facial micro-expression spotting.

In the pioneering work of spotting facial MEs (Shreve et al., 2009), the face was segmented into three regions: the upper part (which includes the forehead), middle part (which includes the nose and cheeks), and the lower part (which include the mouth); and each was analyzed as individual temporal sequences. In their later work (Shreve et al., 2011), the face image is further segmented into eight regions: forehead, left and right of the eye, left and right of cheek, left and right of mouth and chin. Each of the segments is analyzed separately in sequence. With the more localized segments, tiny changes in certain temporal segments could be observed. However, unrelated edged features such as hair, neck, and edge of the face that are present in the localized segments might induce noise and thus affect the extracted features. Instead of splitting the face images into few segments, Shreve et al. (2014) suggested to separate the face images into four quadrants, and each of the quadrant is analyzed individually in the temporal domain. The reason behind this is because of the constraint on locality as facial micro-expressions are restricted to appear in at most two bordering regions (i.e., first and second quadrant, second and third quadrant, third and forth quadrant, and the first and fourth quadrant) of the face (Shreve et al., 2014).

Another popular facial segmentation method is splitting the face into a specific number ($m \times n$) of blocks (Moilanen et al., 2014; Davison et al., 2015, 2016a; Wang et al., 2016a; Li et al., 2017). In the blocking representation, the motion changes in each block could by observed and analysis independently. However, with the increasing in the number of blocks (i.e., $m \times n$), the computation load increases accordingly. Besides, features such as hairs and edges of face that appear in the blocks will affect the final feature vectors as these elements are not related to the facial motions.

A unique approach to facial segmentation for ME spotting is to split the face by Delaunay triangulation (Davison et al., 2016b). It gives more freedom to the shape that defines the regions of the face. Unfortunately, areas of the face that are not useful for ME analysis such as the cheek area may still be captured within the triangular regions. To address this problem, more recent methods partition the face into a few region-of-interests (ROIs) (Polikovsky et al., 2009; Polikovsky and Kameda, 2013; Liong et al., 2015, 2016b,c, 2018; Davison et al., 2016b; Li et al., 2018). The ROIs are regions that correspond to one or more FACS action units (AUs). As such, these regions contain rigid facial motions when the muscles (AUs) are activated. Some studies (Liong et al., 2015, 2016b,c; Davison et al., 2016b) show that ROIs are more effective compared to the use of the entire face in constraining the salient locations for spotting.

3.2. Spotting

Facial micro-expression spotting, or *"micro-movement"* spotting [a term coined by Davison et al. (2016a)] refers to the problem of automatically detecting the temporal interval of a micro-movement in a sequence of video frames. Current approaches for spotting facial micro-movement can be broadly categorized into two groups: classifier-based methods (supervised/unsupervised) and rule-based (use of thresholds or heuristics) methods. There are many possible dichotomies; this survey discusses some early ideas, followed by two distinct groups of works – one on spotting ME movement or window of occurrence, another on spotting the ME apex. A summary of the existing techniques for spotting facial micro-expressions (or micro-movements) are depicted in **Table 3**.

3.2.1. Early Works

In the early works by Polikovsky et al. (2009) and Polikovsky and Kameda (2013), 3D-HOG was adopted to extract the features from each of the regions in the ME videos. Then, *k*-means clustering was used to cluster the features to particular AUs within predefined facial cubes. "Spotting" was approached as a classification task: each frame is classified to neutral, onset, apex or offset, and compared with ground truth labels. The classification rates achieved were satisfactory, in the range of 68–80%. Although their method could potentially contribute to facial micro-movement spotting by locating the four standard phases described by FACS, there are two glaring drawbacks. First, their method was only tested on posed facial ME videos, which are not a good representation of spontaneous (naturally induced) facial MEs. Secondly, the experiment was run as a classification task in which the frames were clustered into one of the four phases; this is highly unsuitable for real-time spotting. The work of Wu et al. (2011) also treats the spotting task as a classification process. Their work uses Gabor filters and the GentleSVM classifier to evaluate the frames. From the resulting label of each frame, the duration of facial micro-expressions were measured according to the transition points and the video frame-rate. Subsequently, they are only considered as ME when their durations last for 1/25–1/5s. They achieved very high spotting performance on the METT training database (Ekman, 2003). However, this was not convincing on two counts; first, only 48 videos were used in the experiments, and second, the videos were synthesized by inserting a flash of micro-expression in the middle of a sequence of neutral face images. In real-world conditions, frame transitions would be much more dynamic compared to the abrupt changes that were artificially added.

Instead of treating the spotting task as frame-by-frame classification, the works of Shreve et al. (2009, 2011) are the first to consider the temporal relation from frame-to-frame and employ a threshold technique to locate spontaneous facial MEs. This follows a more objective method that does not require machine learning. Their works are also the first in the literature to attempt spotting both macro- (i.e., ordinary facial expressions) and micro-expressions from videos. In their work, optical strain, which represents the amount of deformation incurred during motion, was computed from selected facial regions. Then, the

TABLE 3 | Facial micro-expression (or micro-movement) spotting works in literature.

Work	Feature	Feature Analysis	Movement (M) / Apex (A)	Spotting technique	Database
Polikovsky et al., 2009	3D gradient histogram	–		k mean cluster	High-speed ME database (not available)
Shreve et al., 2009	Optical strain	–	M	Threshold technique	USF
Wu et al., 2011	Gabor features	–	M	GentleSVM	METT (48 videos)
Shreve et al., 2011	Optical strain	–	M	Threshold technique	USF-HD Canal-9 (not available) Found videos (not available)
Polikovsky and Kameda, 2013	3D gradient histogram	–		k mean cluster	High-speed ME database (not available)
Shreve et al., 2014	Optical strain	–	M	Threshold technique	USF SMIC
Moilanen et al., 2014	LBP	✓	M	Threshold technique	CASME-A CASME-B SMIC-VIS-E
Davison et al., 2015	HOG	✓	M	Threshold technique	SAMM
Patel et al., 2015	Spatio-temporal integration of OF vectors	–	M	Threshold technique	SMIC-VIS-E
Liong et al., 2015	LBP correlation CLM Optical strain	–	A	Binary search	CASME II
Wang et al., 2016a	MDMD	✓	M	Threshold technique	CAS(ME)^2
Xia et al., 2016	Geometrical motion deformation	–	M	Random walk model	CASME SMIC
Liong et al., 2016b	LBP correlation	–	A	Binary search	CASME II
Liong et al., 2016c	LBP correlation Optical strain	–	A	Binary search	CASME II
Davison et al., 2016a	HOG	✓	M	Threshold technique	SAMM
Davison et al., 2016b	3D HOG LBP OF	✓	M	Threshold technique	SAMM CASME II
Li et al., 2017	HOOF LBP	✓	M	Threshold technique	CASME II SMIC-E-HS SMIC-E-VIS SMIC-E-NIR
Yan and Chen, 2017	LBP correlation CLM HOOF	–	A	Peak detection	CASME II
Ma et al., 2017	RHOOF	–	A	Threshold technique	CASME CASME II
Qu et al., 2017	LBP	✓	M	Threshold technique	CAS(ME)^2
Duque et al., 2018	Riesz Pyramid	✓	M	Threshold technique	SMIC-E-HS CASME II

facial MEs are spotted by tracking the strain magnitudes across frames following these heuristics: (1) strain magnitude exceeds the threshold (calculated from the mean of each video) and is significantly larger than that of the surrounding frames, and (2) the duration of the detected peak can only last at most 1/5th of a second. A 74% true positive rate and 44% false positive rate was achieved in the spotting task. However, a portion of data used in their experiments were posed, while some of them (Canal-9 and Found Videos databases) were not published or are currently defunct. In their later work (Shreve et al., 2014), a peak detector was applied to locate sequences containing MEs based on strain maps. However, the details of the peak detector and threshold value were not disclosed.

3.2.2. Movement Spotting

Micro-expression movements can be located by identifying a "window" of occurrence, typically marked by a starting or *onset* frame, and an ending or *offset* frame. In the work by Moilanen et al. (2014), the facial motion changes were modeled by feature difference (FD) analysis of appearance-based features (i.e., LBP) that incorporates the Chi-Square (χ^2) distance to form the FD magnitudes. Only the top 1/3 of total blocks (per frame) with the greatest FD values were chosen and averaged to an initial feature value representing the frame. The contrasting difference vector is then computed to find relevant peaks from across the sequence. Spotted peak frames (i.e., the peaks that exceed the threshold) are compared with the provided ground truth frames; and considered true positive if they fall within the span of $k/2$ frames (where k is half of the interval frames in the window) before the onset and after the offset. The proposed technique was tested on CASME-A, CASME-B, and SMIC-VIS-E, achieving a true positive rate of 52, 66, and 71%, respectively.

The same spotting approach was adopted by Li et al. (2017) and tested on various spontaneous facial ME databases: CASME II, SMIC-E-HS, SMIC-E-VIS, and SMIC-E-NIR. This work also indicated that LBP consistently outperforms HOOF in all the datasets with higher AUC (area-under-the-ROC-curve) values and lower false positive rates. To spot facial micro-expressions on the new CAS(ME)^2 database, the same spotting approach (Moilanen et al., 2014) is adopted by Wang et al. (2016a). Using their proposed main directional optical flow (MDMD) approach, ME spotting performance on the CAS(ME)^2 is 0.32, 0.35, and 0.33 for recall, precision and F1-score, respectively. For all these works (Moilanen et al., 2014; Wang et al., 2016a; Li et al., 2017; Qu et al., 2017), the threshold value for peak detection is set by taking the difference between the mean and max value of the contrasting difference vector and multiplying it by a fraction in the range of [0,1]. Finally, this value is added with the mean value of the contrasting difference vector to denote the threshold. By these calculations, at least one peak will always be detected as the threshold value will never exceed the maximum value of the contrasting difference vector. This could potentially result in misclassification of non-ME movements since it will *always* detect a peak. Besides, pre-defining the ME window intervals (which obtains the FD values) may not augur well with videos captured at different frame rates. To address the potentiality of a false peak, these works (Moilanen et al., 2014; Davison et al., 2015; Wang et al., 2016a; Li et al., 2017; Qu et al., 2017) proposed to compute the baseline threshold based on a neutral video sequence from each individual subject in the datasets.

In the work of Davison et al. (2015), all detected sequences which are less than 100 frames are denoted as true positives, in which eye blinks and eye gaze are included; while peaks that are detected but not coded as a movement are classed to false positives. The approach achieved scores of 0.84, 0.70, and 0.76 for recall, precision, and F1-measure, respectively on the SAMM database. In their later works, Davison et al. (2016a) and Davison et al. (2016b) introduced "individualized baselines," which are computed by taking a neutral video sequence for the participants and using the χ^2 distance to get an initial feature for the baseline sequence. The maximum value of this baseline feature is identified as the threshold. This improved their previous attempt by a good margin.

A number of innovative approaches were proposed. Patel et al. (2015) computed optical flow vectors over small local regions and integrated them into spatiotemporal regions to find the onset and offset times. In another approach, Xia et al. (2016) applied random walk model to compute the probability of frames containing MEs by considering the geometrical deformation correlation between frames in a temporal window. Duque et al. (2018) designed a system that is able to differentiate between MEs and eye movements by analyzing the phase variations between frames based on the Riesz Pyramid.

3.2.3. Apex Spotting

Besides spotting facial micro-movements, a few other works focused on spotting a specific type of ME phase, particularly the *apex* frame (Liong et al., 2015, 2016b,c; Yan and Chen, 2017). The apex frame, which is the instant indicating the most expressive emotional state in an ME sequence, is believed to be able to effectively reveal the true expression for the particular video. In the work by Yan and Chen (2017), the frame that has the largest feature magnitude was selected as the apex frame. A few interesting findings were revealed: CLM (which provides geometric features) is especially sensitive to contour-based changes such as eyebrow movement, and LBP (which produces appearance features) is more suitable for detecting changes in appearance such as pressing of lips; however, OF is the most all-rounded feature as it is able to spot the apex based on the resultant direction and movement of facial motions. A binary search method was proposed by Liong et al. (2015) to automatically locate the apex frame in a video sequence. By observing that the apex frames are more likely to appear in areas concentrated with peaks, the proposed binary search method iteratively partitions the sequence into two halves, by selecting the half that contains a higher sum of feature difference values. This is repeated until a single peak is left. The proposed method reported a mean absolute error (MAE) of 13.55 frames and standard error (SE) of 0.79 on CASME II using LBP difference features. A recent work by Ma et al. (2017) used Region HOOF (RHOOF) based on 5 regions of interests (ROIs) for apex detection, which resulted in more robust results.

3.3. Performance Metrics

The ME spotting task is akin to a binary detection task (ME is present/not present), hence typical performance metrics can be used. Moilanen et al. (2014) encouraged the use of a Receiver Operating Characteristic (ROC) curve, which was adopted in most subsequent works (Patel et al., 2015; Xia et al., 2016; Li et al., 2017). In essence, the spotted peaks, which are obtained based on a threshold level, will be compared against ground truth labels to determine whether they are true or false spots. If one spotted peak is located within the frame range of [onset - $\frac{N-1}{4}$, offset + $\frac{N-1}{4}$] of a labeled ME clip, the spotted sequence (N frames centered at the peak) will be considered as a true positive ME; otherwise the N frames of spotted sequence will be counted as false positive frames. The specified range considers a tolerance interval of 0.5 s, which corresponds to the presumed maximum duration of MEs. To obtain the ROC curve, true positive rate (TPR), and false positive rate (FPR) are computed as follows:

$$\text{TPR} = \frac{\text{Number of frames of correctly spotted MEs}}{\text{Total number of ground truth ME frames from all samples}} \tag{1}$$

$$\text{FPR} = \frac{\text{Number of incorrectly spotted frames}}{\text{Total number of non-ME frames from all samples}} \tag{2}$$

Recently, Tran et al. (2017) proposed a micro-expression spotting benchmark (MESB) to standardize the performance evaluation of the spotting task. Using a sliding window based multi-scale evaluation and a series of protocols, they recognize the need for a fairer and more comprehensive method of assessment. Taking a leaf out of object detection, the Intersection over Union (IoU) of the detection set and ground truth set was proposed to determine if a sampled sub-sequence window is positive or negative for ME (threshold set at 0.5).

Several works that focused on the spotting of the apex frame (Yan et al., 2014b; Liong et al., 2015, 2016b,c) used Mean Absolute Error (MAE) to compute how close are the estimated apex frames to the ground-truth apex frames:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|e_i| \tag{3}$$

When spotting is performed on the raw long videos, Liong et al. (2016c) introduced another measure called Apex Spotting Rate (ASR), which calculates the success rate in spotting apex frames within the given onset and offset range of a long video. An apex frame is scored 1 if it is located between the onset and offset frames, and 0 otherwise:

$$\text{ASR} = \frac{1}{N}\sum_{i=1}^{N}\delta_i \tag{4}$$

$$\text{where} \quad \delta = \begin{cases} 1, & \text{if } f^* \in (f_{i,\text{onset}}, f_{i,\text{offset}}) \\ 0, & \text{otherwise} \end{cases}$$

4. RECOGNITION OF FACIAL MICRO-EXPRESSIONS

ME recognition is a task that classifies an ME video into one of the universal emotion classes (e.g., Happiness, Sadness, Surprise, Anger, Contempt, Fear, and Disgust). However, due to difficulties in the elicitation of micro-expressions, not all classes are available in the existing datasets. Typically, the emotion classes of the collected samples are unevenly distributed; some are easier to elicit hence they have more samples collected.

Technically, a recognition task involves feature extraction and classification. However, a pre-processing stage could be involved prior to the feature extraction to enhance the availability of descriptive information to be captured by descriptors. In this section, all the aforementioned steps are discussed.

4.1. Pre-processing

A number of fundamental pre-processes such as face landmark detection and tracking, face registration and face region retrieval, have all been discussed in section 3 for the spotting task. Most recognition works employ similar techniques as those used for spotting, i.e., ASM (Milborrow and Nicolls, 2014), DRMF (Asthana et al., 2013), Face++ (Megvii, 2013) for landmark detection; LWM (Goshtasby, 1988) for face registration. Meanwhile, division of the facial area into regions is a step often found within various feature representation techniques (discussed in section 4.2) to further localize features that change subtly. Aside from these known pre-processes, two essential pre-processing techniques have been instrumental in conditioning ME data for the purpose of recognition. We discuss these two steps which involve *magnification* and *interpolation* of ME data.

The uniqueness of facial micro-expressions is in its subtleness, which is one of reasons why recognizing them automatically is very challenging. As the intensity levels of facial ME movements are very low, it is extremely difficult to discriminate ME types among themselves. One solution to this problem is to exaggerate or magnify these facial micro-movements. In recent works (Park et al., 2015; Zarezadeh and Rezaeian, 2016; Li et al., 2017; Wang et al., 2017), the Eulerian Motion Magnification (EMM) (Wu et al., 2012) method was employed to magnify the subtle motions in the ME videos. The EMM method extracts the frequency bands of interest from the different spatial frequency bands obtained from the decomposition of an input video, by using band-pass filters; these extracted bandpass signals at different spatial level are amplified by a magnification factor α to magnify the motions. Li et al. (2017) demonstrated that the EMM method helps to enlarge the difference between different categories of micro-expressions (i.e., inter-class difference); thus the recognition rate is increased. However, larger amplification factors may cause undesirable amplified noise (i.e., motions that are not induced by MEs), which may degrade recognition performance. To prevent over-magnifying ME samples, Le Ngo et al. (2016a) theoretically estimated the upper bounds of effective magnification factors. Besides, the authors also compared the performance of the amplitude-based Eulerian motion magnification (A-EMM) and phase-based Eulerian motion magnification (P-EMM); with the To deal with the distinctive temporal characteristic of different ME classes, a magnification scheme was proposed by Park et al. (2015) to adaptively select the most discriminative frequency band needed for EMM to magnify subtle facial motions. A recent work by Le Ngo et al. (2018) showed that Global Lagrangian Motion Magnification (GLMM) can contribute toward better recognition capability compared to local Eulerian based approaches, particularly at higher magnification factors.

Another concern for ME recognition is with the uneven length (or duration) of ME video samples. In fact, it can contribute to two contrasting scenarios: (a) the case of short duration videos, which restricts the application of the feature extraction techniques which require varied temporal window size (e.g., LBP-based methods that can form binary patterns from varied radius); and (b) the case of long duration videos, whereby redundant or replicated frames (due to high frame rate capture) could deteriorate the recognition performance. To solve the problem, the temporal interpolation method (TIM) is applied to either up-sample (clips that are too short) or down-sample (clips that are too long) clips to produce clips of similar frame lengths.

Briefly, TIM takes original frames as input data to construct a manifold of facial expressions; then it samples on the manifolds for a particular number of output frames (refer to Zhou et al., 2011 for detailed explanation). It is shown by Li et al. (2017)

that modifying the frame length of ME videos can improve the recognition performance if the number of interpolated frames are small. However, when the interpolated frames are increased, the recognition performance is somewhat hampered due to over-interpolation. Therefore, the appropriate interpolation of the ME sequence is vital in preparation for recognition. An alternative technique Sparsity-Promoting Dynamic Mode Decomposition (DMDSP) (Jovanović et al., 2014) was adopted by Le Ngo et al. (2015) and Le Ngo et al. (2016b) to select only significant dynamics in MEs to form sparse structures. From the comprehensive experimental results shown in Le Ngo et al. (2016b), DMDSP achieved better recognition performance compared to TIM (on similar features and classifiers) due to its ability to keep only the significant temporal structures while eliminating irrelevant facial dynamics.

While the aforementioned pre-processing techniques showed positive results in improving ME recognition, yet these methods will notably lengthen the computation time of the overall recognition process. For a real-time system to be feasible, this cost has to be taken into consideration.

4.2. Representations

In the past few years, research in automatic ME analysis have been much focused on the problem of ME recognition: given an ME video sequence/clip, the purpose of recognition is to estimate its emotion label (or class). **Table 4** summarizes the existing ME methods in the literature. From the perspective of feature representations, they can be roughly divided into two main categories: *single-level* approaches and *multi-level* approaches. Single-level approaches refer to frameworks that directly extract feature representations from the video sequences; while for multi-layer approaches, the image sequences are first transformed into another domain or subspace prior to feature representation to exploit other kinds of information to describe MEs.

Feature representation is a transformation of raw input data to a succinct form; typically in face processing, representations can be from two distinct categories: geometric-based or appearance-based (Zeng et al., 2009). Specifically, geometric-based features describe the face geometry such as the shapes and locations of facial landmarks; whereas appearance-based features describe intensity and textural information such as wrinkles, furrows, and other patterns that are caused by emotion. However from previous studies in facial expression recognition (Fasel and Luettin, 2003; Zeng et al., 2009), it is observed that appearance-based features are better than geometric-based features in coping with illumination changes and mis-alignment error. Geometric-based features might not be as stable as appearance-based features as they need precise landmark detection and alignment procedures. For these similar reasons, appearance-based feature representations have become more popular in the literature on ME recognition

4.2.1. LBP-Based Methods

Among appearance-based feature extraction methods, local binary pattern on three orthogonal planes (LBP-TOP) is widely applied in many works (Li et al., 2013; Guo et al., 2014; Le Ngo et al., 2014, 2015, 2016a,b; Yan et al., 2014a; Adegun and Vadapalli, 2016; Zheng et al., 2016; Wang et al., 2017). Most existing datasets (SMIC, CASME II, SAMM) have all reported the LBP-TOP as their baseline evaluation method. LBP-TOP is an extension of its low-level representation, local binary pattern (LBP) (Ojala et al., 2002), which describes local texture variation along a circular region with binary codes which are then encoded into a histogram. LBP-TOP extracts features from local spatio-temporal neighborhoods over three planes: the spatial (XY) plane similarly to the regular LBP, the vertical spatio-temporal (YT) plane and the horizontal spatio-temporal (XT) plane; this enables LBP-TOP to dynamically encode temporal variations.

Subsequently, several variants of LBP-TOP were proposed for the ME recognition task. Wang et al. (2014b) derived Local Binary Pattern— Six Interception Points (LBP-SIP) from LBP-TOP by considering only the 6 unique points lying on three intersecting lines of the three orthogonal planes as neighbor points for constructing the binary patterns. By reducing redundant information from LBP-TOP, LBP-SIP reported better performance than LBP-TOP in this task. A more compact variant, LBP-MOP (Wang et al., 2015b) was constructed by concatenating the LBP features from only three mean images, which are the temporal pooling result of the image stacks along the three orthogonal planes. The performance of LBP-MOP was comparable to LBP-SIP, but with its computation time dramatically reduced. While LBP considers only pixel intensities, spatio-temporal completed local quantized patterns (STCLQP) (Huang et al., 2016) exploited more information containing sign, magnitude, and orientation components. To address the sparseness problem (in most LBP variants), specific codebooks were designed to reduce the number of possible codes to achieve better compactness.

Recent works have yielded some interesting advances. Huang and Zhao (2017) proposed a new binary pattern variant called spatio-temporal local Radon binary pattern (STRBP) that uses Radon transform to obtain robust shape features. Ben et al. (2017) proposed an alternative binary descriptor called Hot Wheel Patterns (HWP) (and its spatio-temporal extension HWP-TOP) to encode the discriminative features of both macro- and micro-expressions images. A coupled metric learning algorithm is then used to model the shared features between micro- and macro-expression information.

4.2.2. Optical Flow-Based Methods

As suggested in several studies (e.g., Li et al., 2017), the temporal dynamics that reside along the video sequences are found to be essential in improving the performance of ME recognition. As such, optical flow (OF) (Horn and Schunck, 1981) based techniques, which measure the spatio-temporal changes in intensity, came into contention as well.

In the work by Xu et al. (2017), a proposal to extract only principal directions of the OF maps was purportedly to eliminate abnormal OF vectors that resulted from noise or illumination changes. A similar concept of exploiting OF in the main direction was employed by Liu et al. (2016) to design main directional mean optical flow (MDMO) features. MDMO is a ROI-based OF feature, which considers both local statistic (i.e., the mean of OF vectors in the bin with the maximum count in each ROI) and its spatial location (i.e., the ROI to which it belongs). Unlike the aforementioned works which exploited only the single dominant direction of OF in each facial region, Allaert et al. (2017) determined the consistent facial motion, which could be in multiple directions from a single facial region. The assumption was made based on the fact that facial motions spread progressively due to skin elasticity, hence only the directions that are coherent in

TABLE 4 | Benchmarking facial micro-expression recognition works in literature.

Papers	Pre-processing	Features	Classifier	Accuracy (%)		F1-score (%)	
				CASME II	SMIC	CASME II	SMIC
LOSO							
Li et al., 2013	–	LBP-TOP	SVM	–	48.78	–	–
Liong et al., 2016a	–	OSF + OS and weighted LBP-TOP	SVM	–	52.44	–	–
Liong et al., 2014a	–	OS	SVM	–	53.56	–	–
Liong et al., 2014b	–	OS weighted LBP-TOP	SVM	42.00	53.66	0.38	0.54
Le Ngo et al., 2014	–	STM	Adaboost	43.78	44.34	0.3337	0.4731
Wang et al., 2015b	–	LBP-MOP	SVM	44.13	50.61	–	–
Xu et al., 2017	–	Facial Dynamics Map	SVM	45.93	54.88	0.4053	0.538
Oh et al., 2016	–	Monogenic + LBP-TOP	SVM	–	–	0.41	0.44
Oh et al., 2015	–	Riesz wavelet + LBP-TOP	SVM	–	–	0.43	–
Liong et al., 2018	ROIs	LBP-TOP	SVM	46.00	54.00	0.32	0.52
Wang et al., 2014b	–	LBP-SIP	SVM	46.56	44.51	0.448	0.4492
Le Ngo et al., 2016a	A-EMM	LBP-TOP	SVM	–	–	0.51	–
Le Ngo et al., 2016b	DMDSP	LBP-TOP	SVM	49.00	58.00	0.51	0.60
Park et al., 2015	Adaptive MM	LBP-TOP	SVM	51.91	–	–	–
Happy and Routray, 2017	–	HFOFO	SVM	56.64	51.83	0.5248	0.5243
Liong et al., 2016b	–	Bi-WOOF	SVM	–	–	0.56	0.53
Huang et al., 2016	–	STCLQP	SVM	58.39	64.02	0.5836	0.6381
Huang et al., 2015	–	STLBP-IP	SVM	59.51	57.93	0.57*	0.58*
Liong et al., 2016c	–	Bi-WOOF (apex frame)	SVM	–	–	0.61	0.62
He et al., 2017	–	MMFL	SVM	59.81	63.15	–	–
Kim et al., 2016	–	CNN + LSTM	Softmax	60.98	–	–	–
Liong and Wong, 2017	–	Bi-WOOF + Phase	SVM	62.55	68.29	0.65	0.67
Zheng et al., 2016	–	LBP-TOP	RK-SVD	63.25		–	–
Zong et al., 2018a	–	Hierarchical STLBP-IP	KGSL	63.83	60.78	0.6110	0.6126
Huang and Zhao, 2017	TIM	STRBP	SVM	64.37	60.98	–	–
Huang et al., 2017	–	Discriminative STLBP-IP	SVM	64.78	63.41	–	–
Allaert et al., 2017	–	OF Maps	SVM	65.35	–	–	–
Li et al., 2017	TIM+EVM	HIGO	SVM	67.21	68.29	–	–
Zheng, 2017 †‡	–	2DSGR	SRC	–	71.19	–	–
Liu et al., 2016 †	–	MDMO	SVM	67.37	80.00	–	–
Davison et al., 2017 ‡	–	HOOF	SVM	76.60	–	0.55	–
LOVO							
Wang et al., 2015a †‡	TIM	LBP-TOP on TICS	SVM	62.30	–	–	–
Yan et al., 2014a	–	LBP-TOP	SVM	63.41	–	–	–
Wang et al., 2014a	TIM	DLSTD	SVM	63.41	68.29	–	–
Happy and Routray, 2017	–	HFOFO	SVM	64.06	56.10	0.6025	0.5536
Liong et al., 2014b	–	OS weighted LBP-TOP	SVM	65.59	–	–	–
Wang et al., 2015b	–	LBP-MOP	SVM	66.80	60.98	–	–
Wang et al., 2014b	–	LBP-SIP	SVM	67.21	–	–	–
Ping et al., 2016		LBP-TOP	GSLSR	67.89	70.12	–	–
Park et al., 2015	Adaptive MM	LBP-TOP	SVM	69.63	–	–	–
Wang et al., 2017	EVM	LBP-TOP	SVM	75.30	–	–	–
Li et al., 2017	TIM+EVM	HIGO	SVM	78.14	75.00	–	–
OTHER PROTOCOLS							
Zhang et al., 2017 *Evenly Distributed*	–	LBP-TOP and HOOF	RF	62.5	–	-–	–
Jia et al., 2017 *Random Test (20 times)*	–	SVD+ LBP/LBP-TOP	KNN	65.5	–	–	–
Peng et al., 2017 §‡ *three-fold cross-validation*	–	DTSCNN	SVM	66.67	–	–	–
Adegun and Vadapalli, 2016 † *five-fold cross-validation*	–	LBP-TOP	ELM	96.12	–	–	–

†*Not all the samples in the dataset were used in the experiments.*

‡*Different number of emotion classes were used in the experiments.*

§*Combined CASME I/II database was used.*

**Not reported in paper, but computed from confusion table provided.*

the neighboring facial regions are extracted to construct a consistent OF map representation.

Motivated by the use of optical strain (OS) for ME spotting (Shreve et al., 2009, 2014), Liong et al. (2014a) proposed to leverage on its strengths for ME recognition. OS is derived from OF by computing the normal and shear strain tensor components of the OF. This enables the capture of small and subtle facial deformation. In their work, the OS magnitude images are temporally pooled to form a single pooled OS map; then the resulting map is max-normalized and resized to a fixed smaller resolution before transforming into a feature vector that represent the video. To emphasize the importance of active regions, the authors (Liong et al., 2014b) proposed to weight local LBP-TOP features with different weights which were generated from the temporal mean-pooled OS map. This allows regions that actively exhibit MEs to be given more significance, hence increasing the discrimination between emotion types. In a more recent attempt, Liong et al. (2016b) proposed a Bi-Weighted Oriented Optical Flow (BI-WOOF) descriptor which applies two schemes to weight the HOOF descriptor locally and globally. Locally, the magnitude components were used to weight the orientation bins within each ROI; the resultant locally weighted histograms are then weighted again (globally) by multiplying with the mean optical strain (OS) magnitude of each ROI. Intuitively, a larger change in the pixel's movement or deformation will contribute toward a more discriminative histogram. Instead of considering the whole image sequences, the authors also demonstrated promising recognition performance using only two frames (i.e., the onset frame and the apex frame) instead of using whole sequences. This was able to reduce the processing time by a large margin.

Zhang et al. (2017) proposed to aggregate the histogram of the oriented optical flow (HOOF) (Chaudhry et al., 2009) with LBP-TOP features region-by-region to generate local statistical features. In their work, they revealed that fusing of local features within each ROI can capture more detailed and representative information than globally done. In the work by Happy and Routray (2017), fuzzy histogram of optical flow orientation (FHOFO) was proposed for ME recognition. In HFOFO, the histograms are only the collection of orientations without being weighted by the optical flow magnitudes; the assumption was made that MEs are so subtle that the induced magnitudes should be ignored. They also introduced a fuzzification process that considers the contribution of an orientation angle to its surrounding bins based on fuzzy membership functions; as such smooth histograms for motion vector are created.

4.2.3. Other Methods

Aside from methods based on low-level features, there are also numerous techniques proposed to extract other types of feature representations. Lu et al. (2014) proposed a Delaunay-based temporal coding model (DTCM) to encode the local temporal variation (in grayscale values) in each subregion obtained by Delaunay triangulation and preserve the ones with high saliency as features. In the work of Li et al. (2017), the histogram of image gradient orientation (HIGO), which is a degenerate variant of HOG, was employed in the recognition task. It uses simple vote rather than weighted vote when counting the responses of the gradient orientations. As such, it could depress the influence of illumination contrast by ignoring the magnitude. The use of color space was also experimented in the work of Wang et al. (2015a), where LBP-TOP features were extracted from Tensor Independent Color Space (TICS). In TICS, the three color components (R, G, and B) were transformed into three uncorrelated components which are as independent as possible to avoid redundancy and thus increase the recognition performance. The Sparse Tensor Canonical Correlation Analysis (STCCA) representation proposed by Wang et al. (2016b) offers a solution to mitigate the sparsity of spatial and temporal information in a ME sequence.

Signal components such as magnitude, phase and orientation can be exploited as features for ME recognition. Oh et al. (2015) proposed a monogenic Riesz wavelet framework, where the decomposed magnitude, phase, and orientation components (which represent energy, structural and geometric information respectively) are concatenated to describe MEs. In their extended work (Oh et al., 2016), higher-order Riesz transform was adopted to exploit the intrinsic two-dimensional (i2D) local structures such as corners, junctions, and other complex contours. They demonstrated that i2D structures are better representative parts than i1D structures (i.e., simple structures such as lines and straight edges) in describing MEs. By supplementing the robust Bi-WOOF descriptor (Liong et al., 2016b) with Riesz monogenic phase information derived from the onset-apex difference image (Liong and Wong, 2017), recognition performance can be further boosted.

Integral projections are an easy way of simplifying spatial data to obtain shape information along different directions. The LBP-Integral Projection (IP) technique proposed by Huang et al. (2015) applies the LBP operator on these projections. A difference image is first computed from successive frames (to remove face identity) before it is projected into two parts: vertical projection and horizontal projection. This method was found to be more effective than directly using features derived from the original appearance information. In their extended work (Huang et al., 2017), original pixel information is replaced by extracted subtle emotion information as input for generating spatio-temporal local binary pattern with revisited integral projection (STLBP-RIP) features. To further enhance the discriminative power of these features, only features with the smallest Laplacian scores are selected as the final feature representation.

A few works increase the significance of features by means of excluding irrelevant information such as pose and subject identity, which may obstruct salient emotion information. Robust principal component analysis (RPCA) (Wright et al., 2009) was adopted in Wang et al. (2014a) and Huang et al. (2016) to extract subtle emotion information for feature extraction. In Wang et al. (2014a), the extracted subtle emotion information was encoded by local spatio-temporal directional (LSTD) to extract more detailed spatio-temporal directional changes on the *x*, *y*, and *t* directions from each plane (XY, XT, and YT). Lee et al. (2017) proposed an interesting use of Multimodal Discriminant Analysis (MMDA) to orthogonally decompose a sample into three modes or "identity traits" (emotion, gender and race) in a simultaneous manner. Only the essential emotion components are magnified before the samples are synthesized and reconstructed.

Recently, numerous new works have begun exploring other forms of representation and mechanisms. He et al. (2017) proposed a strategy to extract low-level features from small regions (or cubes) of a video by learning a set of class-specific feature mappings.

Jia et al. (2017) devised a macro-to-micro transformation model based on singular value decomposition (SVD) to recognize MEs by utilizing macro-expressions as part of the training data. This overcomes the lack of labeled data in MEs databases. There were various recent attempts at casting the recognition task as one arising from a different problem. Zheng (2017) formulated it as a sparse approximation problem and presented the 2D Gabor filter and sparse representation (2DSGR) technique for feature extraction. Zhu et al. (2018) drew inspiration from similarities between MEs and speech to propose a transfer learning method that projects both domain signals to a common subspace. In a radical move, Davison et al. (2017) proposed to re-group MEs based on Action Units (AUs) instead of by emotion categories, which are arguably susceptible to bias in self-reports used during the construction of dataset. Their experimental results on CASME II and SAMM showed that recognition performance should be higher than what is currently expected from other works that used emotion labels.

4.3. Classification

The last stage in an ME recognition task involves the classification of the emotion type. Various types of classifiers have been used for the task of ME recognition such as *k*-Nearest Neighbor (*k*-NN), support vector machine (SVM), random forest (RF), sparse representation classifier (SRC), Relaxed K-SVD, group sparse learning (GSL) and extreme learning machine (ELM). From the literature, the most widely used classifier is the SVM. SVMs are computational algorithms that construct a hyperplane or a set of hyperplanes in a high or infinite dimensional space (Cortes and Vapnik, 1995). During the training of SVM, the margins between the borders of different classes are sought to be maximal. Compared to other classifiers, SVMs are robust, accurate, and very effective even in cases where the number of training samples is small. On the contrary, two other notable classifiers—RF and *k*-NN are seldom used in the ME recognition task. Although the RF is generally quicker than SVM, it is prone to overfit when dealing with noisy data. The *k*-NN uses an instance-based learning process which may not be suitable for sparse high-dimensional data such as face data.

To deal with the sparseness of MEs, several works tried using relaxed K-SVD, SRC, and GSL techniques for classification. However, each of these methods tackle the sparseness of MEs differently. The relaxed K-SVD (Zheng et al., 2016) learns a sparse dictionary to distinguish different MEs by minimizing the variance of sparse coefficients. The SRC (Yang et al., 2012) used in Zheng (2017) represents a given test sample as a sparse linear combination of all training samples; hence the sparse nonzero representation coefficients are likely to concentrate on training samples that are of the same class as the test sample. A Kernelized GSL (Zong et al., 2018a) is proposed to facilitate the process of learning a set of importance weights from hierarchical spatiotemporal descriptors that can aid the selection of the important blocks from various facial blocks. Neural networks can offer a one-shot process (feature extraction and classification), with a remarkable ability to extract complex patterns from data. However, a substantial amount of labeled data is required to properly train a neutral network without overfitting it, resulting in it being less favorable for ME recognition since labeled data is limited. The ELM (Huang et al., 2006), which is naturally just feed-forward network with a single hidden layer was used by Adegun and Vadapalli (2016) to classify MEs.

4.4. Experimental Protocol and Performance Metrics

The original dataset papers (Li et al., 2013; Yan et al., 2014a; Davison et al., 2016a) all propose the adoption of the Leave-One-Subject-Out (also known as "LOSO") cross-validation as the default experimental protocol. This is done with consideration that the samples were collected by eliciting the emotions from a number of different participants (i.e., S number of subjects). As such, cross validation should be carried out by withholding a particular subject s while the other $S-1$ subjects are used in the training step. This removes the potential identity bias that may arise during the learning process; a subject that is being evaluated could have been seen and learned in the training step. A number of other works used the Leave-One-Video-Out ("LOVO") cross-validation protocol instead, which exhaustively divides all samples into S number of possible train-test partitions. This protocol is deemed to avoid irregular partitioning but is often likely to overestimate the performance of the classifier. A few works opted to report their results using their own choice of evaluation protocol, such as an evenly distributed sets (Zhang et al., 2017), random sampling of test partition (Jia et al., 2017), and five-fold cross validation (Adegun and Vadapalli, 2016). Generally, the works in literature can be categorized into these three groups, as shown in **Table 4**.

The ME recognition task reports the typical performance metric of *Accuracy*, which is commonly used in other image/video recognition problems. A majority of works in the literature report the Accuracy metric, which is simply the number of correctly classified video sequences over the total number of video sequences in the dataset. However, due to the imbalanced nature of the ME datasets which was first discussed by Le Ngo et al. (2014), Accuracy scores can be highly skewed toward classes that are larger as classifiers tend to learn poorly from classes that are less represented. Consequently, it makes more sense to report the *F1-Score* (or F-measure), which is the harmonic mean of the *Precision* and *Recall*:

$$\textit{F1-Score} = 2 \cdot \frac{\textit{Precision} \cdot \textit{Recall}}{\textit{Precision} + \textit{Recall}} \tag{5}$$

$$\textit{Precision} = \frac{tp}{tp + fp} \tag{6}$$

$$\textit{Recall} = \frac{tp}{tp + fn} \tag{7}$$

where tp, fp, and fn are the number of true positives, false positives, false negatives, respectively. The overall performance of a method can be reported by *macro-averaging* across all classes (i.e., compute scores for each class, then average them), or by *micro-averaging* across all classes (i.e., summing up the individual tp, fp, and fn in the entire set before computing scores).

5. CHALLENGES

The studies reviewed in sections 2, 3, and 4 show the progress in the research work in ME analysis. However, there is still considerable room for improvement in the performance of ME spotting and recognition. In this section, some recognized problems in existing databases and challenging issues in both tasks are discussed in detail.

5.1. Databases

Acquiring valuable spontaneous ME data and their ground truth is far from being solved. Among the various affective states, certain emotions (such as happiness) are relatively easier to be elicited compared to others (e.g., fear, sadness, anger) (Coan and Allen, 2007). Consequently, there is an imbalanced distribution of samples per emotion and number of samples per subject. This could be biased toward particular emotions that constitute a larger portion of the training set. To address this issue, a more effective way of eliciting affective MEs (especially to those are relatively difficult) should be discovered. Social psychology has suggested creative strategies for inducing affective expressions that are difficult to elicit (Coan and Allen, 2007). Some works have underlined the possibility of using other complementary information from the body region (Song et al., 2013) or instantaneous heart rate from skin variations (Gupta et al., 2018) to better analyze micro-expressions.

Almost all the existing datasets contain a majority of subjects from one particular country or ethnicity. Though it is common knowledge that basic facial expressions are universal across the cultural background, nevertheless subjects from different backgrounds may express differently toward the same elicitation, or at least with different intensity level as they may have different ways of expressing an emotion. Thus, a well-established database should comprise a diverse range of ethnic groups to provide better generalization for experiments.

Although much effort has been paid toward the collection of databases of spontaneous MEs, some databases (e.g., SMIC) lack important metadata such as FACS. It is generally accepted that human facial expression data need to be FACS coded. The main reason being that FACS AUs are objective descriptors and independent of subjective interpretation. Moreover, it is also essential to report the reliability measure of the inter-observers (or inter-coders) involved in the labeling of data.

Considering the implementation of real-life applications of ME recognition in the near future, existing databases that are constructed under studio environments, may not best represent MEs exhibited in real-life situations. Thus, developing and introducing real-world ME databases could bring about a leap of progress in this domain.

5.2. Spotting

Recent work on the spotting of MEs have achieved promising results on successfully locating the temporal dynamics of micro-movements; however, there is room for improvement as the problem of spotting MEs remains a challenging task to date.

5.2.1. Landmark Detection

Even though the facial landmark detection algorithms have made remarkable progress over the past decade, the available landmark detectors are not always accurate or steady. The unsteadiness of face alignment based on imprecise facial landmarks may result in significant noise (i.e., rigid head movements and eye gaze) associated with dynamic facial expressions. This in turn increases the difficulty in detecting the correct MEs. Thus, a more advanced robust facial landmark detection is required to correctly and precisely locate the landmark points on the face.

5.2.2. Eyes: To Keep or Not Keep?

To avoid the intrusion of eye blinks, majority of works in the literature simply mask out the eye regions. However, according to some findings (Zhao et al., 2011; Vaidya et al., 2014; Lu et al., 2015; Duan et al., 2016), the eye region is one of the most discriminative regions for affect recognition. As many spontaneous MEs involving muscles around eye regions, there is a need to differentiate between the eye blinks corresponding to certain expressions and those that are merely irrelevant facial motions. In addition, the onsets of the many MEs also temporally overlap with eye blinks (Li et al., 2017). Thus, this warrants a more robust approach at dealing with overlapping occurrences of facial motions.

5.2.3. Feature-Based or Rule-Based?

The few studies (Liong et al., 2015; Yan and Chen, 2017) investigated the effectiveness of individual feature descriptors in capturing the micro-movements for the ME spotting task. They have showed that micro-movements that are induced from different facial components actually resulted in motion changes from different perspectives such as appearance, geometric, and etc. For example, lifting up or down the eyebrows results in a clear contour change (geometrical), which could be effectively captured by geometric-based feature descriptors; pressing of lips could cause the variation in appearance but not the position, and thus appearance-based feature descriptors can capture these changes. Interestingly, they reported that motion-based features such as optical flow based features outperformed appearance-based and geometric-based features in the ME spotting. The problem remains that the assumptions made by optical flow methods are likely to be violated in unconstrained environments, rendering real-time implementation challenging.

Majority of existing efforts toward the spotting of MEs employ rule-based approaches that rely on thresholds. Frames with magnitude exceeding the pre-defined threshold value are the frames (i.e., the temporal dynamics) where ME appears. However, prior knowledge is required to set the appropriate threshold for distinguishing the relevant peaks from local magnitude variation and background noise. This is not really practical in the real-time domain. Instead, Liong et al. (2015) designed a simple divide-and-conquer strategy, which does not require a threshold to locate the temporal dynamics of MEs. Their method finds the apex frame based on a high concentration of peaks.

5.2.4. Onset and Offset Detection

Further steps should also be considered to locate the onset and offset frames of these ME occurrences. While it is relatively easier to identify the peaks and valleys of facial movements, the onset and offset frames are much more difficult to determine. The task of locating the onset and offset frames will be even tougher when dealing with real-life situations where facial movements are continuously changing. Thus, the indicators and criteria for determining the onset and offset frames need to be properly defined and further studied. Spotting the ME onset and offset frames is a crucial step which can lead to automatic ME analysis.

5.3. Recognition

In the past few years, much effort has been done toward ME recognition, including developing new features to better describe MEs. However, due to the short elapsed duration and low intensity of MEs, there is still room for improvement toward achieving satisfactory accuracy rates. This could be due to several possible reasons.

5.3.1. Block Selection

In most works, block-based segmentation of a face to extract local information is a common practice. Existing efforts employed block-based segmentation of a face without considering the contribution from each of the blocks. Ideally, the contribution from all blocks should be varied, whereby the blocks containing the key facial components such as eyebrows, eyes, mouth, and cheek should be highlighted as the motion changes at these regions convey meaningful information from differentiating different MEs. Higher weights can be assigned to those regions that contain key facial components to enhance the discriminative power. Alternatively, the discriminative features from the facial blocks can be selected through a learning process; the recent work of Zong et al. (2018a) offers a solution to this issue.

5.3.2. Type of Features

Since the emergence of the ME recognition works, many different feature descriptors have been proposed for MEs. Due to the characteristic of the feature descriptors, the extracted features might carry different information (e.g., appearance, geometric, motion, etc). For macro-expressions, it has been shown in (Fasel and Luettin, 2003) and Zeng et al. (2009) that geometric-based features performed poorer than appearance- and motion-based features as they are highly dependent on the precision of facial landmark points. However, recent ME works (Huang et al., 2015, 2017) show that shape information is arguably more discriminative for identifying certain MEs. Perhaps different features may carry meaningful information for different expression types. This should be carefully exploited and taken into consideration during feature extraction process.

5.3.3. Deep Learning

The advancement of Deep Learning has prompted the community to look for new ways of extracting better features. However, a crucial ingredient to this remains as to the feasible amount of data necessary to train a model that does not over-fit easily; the small scale of data (lack of ME samples per category) and the imbalanced distribution of samples are the primary obstacles. Recently an approach by Patel et al. (2016) made an attempt to utilize deep features transferred from pretrained ImageNet models. The authors deemed that fine-tuning the network to the ME datasets is not plausible (insufficient data) and opted for a feature selection scheme. Some other works (Kim et al., 2016; Peng et al., 2017) have also begun exploring the use of deep neural networks by encoding spatial and temporal features learned from network architectures that are relatively "shallower" than those used in the ImageNet challenge (Russakovsky et al., 2015). This may be a promising research direction in terms of advancing the features used for this task.

5.3.4. Cross-Database Recognition

Another on-going development that challenges existing experimental presuppositions is cross-database recognition. This setup mimics a realistic setting where training and test samples may come from different environments. Current recognition performance based on single databases, is expected to plunge under such circumstances. Zong et al. (2017, 2018b) proposed a domain regeneration (DR) framework, which aims to regenerate micro-expression samples from source and target databases. The authors aptly point out that much is still to be done to discover more robust algorithms that work well across varying domains. The first ever Micro-Expression Grand Challenge (Yap et al., 2018) was held with special attention given to the importance of cross-database recognition settings. Two protocols – Hold-out Database Evaluation (HDE) and Composite Database Evaluation (CDE), were proposed in the challenge, using the CASME II and SAMM databases. The reported performances (Khor et al., 2018; Merghani et al., 2018; Peng et al., 2018) were poorer than most other works that apply only to single databases, indicating that future methods need to be more robust across domains.

5.4. Experiment Related Issues

5.4.1. Evaluation Protocol

An important issue that should be addressed in ME recognition is how the data is evaluated. Due to the different evaluation protocols used in existing works, a fair comparison among these works could not be adequately established. Currently, the two popular evaluation protocols that are widely applied in ME recognition are: leave-one-video-out cross-validation (LOVOCV) and leave-one subject-out cross validation (LOSOCV). The common *k*-fold cross-validation is not suitable as the current publicly available spontaneous ME datasets are highly imbalanced (Le Ngo et al., 2014). The number of samples per subject and the number of samples per emotion class in these datasets vary quite considerably. For instance, in the CASME II dataset, the number of samples that belong to the "Surprise" class is 25 compared to the 102 samples of the "Others" class; while the difference in the number of samples for "Subject 08" and "Subject 17" are 8 and 34, respectively. As such, with *k*-fold cross-validation, the fairness in evaluation is likely to be questionable. The same goes with employing LOVOCV, where only one video sample is left out as the test sample while the remaining samples are used for training; subsequently, the average accuracy across all folds is taken as the final result. This can possibly introduce additional biases on certain subjects that have more representation during the evaluation process. Moreover, the performance of such a protocol typically over-estimates the actual classifier performance due to a substantially large training set. It is paramount to stress that the LOSOCV protocol is a more convincing evaluation protocol as it separates the samples of the test set based on the subject identity. As such, the training model is not biased toward the identity of the subject (akin to face recognition instead). Naturally, this protocol also limits the ability of methods to learn the intrinsic micro-expression dynamics of each subject. The intensity and manner of which micro-expressions are shown may differ from person to person, hence compartmentalizing a subject altogether may inhibit the modeling process.

5.4.2. Performance Metrics

Besides the usage of evaluation protocol, the choice of performance metrics is also crucial to understanding the actual performance of automatic ME analysis. Currently, two performance metrics are used most widely: the Accuracy rate and F1-score. While the Accuracy rate is straightforward in calculation, it does not give an adequate reflection of the effectiveness of a classifier as it is susceptible to heavily skewed data (uneven distribution of samples per emotion class), a characteristic found in most current datasets. Also, the Accuracy rate merely shows the average "hit rate" across all classes; thus the performance of the classifier that deals with each emotion class is not revealed. It is a much preferred practice to report confusion matrices for better understanding of its per-class performances. From thereafter, performance metrics such as F1-score, Precision and Recall provide a better measure of a classifier's performance when dealing with imbalanced datasets (Sokolova and Lapalme, 2009; Le Ngo et al., 2014). The overall F1-score, Precision and Recall scores should be micro-averaged based on the total number of true positives, false positives, and false negatives.

5.4.3. Emotion Class

There are several existing works considering different number of emotion classes instead of using the emotion classes provided by the databases. For instance, in the works by Wang et al. (2015a) and Zheng (2017), the authors considered only three or four emotion labels (i.e., Positive, Negative, Surprise, and/or Others) instead of the original emotion labels of the CASME II (i.e., Happiness, Surprise, Disgust, Repression, and Others). Due to the reduction in the number of emotion classes considered, the classification task could be relatively simpler compared to those that have more emotion classes. As a result, higher performances were reported but this also inhibits these works from fair benchmarking against other works on the merit of their methods. It is important to note also that the grouping of classes may be biased toward negative categories since there is only one positive category (Happiness).

Recently, Davison et al. (2017) challenged the current use of emotion classes by proposing the use of *objective classes*, which are determined by restructuring these new categories around the Action Units (AUs) that have been FACS coded. Samples from the two most recent FACS coded datasets, CASME II and SAMM, were re-grouped into these objective classes for their use. The authors argued that emotion classification requires the context of the situation for an interpreter to make a meaningful interpretation, while relying on self-reports (Yan et al., 2014a) can also cause further unpredictability and bias. Although FACS coding can objectively assign AUs to specific muscle movements of the face but the emotion type becomes less obvious. Lim and Goh (2017), through their fuzzy modeling, provided some insights as to why the emotional content in ME samples are non-mutually exclusive as they may contain traces of more than one emotion type.

6. CONCLUSION

Research on the machine analysis of facial MEs has witnessed substantial progress in the last few years as several new spontaneous facial MEs databases were made available to aid automatic analysis of MEs. This has spiked the interest of the affective and visual computing community with a good number of promising methods making headways in both automatic ME spotting and recognition tasks. This necessitates a comprehensive review of recent advances to better taxonomize the increasing number of existing works. In addition, this paper also summarizes the issues that have not received sufficient attention, but are crucial for feasible machine interpretation of MEs. Among the important issues that are yet to be addressed in the field of ME spotting:

- Handling macro movements: Differentiating between larger, macro facial movements such as eye blinks and twitches, for better spotting of the onset of MEs,
- Developing more precise spotting techniques that can cope with various head poses and camera views: Extension of current constrained environments toward more real-time "in-the-wild" settings will provide a major leap in practicality.
- Establishing a firm criteria for defining the onset and offset frames for MEs: This allows ME short sequences to be extracted from long videos, which in turn, can be classified into emotion classes.

For the ME recognition task, there are a few issues that deserve the community's attention:

- Excluding irrelevant facial information: As MEs are very subtle, it is a great challenge to remove other image perturbations caused by face alignment and slight head rotations which may interfere with the MEs.
- Improving feature representations: Encoding subtle movements are difficult even when feature representations are rich. This is due to limitations in the amount of data that is currently available.
- Encouraging cross-database evaluation: Evaluating within single databases often gives a false impression of a method's performance, especially when existing databases lack diversity.

AUTHOR CONTRIBUTIONS

Y-HO and JS compiled and analyzed the works reviewed in this article, ACLN organized the structure of the review, and RCWP provided the critical analysis and necessary proof reading. All authors took part in the writing of the article.

FUNDING

This work is supported in part by Telekom Malaysia Research & Development (TM R&D) Fund-UbeAware, 2beAware and ParaDigm projects, Ministry of Higher Education Grant (FRGS/1/2016/ICT02/MMU/02/2) Malaysia, The Belt and Road Young Scholar Exchange Grant (17510740100), China and Multimedia University, Malaysia.

REFERENCES

Adegun, I. P., and Vadapalli, H. B. (2016). "Automatic recognition of micro-expressions using local binary patterns on three orthogonal planes and extreme learning machine," in *Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)* (Stellenbosch), 2016, 1–5.

Allaert, B., Bilasco, I. M., Djeraba, C., Allaert, B., Mennesson, J., Bilasco, I. M., et al. (2017). "Consistent optical flow maps for full and micro facial expression recognition," in *VISAPP, Proc. of the 12th Int. Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications* (Porto), 235–242.

Asthana, A., Zafeiriou, S., Cheng, S., and Pantic, M. (2013). "Robust discriminative response map fitting with constrained local models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Portland), 3444–3451.

Ben, X., Jia, X., Yan, R., Zhang, X., and Meng, W. (2017). Learning effective binary descriptors for micro-expression recognition transferred by macro-information. *Pattern Recogn. Lett.* 107, 50–58. doi: 10.1016/j.patrec.2017.07.010

Bettadapura, V. (2012). Face expression recognition and analysis: the state of the art. *arXiv preprint arXiv:1203.6722*.

Chaudhry, R., Ravichandran, A., Hager, G., and Vidal, R. (2009). "Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions," in *IEEE Conference on Computer Vision and Pattern Recognition, 2009* (Miami), 1932–1939.

Coan, J. A., and Allen, J. J. (2007). *Handbook of Emotion Elicitation and Assessment*. Oxford University Press.

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297.

Cristinacce, D., and Cootes, T. F. (2006). "Feature detection and tracking with constrained local models," in *BMVC* (Edinburgh).

Davison, A., Lansley, C., Costen, N., Tan, K., and Yap, M. H. (2016a). SAMM: a spontaneous micro-facial movement dataset. *IEEE Trans. Affect. Comput*. 9, 116–129. doi: 10.1109/TAFFC.2016.2573832

Davison, A. K., Lansley, C., Ng, C. C., Tan, K., and Yap, M. H. (2016b). Objective micro-facial movement detection using facs-based regions and baseline evaluation. *arXiv preprint arXiv:1612.05038*.

Davison, A. K., Merghani, W., and Yap, M. H. (2017). Objective classes for micro-facial expression recognition. *arXiv preprint arXiv:1708.07549*.

Davison, A. K., Yap, M. H., and Lansley, C. (2015). "Micro-facial movement detection using individualised baselines and histogram-based descriptors," in *2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (Kowloon), 1864–1869.

Duan, X., Dai, Q., Wang, X., Wang, Y., and Hua, Z. (2016). Recognizing spontaneous micro-expression from eye region. *Neurocomputing* 217, 27–36.

Duque, C., Alata, O., Emonet, R., Legrand, A.-C., and Konik, H. (2018). "Micro-expression spotting using the Riesz pyramid," in *WACV 2018* (Lake Tahoe).

Ekman, P. (2002). *Microexpression Training Tool (METT)*. University of California, San Francisco, CA.

Ekman, P. (2003). *Micro Expression Training Tool (METT) and Subtle Expression Training Tool (SETT)*. San Francisco, CA: Paul Ekman Company.

Ekman, P. (2009a). "Lie catching and microexpressions," in *The Philosophy of Deception*, ed C. Martin (Oxford University Press), 118–133.

Ekman, P. (2009b). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (revised edition)*. WW Norton & Company.

Ekman, P., and Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry* 32, 88–106.

Fasel, B., and Luettin, J. (2003). Automatic facial expression analysis: a survey. *Pattern Recogn.* 36, 259–275. doi: 10.1016/S0031-3203(02)00052-3

Frank, M., Herbasz, M., Sinuk, K., Keller, A., and Nolan, C. (2009a). "I see how you feel: training laypeople and professionals to recognize fleeting emotions," in *The Annual Meeting of the International Communication Association* (New York City, NY: Sheraton New York.

Frank, M. G., Maccario, C. J., and Govindaraju, V. (2009b). "Behavior and security," in *Protecting Airline Passengers in the Age of Terrorism*, eds P. Seidenstat and X. Francis, and F. X. Splane (Santa Barbara, CA: Greenwood Pub Group), 86–106.

Goshtasby, A. (1988). Image registration by local approximation methods. *Image Vis. Comput.* 6, 255–261.

Guo, Y., Tian, Y., Gao, X., and Zhang, X. (2014). "Micro-expression recognition based on local binary patterns from three orthogonal planes and nearest neighbor method," in *International Joint Conference on Neural Networks (IJCNN), 2014* (Beijing), 3473–3479.

Gupta, P., Bhowmick, B., and Pal, A. (2018). "Exploring the feasibility of face video based instantaneous heart-rate for micro-expression spotting," in *Proceeding of IEEE CVPR Workshops* (Salt Lake City), 1316–1323.

Haggard, E. A., and Isaacs, K. S. (1966). "Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy," in *Methods of Research in Psychotherapy*, eds L. A. Gottschalk and H. Auerbach (Boston, MA: Springer), 154–165.

Happy, S., and Routray, A. (2017). Fuzzy histogram of optical flow orientations for micro-expression recognition. *IEEE Trans. Affect. Comput*. doi: 10.1109/TAFFC.2017.2723386

He, J., Hu, J.-F., Lu, X., and Zheng, W.-S. (2017). Multi-task mid-level feature learning for micro-expression recognition. *Pattern Recogn.* 66, 44–52. doi: 10.1016/j.patcog.2016.11.029

Hess, U., and Kleck, R. E. (1990). Differentiating emotion elicited and deliberate emotional facial expressions. *Eur. J. Soc. Psychol.* 20, 369–385.

Horn, B. K., and Schunck, B. G. (1981). Determining optical flow. *Artif. Intell.* 17, 185–203.

Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K. (2006). Extreme learning machine: theory and applications. *Neurocomputing* 70, 489–501. doi: 10.1016/j.neucom.2005.12.126

Huang, X., Wang, S.-J., Liu, X., Zhao, G., Feng, X., and Pietikainen, M. (2017). Discriminative spatiotemporal local binary pattern with revisited integral projection for spontaneous facial micro-expression recognition. *IEEE Trans. Affect. Comput.* doi: 10.1109/TAFFC.2017.2713359

Huang, X., Wang, S.-J., Zhao, G., and Piteikainen, M. (2015). "Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection," in *Proceedings of the IEEE International Conference on Computer Vision Workshops* (Santiago), 1–9.

Huang, X., and Zhao, G. (2017). "Spontaneous facial micro-expression analysis using spatiotemporal local radon-based binary pattern," in *2017 International Conference on The Frontiers and Advances in Data Science (FADS)* (Xi'an), 159–164.

Huang, X., Zhao, G., Hong, X., Zheng, W., and Pietikäinen, M. (2016). Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns. *Neurocomputing* 175, 564–578. doi: 10.1016/j.neucom.2015.10.096

Husak, P., Cech, J., and Matas, J. (2017). Spotting facial micro-expressions "in the wild". In *22nd Computer Vision Winter Workshop* (Retz).

Jia, X., Ben, X., Yuan, H., Kpalma, K., and Meng, W. (2017). Macro-to-micro transformation model for micro-expression recognition. *J. Comput. Sci.* 25, 289–297. doi: 10.1016/j.jocs.2017.03.016

Jovanović, M. R., Schmid, P. J., and Nichols, J. W. (2014). Sparsity-promoting dynamic mode decomposition. *Phys. Fluids* 26:024103. doi: 10.1063/1.4863670

Khor, H.-Q., See, J., Phan, R. C. W., and Lin, W. (2018). "Enriched long-term recurrent convolutional network for facial micro-expression recognition," in *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on* (Xi′an: IEEE), 667–674.

Kim, D. H., Baddar, W. J., and Ro, Y. M. (2016). "Micro-expression recognition with expression-state constrained spatio-temporal feature representations," in *Proceedings of the 2016 ACM on Multimedia Conference* (Amsterdam), 382–386.

Lee, Z.-C., Phan, R. C.-W., Tan, S.-W., and Lee, K.-H. (2017). "Multimodal decomposition for enhanced subtle emotion recognition," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017* (Kuala Lumpur: IEEE), 665–671.

Le Ngo, A. C., Liong, S.-T., See, J., and Phan, R. C.-W. (2015). "Are subtle expressions too sparse to recognize?" in *2015 IEEE International Conference on Digital Signal Processing (DSP)* (Singapore), 1246–1250.

Le Ngo, A. C., Oh, Y.-H., Phan, R. C.-W., and See, J. (2016a). "Eulerian emotion magnification for subtle expression recognition," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Shanghai), 1243–1247.

Le Ngo, A. C., Johnston, A., Phan, R. C.-W., and See, J. (2018). "Micro-expression motion magnification: Global Lagrangian vs. Local Eulerian Approaches," in *Automatic Face & Gesture Recognition (FG 2018) Workshops, 2018 13th IEEE International Conference on* (Xi'an: IEEE), 650–656.

Le Ngo, A. C., Phan, R. C.-W., and See, J. (2014). "Spontaneous subtle expression recognition: imbalanced databases and solutions," in *Asian Conference on Computer Vision* (Singapore: Springer), 33–48.

Le Ngo, A. C., See, J., and Phan, C.-W. R. (2016b). Sparsity in dynamics of spontaneous subtle emotion: analysis & application. *IEEE Trans. Affect. Comput*. doi: 10.1109/TAFFC.2016.2523996

Li, J., Soladie, C., and Seguier, R. (2018). "LTP-ML: micro-expression detection by recognition of local temporal pattern of facial movements," in *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on* (Xi'an: IEEE).

Li, X., Pfister, T., Huang, X., Zhao, G., and Pietikainen, M. (2013). "A spontaneous micro-expression database: inducement, collection and baseline," in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 1–6.

Li, X., Xiaopeng, H., Moilanen, A., Huang, X., Pfister, T., Zhao, G., et al. (2017). Towards reading hidden emotions: a comparative study of spontaneous micro-expression spotting and recognition methods. *IEEE Trans. Affect. Comput*. doi: 10.1109/TAFFC.2017.2667642

Lim, C. H., and Goh, K. M. (2017). "Fuzzy qualitative approach for micro-expression recognition," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017* (Kuala Lumpur: IEEE), 1669–1674.

Liong, S.-T., Phan, R. C.-W., See, J., Oh, Y.-H., and Wong, K. (2014a). "Optical strain based recognition of subtle emotions," in *2014 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)* (Kuching), 180–184.

Liong, S.-T., See, J., Phan, R. C.-W., Le Ngo, A. C., Oh, Y.-H., and Wong, K. (2014b). "Subtle expression recognition using optical strain weighted features," in *Computer Vision-ACCV 2014 Workshops* (Singapore: Springer), 644–657.

Liong, S.-T., See, J., Phan, R. C.-W., Oh, Y.-H., Le Ngo, A. C., Wong, K., et al. (2016a). Spontaneous subtle expression detection and recognition based on facial strain. *Signal Process. Image Commun.* 47, 170–182. doi: 10.1016/j.image.2016.06.004

Liong, S.-T., See, J., Phan, R. C.-W., and Wong, K. (2016b). Less is more: micro-expression recognition from video using apex frame. *arXiv preprint arXiv:1606.01721*.

Liong, S.-T., See, J., Phan, R. C.-W., Wong, K., and Tan, S.-W. (2018). Hybrid facial regions extraction for micro-expression recognition system. *J. Signal Process. Syst.* 90, 601–617. doi: 10.1007/s11265-017-1276-0

Liong, S.-T., See, J., Wong, K., Le Ngo, A. C., Oh, Y.-H., and Phan, R. (2015). "Automatic apex frame spotting in micro-expression database," in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)* (Kuala Lumpur), 665–669.

Liong, S.-T., See, J., Wong, K., and Phan, R. C.-W. (2016c). "Automatic micro-expression recognition from long video using a single spotted apex," in *Asian Conference on Computer Vision (ACCV) Workshops* (Taipei: Springer), 345–360.

Liong, S.-T., and Wong, K. (2017). "Micro-expression recognition using apex frame with phase information," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017* (Kuala Lumpur: IEEE), 534–537.

Liu, Y.-J., Zhang, J.-K., Yan, W.-J., Wang, S.-J., Zhao, G., and Fu, X. (2016). A main directional mean optical flow feature for spontaneous micro-expression recognition. *IEEE Trans. Affect. Comput.* 7, 299–310. doi: 10.1109/TAFFC.2015.2485205

Lu, Y., Zheng, W.-L., Li, B., and Lu, B.-L. (2015). "Combining eye movements and EEG to enhance emotion recognition," in *IJCAI* (Buenos Aires), 1170–1176.

Lu, Z., Luo, Z., Zheng, H., Chen, J., and Li, W. (2014). "A delaunay-based temporal coding model for micro-expression recognition," in *Asian Conference on Computer Vision (ACCV) Workshops* (Singapore: Springer), 698–711.

Ma, H., An, G., Wu, S., and Yang, F. (2017). "A region histogram of oriented optical flow (RHOOF) feature for apex frame spotting in micro-expression," in *2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)* (Xiamen), 281–286.

Megvii, I. (2013). *Face++ Research Toolkit.* Available online at: www.faceplusplus.com

Merghani, W., Davison, A., and Yap, M. (2018). "Facial Micro-expressions Grand Challenge 2018: evaluating spatio-temporal features for classification of objective classes," in *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on* (Xi'an: IEEE), 662–666.

Milborrow, S., and Nicolls, F. (2014). "Active shape models with SIFT descriptors and MARS," in *VISAPP (2)* (Lisbon), 380–387.

Moilanen, A., Zhao, G., and Pietikainen, M. (2014). "Spotting rapid facial movements from videos using appearance-based feature difference analysis," in *2014 22nd International Conference on Pattern Recognition (ICPR)* (Stockholm), 1722–1727.

Oh, Y.-H., Le Ngo, A. C., Phan, R. C.-W., See, J., and Ling, H.-C. (2016). "Intrinsic two-dimensional local structures for micro-expression recognition," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Shanghai), 1851–1855.

Oh, Y.-H., Le Ngo, A. C., See, J., Liong, S.-T., Phan, R. C.-W., and Ling, H.-C. (2015). "Monogenic riesz wavelet representation for micro-expression recognition," in *2015 IEEE International Conference on Digital Signal Processing (DSP)* (Singapore), 1237–1241.

Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* 24, 971–987. doi: 10.1109/TPAMI.2002.1017623

Park, S. Y., Lee, S. H., and Ro, Y. M. (2015). "Subtle facial expression recognition using adaptive magnification of discriminative facial motion," in *Proceedings of the 23rd Annual ACM Conference on Multimedia Conference* (Brisbane), 911–914.

Patel, D., Hong, X., and Zhao, G. (2016). "Selective deep features for micro-expression recognition," in *23rd International Conference on Pattern Recognition (ICPR), 2016* (Cancun), 2258–2263.

Patel, D., Zhao, G., and Pietikäinen, M. (2015). "Spatiotemporal integration of optical flow vectors for micro-expression detection," in *International Conference on Advanced Concepts for Intelligent Vision Systems* (Catania: Springer), 369–380.

Peng, M., Wang, C., Chen, T., Liu, G., and Fu, X. (2017). Dual temporal scale convolutional neural network for micro-expression recognition. *Front. Psychol.* 8:1745. doi: 10.3389/fpsyg.2017.01745

Peng, M., Wu, Z., Zhang, Z., and Chen, T. (2018). "From macro to micro expression recognition: deep learning on small datasets using transfer learning," in *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on* (Xi′an: IEEE), 657–661.

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

Ping, L., Zheng, W., Ziyan, W., Qiang, L., Yuan, Z., Minghai, X., et al. (2016). Micro-expression recognition by regression model and group sparse spatio-temporal feature learning. *IEICE Trans. Inform. Syst.* 99, 1694–1697. doi: 10.1587/transinf.2015EDL8221

Pitt, M. K., and Shephard, N. (1999). Filtering via simulation: auxiliary particle filters. *J. Am. Stat. Assoc.* 94, 590–599.

Polikovsky, S., and Kameda, Y. (2013). Facial micro-expression detection in hi-speed video based on facial action coding system (facs). *IEICE Trans. Inform. Syst.* 96, 81–92. doi: 10.1587/transinf.E96.D.81

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

Porter, S., and Ten Brinke, L. (2008). Reading between the lies identifying concealed and falsified emotions in universal facial expressions. *Psychol. Sci.* 19, 508–514. doi: 10.1111/j.1467-9280.2008.02116.x

Qu, F., Wang, S.-J., Yan, W.-J., Li, H., Wu, S., and Fu, X. (2017). CAS(ME)^2: a database for spontaneous macro-expression and micro-expression spotting and recognition. *IEEE Trans. Affect. Comput.* doi: 10.1109/TAFFC.2017.2654440

Radlak, K., Bozek, M., and Smolka, B. (2015). "Silesian Deception Database: presentation and analysis," in *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection* (Seattle, DC), 29–35.

Ross, A. (2004). *Procrustes Analysis.* Course report, Department of Computer Science and Engineering, University of South Carolina.

Rothwell, J., Bandar, Z., O'Shea, J., and McLean, D. (2006). Silent talker: a new computer-based system for the analysis of facial cues to deception. *Appl. Cogn. Psychol.* 20, 757–777. doi: 10.1002/acp.1204

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). ImageNET Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* 115, 211–252. doi: 10.1007/s11263-015-0816-y

Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., and Pantic, M. (2013). "300 faces in-the-wild challenge: the first facial landmark localization challenge," in *Proceedings of the IEEE International Conference on Computer Vision Workshops* (Sydney), 397–403.

Saragih, J. M., Lucey, S., and Cohn, J. F. (2009). "Face alignment through subspace constrained mean-shifts," in *2009 IEEE 12th International Conference on Computer Vision* (Kyoto), 1034–1041.

Sariyanidi, E., Gunes, H., and Cavallaro, A. (2015). Automatic analysis of facial affect: a survey of registration, representation, and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 1113–1133. doi: 10.1109/TPAMI.2014.2366127

Shreve, M., Brizzi, J., Fefilatyev, S., Luguev, T., Goldgof, D., and Sarkar, S. (2014). Automatic expression spotting in videos. *Image Vis. Comput.* 32, 476–486. doi: 10.1016/j.imavis.2014.04.010

Shreve, M., Godavarthy, S., Goldgof, D., and Sarkar, S. (2011). "Macro-and micro-expression spotting in long videos using spatio-temporal strain," in *2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011)* (Santa Barbara), 51–56.

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

Sokolova, M., and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Inform. Process. Manage.* 45, 427–437. doi: 10.1016/j.ipm.2009.03.002

Song, Y., Morency, L.-P., and Davis, R. (2013). "Learning a sparse codebook of facial and body microexpressions for emotion recognition," in *Proceedings of the 15th ACM on International Conference on Multimodal Interaction* (Sydney: ACM), 237–244.

Tomasi, C., and Kanade, T. (1991). Detection and tracking of point features. *Int. J. Comput. Vis.* 9, 137–154.

Tran, T.-K., Hong, X., and Zhao, G. (2017). "Sliding window based micro-expression spotting: a benchmark," in *Advanced Concepts for Intelligent Vision Systems (ACIVS), 18th International Conference on* (Antwerp: Springer), 542–553.

Vaidya, A. R., Jin, C., and Fellows, L. K. (2014). Eye spy: the predictive value of fixation patterns in detecting subtle and extreme emotions from faces. *Cognition* 133, 443–456. doi: 10.1016/j.cognition.2014.07.004

Valstar, M. F., and Pantic, M. (2012). Fully automatic recognition of the temporal phases of facial actions. *IEEE Trans. Syst. Man Cybern. Part B* 42, 28–43. doi: 10.1109/TSMCB.2011.2163710

Wang, S.-J., Wu, S., Qian, X., Li, J., and Fu, X. (2016a). A main directional maximal difference analysis for spotting facial movements from long-term videos. *Neurocomputing* 230, 382–389. doi: 10.1016/j.neucom.2016.12.034

Wang, S.-J., Yan, W.-J., Sun, T., Zhao, G., and Fu, X. (2016b). Sparse tensor canonical correlation analysis for micro-expression recognition. *Neurocomputing* 214, 218–232.

Wang, S. J., Yan, W. J., Li, X., Zhao, G., Zhou, C.-G., Fu, X., et al. (2015a). Micro-expression recognition using color spaces. *IEEE Trans. Image Process.* 24, 6034–6047. doi: 10.1109/TIP.2015.2496314

Wang, S.-J., Yan, W.-J., Zhao, G., Fu, X., and Zhou, C.-G. (2014a). "Micro-expression recognition using robust principal component analysis and local spatiotemporal directional features," in *Workshop at the European Conference on Computer Vision* (Zurich: Springer), 325–338.

Wang, Y., See, J., Oh, Y.-H., Phan, R. C., Rahulamathavan, Y., Ling, H. C., et al. (2017). "Effective recognition of facial micro-expressions with video motion magnification," in *Multimedia Tools and Applications*. 76, 21665–21690. doi: 10.1007/s11042-016-4079-6

Wang, Y., See, J., Phan, R. C. W., and Oh, Y. H. (2014b). "LBP with Six Intersection Points: reducing redundant information in LBP-TOP for micro-expression recognition," in *Computer Vision–ACCV 2014* (Singapore: Springer), 525–537.

Wang, Y., See, J., Phan, R., and Oh, Y. H. (2015b). Efficient spatio-temporal local binary patterns for spontaneous facial micro-expression recognition. *PLoS ONE* 10:e0124674. doi: 10.1371/journal.pone.0124674

Warren, G., Schertler, E., and Bull, P. (2009). Detecting deception from emotional and unemotional cues. *J. Nonverb. Behav.* 33, 59–69. doi: 10.1007/s10919-008-0057-7

Weinberger, S. (2010). Airport security: intent to deceive? *Nature* 465, 412–415. doi: 10.1038/465412a

Wright, J., Ganesh, A., Rao, S., Peng, Y., and Ma, Y. (2009). "Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization," in *Advances in Neural Information Processing Systems* (Vancouver), 2080–2088.

Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J. V., Durand, F., and Freeman, W. T. (2012). Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph.* 31, 1–8. doi: 10.1145/2185520.2185561

Wu, Q., Shen, X., and Fu, X. (2011). "The machine knows what you are hiding: an automatic micro-expression recognition system," in *Affective Computing and Intelligent Interaction ACII 2011* (Memphis), 152–162.

Xia, Z., Feng, X., Peng, J., Peng, X., and Zhao, G. (2016). Spontaneous micro-expression spotting via geometric deformation modeling. *Comput. Vis. Image Understand.* 147, 87–94. doi: 10.1016/j.cviu.2015.12.006

Xu, F., Zhang, J., and Wang, J. (2017). Microexpression identification and categorization using a facial dynamics map. *IEEE Trans. Affect. Comput.* 8, 254–267. doi: 10.1109/TAFFC.2016.2518162

Yan, W.-J., Wang, S.-J., Chen, Y.-H., Zhao, G., and Fu, X. (2014b). "Quantifying micro-expressions with constraint local model and local binary pattern," in *Workshop at the European Conference on Computer Vision* (Zurich: Springer), 296–305.

Yan, W.-J., and Chen, Y.-H. (2017). Measuring dynamic micro-expressions via feature extraction methods. *J. Comput. Sci.* 25, 318–326. doi: 10.1016/j.jocs.2017.02.012

Yan, W.-J., Li, X., Wang, S.-J., Zhao, G., Liu, Y.-J., Chen, Y.-H., et al. (2014a). CASME II: an improved spontaneous micro-expression database and the baseline evaluation. *PLoS ONE* 9:e86041. doi: 10.1371/journal.pone.0086041

Yan, W.-J., Wu, Q., Liang, J., Chen, Y.-H., and Fu, X. (2013a). How fast are the leaked facial expressions: the duration of micro-expressions. *J. Nonverb. Behav.* 37, 217–230. doi: 10.1007/s10919-013-0159-8

Yan, W.-J., Wu, Q., Liu, Y.-J., Wang, S.-J., and Fu, X. (2013b). "Casme database: a dataset of spontaneous micro-expressions collected from neutralized faces," in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (Shanghai), 1–7.

Yang, J., Zhang, L., Xu, Y., and Yang, J.-Y. (2012). Beyond sparsity: the role of l 1-optimizer in pattern classification. *Pattern Recogn.* 45, 1104–1118.

Yap, M. H., See, J., Hong, X., and Wang, S.-J. (2018). "Facial Micro-Expressions Grand Challenge 2018 Summary," *in Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on* (Xi'an: IEEE), 675–678.

Zarezadeh, E., and Rezaeian, M. (2016). Micro expression recognition using the eulerian video magnification method. *Brain* 7, 43–54. Available online at: http://www.brain.edusoft.ro/index.php/brain/article/view/623

Zeng, Z., Pantic, M., Roisman, G. I., and Huang, T. S. (2009). A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* 31, 39–58. doi: 10.1109/TPAMI.2008.52

Zhang, J., Shan, S., Kan, M., and Chen, X. (2014). "Coarse-to-fine auto-encoder networks (cfan) for real-time face alignment," in *European Conference on Computer Vision* (Zurich: Springer), 1–16.

Zhang, S., Feng, B., Chen, Z., and Huang, X. (2017). "Micro-expression recognition by aggregating local spatio-temporal patterns," in *International Conference on Multimedia Modeling* (Reykjavik: Springer), 638–648.

Zhao, Y., Wang, X., and Petriu, E. M. (2011). "Facial expression anlysis using eye gaze information," in *2011 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA)*, 1–4. IEEE.

Zheng, H. (2017). "Micro-expression recognition based on 2d gabor filter and sparse representation," in *Journal of Physics: Conference Series Vol. 787* (IOP Publishing).

Zheng, H., Geng, X., and Yang, Z. (2016). "A relaxed K-SVD algorithm for spontaneous micro-expression recognition," in *Pacific Rim International Conference on Artificial Intelligence* (Phuket: Springer), 692–699.

Zhou, Z., Zhao, G., and Pietikäinen, M. (2011). "Towards a practical lipreading system," in *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Colorado Springs), 137–144.

Zhu, X., Ben, X., Liu, S., Yan, R., and Meng, W. (2018). Coupled source domain targetized with updating tag vectors for micro-expression recognition. *Multimedia Tools Appl.* 77, 3105–3124. doi: 10.1007/s11042-017-4943-z

Zitova, B., and Flusser, J. (2003). Image registration methods: a survey. *Image Vis. Comput.* 21, 977–1000. doi: 10.1016/S0262-8856(03)00137-9

Zong, Y., Huang, X., Zheng, W., Cui, Z., and Zhao, G. (2017). "Learning a target sample re-generator for cross-database micro-expression recognition," in *Proceedings of the 2017 ACM on Multimedia Conference* (Mountain View, CA), 872–880.

Zong, Y., Huang, X., Zheng, W., Cui, Z., and Zhao, G. (2018a). Learning from hierarchical spatiotemporal descriptors for micro-expression recognition. *IEEE Trans. Multimed*. doi: 10.1109/TMM.2018.2820321

Zong, Y., Zheng, W., Huang, X., Shi, J., Cui, Z., and Zhao, G. (2018b). Domain regeneration for cross-database micro-expression recognition. *IEEE Trans. Image Process*. 27, 2484–2498. doi: 10.1109/TIP.2018.2797479

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer XL and handling Editor declared their shared affiliation.

OPINION
published: 07 August 2019
doi: 10.3389/fpsyg.2019.01833

Automatic Micro-Expression Analysis: Open Challenges

Guoying Zhao[1,2]* *and Xiaobai Li*[2]

[1] *School of Information and Technology, Northwest University, Xi'an, China,* [2] *Center for Machine Vision and Signal Analysis, University of Oulu, Oulu, Finland*

Keywords: micro-expression, facial expression, automatic analysis, computer vision, machine learning

OPEN ACCESS

Edited by:
Carl Senior,
Aston University, United Kingdom

Reviewed by:
Amy Dawel,
Australian National University, Australia

Correspondence:
Guoying Zhao
guoying.zhao@oulu.fi

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *06 May 2019*
Accepted: *24 July 2019*
Published: *07 August 2019*

Citation:
Zhao G and Li X (2019) Automatic Micro-Expression Analysis: Open Challenges. Front. Psychol. 10:1833. doi: 10.3389/fpsyg.2019.01833

Micro-expressions, the fleeting and involuntary facial expression, often occurring in high-stake situations when people try to conceal or mask their true feelings, became well-known since 1960s, from the work of Haggard and Isaacs (1966) in which micro-expression was firstly termed as micromomentary facial expressions, and later from the work of Ekman and Friesen (1969).

Micro-expressions are too short (1/25 to 1/2 s) and subtle for human eyes to perceive. Study (Ekman, 2002) shows that for micro-expression recognition tasks, ordinary people without training only perform slightly better than chance on average. So computer vision and machine learning methods for automatic micro-expression analysis become appealing. Pfister et al. (2011) started pioneering research on spontaneous micro-expression recognition with the first publically available spontaneous micro-expression dataset: SMIC, and achieved very promising results that compare favorably with the human accuracy. Since then micro-expression study in computer vision field has been attracting attentions from more and more researchers. A number of works have been contributing to the automatic micro-expression analysis from the aspects of new datasets collection (from emotion level annotation to action unit level annotation; Li et al., 2013; Davison et al., 2018), micro-expression recognition (from signal apex frame recognition to whole video recognition; Wang et al., 2015; Liu et al., 2016; Li Y. et al., 2018; Huang et al., 2019) and micro-expression detection (from micro-expression peak detection to micro-expression onset and offset detection; Patel et al., 2015; Xia et al., 2016; Jain et al., 2018). First completed system integrating micro-expression recognition and detection toward reading hidden emotions (Li X. et al., 2018) has been reported by MIT Technology Review (2015) and achieved increasing attention, in which the machine learning method obtained 80.28% for three class (positive/negative/surprise) recognition for 71 micro-expression video clips recorded from eight subjects and 57.49% for five class (happiness, disgust, surprise, repression, and other) recognition for 247 micro-expression video clips recorded from 26 subjects (Li X. et al., 2018), which has outperformed the recognition capability of human subjects (Li X. et al., 2018).

However, there are still many open challenges which need to be considered in the future research. Several main challenges related with micro-expression study are discussed in details in the following.

DATASETS

Data are a central part in micro-expression research. Even though there have been more datasets collected and released, from the first SMIC (Li et al., 2013), to CASME (Yan et al., 2013), CASME II (Yan et al., 2014), SAMM (Davison et al., 2018), MEVIEW dataset (Husak et al., 2017), and CAS(ME)^2 (Qu et al., 2018), including more subjects, higher resolution, and more videos, the scale of current datasets is just hundreds of micro-expression videos captured from 30 to 40 subjects, and there still lacks high quality, naturally collected and well-annotated large scale micro-expression data captured by different sensors for training efficient deep learning methods, which is a big obstacle for the research. As inducing and labeling micro-expression data from scratch is extremely challenging and time consuming, it is not feasible for any single research group to gather data scale of larger than tens of thousands of samples. One possible option for future micro-expression data construction work could be utilizing the vast source of YouTube videos and mining with some video tagging techniques for candidate clips then follow with human labeling. Another option could be collaborative and parallel data collection and labeling through cloud sourcing.

Moreover, one potential application of micro-expression analysis is lie detection. When lying, more contradictory behaviors could be found in verbal and non-verbal signals (Navarro and Karlins, 2008), perhaps more micro-expressions could appear. Therefore, new datasets containing not only facial expression and micro-expression, but also audio speech could be beneficial for micro-expression study.

ACTION UNITS DETECTION OF MICRO-EXPRESSIONS

Facial Action Coding System (FACS) is an anatomically based system for measuring facial movements (Ekman and Friesen, 1978), which is used to describe visually distinguishable facial activity on the basis of many unique action units (AUs). In most of the previous work (Wang et al., 2015; Li X. et al., 2018), micro-expressions were recognized from the whole face without action unit study, and only positive and negative micro-expressions, or limited number of micro-expressions were classified. Instead of directly recognizing a certain number of prototypical expressions as in most of the previous research, AUs can provide an intermediate meaningful abstraction of facial expressions, and carry lots of information which can help better detect and understand people's feelings. Even though AU detection has been taken into consideration for macro-expression analysis (Zhao et al., 2016, 2018; Han et al., 2018; Zhang et al., 2018), including pain detection and pain intensity estimation (Prkachin and Solomon, 2008; Lucey et al., 2011), rare work has been done for AUs in micro-expressions. Future study could pay more attention to explore the relationship between AUs and micro-expressions. For example: is there fixed mapping between the onset of a certain AU (or a sequence of AU combinations) and one micro-expression category, just like the criteria for AU and facial expression correspondence listed in FACS manual? The category of concerned micro-expression emotions is not necessarily limited to the prototypical basic emotions, i.e., happiness, sadness, surprise, anger, disgust and fear, but could also consider other emotions which are out of the above mentioned basic emotion scope, yet very useful for real-world applications, like nervousness, disagreement and contempt. Besides, except those most common emotional AUs (that are considered to be closely related with emotional expressions), e.g., AU1, AU4, and AU12, other AUs which were formally considered as "irrelevant to emotions" also worth more exploration, as studies found that some (e.g., eye blinks and eye gaze change) are employed as disguise behaviors to cover true feelings thus frequently occur WITH the onset of micro-expressions.

REALISTIC SITUATIONS

Most of the existing efforts on micro-expression analysis have been made to classify the basic micro-expressions collected in highly controlled environments, e.g., from frontal view (without view changes), with stable and bright lighting conditions (without illumination variations), whole face visible (without occlusion). Such conditions are very difficult to reproduce in real-world applications and tools trained on such data usually do not generalize well to natural recordings made in unconstrained settings. Effective algorithms for recognizing naturally occurring micro-expressions which are robust to realistic situations with the capability to deal with pose changes, illumination variations and poor quality of videos, recorded in-the-wild environment must be developed.

MACRO- AND MICRO- EXPRESSIONS

Previous work about facial expression has concerned with either micro- or macro-expressions. For most early micro-expression works, it has been assumed that there are just micro-expressions in a video clip. For example, in the collection of most micro-expression datasets (Li et al., 2013; Yan et al., 2013, 2014; Davison et al., 2018; Qu et al., 2018), subjects were asked to try their best to keep a neutral face when watching emotional movie clips. In this way, the conflict of felt emotion elicited by the movie clip and the strong intention to suppress any facial expression could induce micro-expressions. The consequence in the collected videos is that, if there is micro-expression in the recorded video, it is unlikely to have other natural facial expressions. But in most cases in real life, this is not true. Micro-expressions can appear when there is a macro-expression as well, for example, when people smile, they might furrow forehead very quickly and shortly, which show their true feeling (Ekman and Friesen, 1969). Future studies could also concern the relationship of macro and micro-expressions, and explore methods that can detect and distinguish these two when they co-occur or even overlap with each other in one scenario, which would be very helpful to understand people's feelings and intentions more accurately.

CONTEXT CLUES AND MULTI-MODALITY LEARNING

In social interactions, people interpret other's emotions and situations based on many things (Huang et al., 2018): people in the interaction, their speech, facial expression, cloths, body pose, gender, age, surrounding environments, social parameters, and so on. All these can be considered as contextual information. Some people are better emotion readers, as they can sense others' emotion more accurately than the rest. These people usually pick up subtle clues from multiple aspects, not only the facial expressions (Navarro and Karlins, 2008). One original motivation for the study of micro-expression is to explore people's suppressed and concealed emotions, but we shouldn't forget that micro-expression is only one of the many clues for such purpose. Future studies should try broaden the scope and consider combining micro-expression with other contextual behaviors, e.g., eye blink, eye gaze change, hand gesture change, or even whole body posture, in order to achieve better understanding of people's hidden emotions on a fuller scope.

Recent psychological research demonstrates that emotions are a multimodal procedure which can be expressed in various ways. "Visual scenes, voices, bodies, other faces, cultural orientation, and even words shape how emotions is perceived in a face" (Barrett et al., 2011). As well emotional data can be recorded with different sensors, e.g., color camera, near infrared camera, depth camera, or physiological sensors, for recording emotional behaviors or bodily changes. This also applies to the study of micro-expression and suppressed or hidden emotion. One single modality could be unreliable, as one certain behavior pattern could be just related to physiological uncomfort or personal habit, but has nothing to do with emotional states. So only when multiple cues are considered together we could achieve more reliable emotion recognition. There is very little investigation in this respect so far, and future micro-expression studies could consider combining multi-modality data for micro-expression and hidden emotion recognition.

ANALYSIS FOR MULTIPLE PERSONS IN INTERACTIONS

The current micro-expressions research focuses on single person watching affective movies or advertisements, which is reasonable in the early stage for making challenging tasks easier and more feasible. Later it is surely that the research will be shifting toward more realistic and challenging interaction environments where multiple persons are involved. Natural interactions will induce more natural and spontaneous emotional responses in terms of facial expressions and micro-expressions, but the scenario will also become very complicated. It would be very interesting to explore not only individual level of emotional changes, but also the interpersonal co-occurrence (e.g., mimicry or contagion), and the affective dynamics of the whole group.

DISCUSSION

We have discussed the progress and the open challenges in automatic micro-expression analysis. Solving these issues needs interdisciplinary expertise. The collaboration of machine learning, psychology, cognition and social behavior is necessary for advancing the in-depth investigation of micro-expressions and related applications in real world.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was partially supported by the National Natural Science Foundation of China (Grants No. 61772419), Infotech Oulu and Academy of Finland (ICT 2023 project with grant No. 313600).

REFERENCES

Barrett, L. F., Mesquita, B., and Gendron, M. (2011). Context in emotion perception. *Curr. Dir. Psychol. Sci.* 20, 286–290. doi: 10.1177/0963721411422522

Davison, A. K., Lansley, C., Costen, N., Tan, K., and Yap, M. H. (2018). SAMM: a spontaneous micro-facial movement dataset. *IEEE Trans. Affect. Comput.* 9, 116–129. doi: 10.1109/TAFFC.2016.2573832

Ekman, P. (2002). *Microexpression Training Tool (METT)*. San Francisco, CA: University California.

Ekman, P., and Friesen, W. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement Consulting.* Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., and Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry* 32, 88–106.

Haggard, E., and Isaacs, K. (1966). "Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy," in *Methods of Research in Psychotherapy,* eds L. A. Gottschalk and A. H. Auerbach (New York, NY: Appleton-Century-Crofts), 154–165.

Han, S., Meng, Z., Li, Z., Reilly, J., Cai, J., Wang, X., et al. (2018). "Optimizing filter size in convolutional neural networks for facial action unit recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 5070–5078.

Huang, X., Dhall, A., Goecke, R., Pietikäinen, M., and Zhao, G. (2018). Multimodal framework for analyzing the affect of a group of people. *IEEE Trans. Multimedia* 20, 2706–2721. doi: 10.1109/TMM.2018.2818015

Huang, X., Wang, S.-J., Liu, X., Zhao, G., Feng, X., and Pietikäinen, M. (2019). Discriminative spatiotemporal local binary pattern with revisited integral projection for spontaneous facial micro-expression recognition. *IEEE Trans. Affect. Comput.* 10, 32–47. doi: 10.1109/TAFFC.2017.2713359

Husak, P., Cech, J., and Matas, J. (2017). "Spotting facial micro-expressions "In the Wild"," in *Proceedings of the 22nd Computer Vision Winter Workshop*, eds N. M. Artner, I. Janusch, and W. G. Kropatsch (Retz).

Jain, D. K., Zhang, Z., and Huang, K. (2018). Random walk-based feature learning for micro-expression recognition. *Pattern Recogn. Lett.* 115, 92–100. doi: 10.1016/j.patrec.2018.02.004

Li, X., Hong, X., Moilanen, A., Huang, X., Pfister, T., Zhao, G., et al. (2018). Towards reading hidden emotions: a comparative study of spontaneous micro-expression spotting and recognition methods. *IEEE Trans. Affect. Comput.* 9, 563–577. doi: 10.1109/TAFFC.2017.2667642

Li, X., Pfister, T., Huang, X., Zhao, G., and Pietikäinen, M. (2013). "A spontaneous micro facial expression database: inducement, collection and baseline," in *Proceedings of the IEEE International Conference on Face and Gesture Recognition* (Shanghai: FG 2013).

Li, Y., Huang, X., and Zhao, G. (2018). "Can micro-expression be recognized based on single apex frame?" in *International Conference on Image Processing* (Athens: ICIP).

Liu, Y.-J., Zhang, J.-K., Yan, W.-J., Wang, S.-J., Zhao, G., and Fu, X. (2016). A main directional mean optical flow feature for spontaneous micro-expression recognition. *IEEE Trans. Affect. Comput.* 7, 299–310. doi: 10.1109/TAFFC.2015.2485205

Lucey, P., Cohn, J. F., Prkachin, K. M., Solomon, P. E., and Matthews, I. (2011). "Painful data: the unbc-mcmaster shoulder pain expression archive database," in *Proceedings of the IEEE International Conference on Face and Gesture Recognition* (Santa Barbara, CA: FG 2011).

Navarro, J., and Karlins, M. (2008). *What Every BODY Is Saying: An ex-FBI Agent's Guide to Speed Reading People*. New York, NY: Collins.

Patel, D., Zhao, G., and Pietikäinen, M. (2015). "Spatiotemporal integration of optical flow vectors for micro-expression detection," in *Proceedings of the International Conference on Advanced Concepts for Intelligent Vision Systems* (Catania: ACIVS).

Pfister, T., Li, X., Zhao, G., and Pietikainen, M. (2011). "Recognising spontaneous facial micro-expressions," in *Proceedings of the IEEE International Conference on Computer Vision* 2011 (Barcelona).

Prkachin, K. M., and Solomon, P. E. (2008). The structure, reliability and validity of pain expression: evidence from patients with shoulder pain. *Pain* 139, 267–274. doi: 10.1016/j.pain.2008.04.010

Qu, F., Wang, S.-J., Yan, W.-J., Li, H., Wu, S., and Fu, X. (2018). CAS(ME)^2: A database for spontaneous macro-expression and micro-expression spotting and recognition. *IEEE Trans. Affect. Comput*. 9, 424–436. doi: 10.1109/TAFFC.2017.2654440

Wang, S.-J., Yan, W.-J., Li, X., Zhao, G., Zhou, C.-G., Fu, X., et al. (2015). Micro-expression recognition using color spaces. *IEEE Trans. Image Process.* 24, 6034–6047. doi: 10.1109/TIP.2015.2496314

Xia, Z., Feng, X., Peng, J., Peng, X., and Zhao, G. (2016). Spontaneous micro-expression spotting via geometric deformation modeling. *Comput. Vision Image Understand.* 147, 87–94. doi: 10.1016/j.cviu.2015.12.006

Yan, W.-J., Li, X., Wang, S.-J., Zhao, G., Liu, Y.-J., Chen, Y.-H., et al. (2014). CASME II: An improved spontaneous micro-expression database and the baseline evaluation. *PLoS ONE* 9:e86041.

Yan, W.-J., Wu, Q., Liu, Y.-J., Wang, S.-J., and Fu, X. (2013). "CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces," in *Proceedings of the IEEE International Conference Automatic Face and Gesture Recognition 2013* (Shanghai).

Zhang, Y., Dong, W., Hu, B.-G., and Ji, Q. (2018). "Classifier learning with prior probabilities for facial action unit recognitionm," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 5108–5116.

Zhao, K., Chu, W.-S., and Martinez, A. M. (2018). Learning facial action units from web images with scalable weakly supervised clustering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 2090–2099.

Zhao, K., Chu, W.-S., and Zhang, H. (2016). "Deep region and multi-label learning for facial action unit detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV), 3391–3399.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 23 May 2018
doi: 10.3389/fpsyg.2018.00772

Fairness and Smiling Mediate the Effects of Openness on Perceived Fairness: Beside Perceived Intention

Zhifang He[1,2], Jianping Liu[1], Zhiming Rao[3] and Lili Wan[2]*

[1] School of Psychology, Jiangxi Normal University, Nanchang, China, [2] School of Humanities, Jiangxi University of Traditional Chinese Medicine, Nanchang, China, [3] School of Physics and Communication Electronics, Jiangxi Normal University, Nanchang, China

Previous studies have shown that smiling, fairness, intention, and the results being openness to the proposer can influence the responses in ultimatum games, respectively. But it is not clear that how the four factors might interact with each other in twos or in threes or in fours. This study examined the way that how the four factors work in resource distribution games by testing the differences between average rejection rates in different treatments. Two hundred and twenty healthy volunteers participated in an intentional version of the ultimatum game (UG). The experiment used a 2 × 2 × 2 × 2 mixed design with "openness" as a between subjects factor and the other three as within subjects factors, and the participants were assigned as recipients. The results revealed that fairness or perceived good intention reduced the subject's average rejection rates. There was a significant interaction between facial expressions and openness. With fair offers, the average rejection rate for informed was lower than that of uninformed; but when unfair, no difference between the corresponding average rejection rates was found. The interaction effect of smiling and openness was also significant, the average rejection rate for informed offers was lower when the proposer was smiling and no rejection rate difference between uninformed offers and informed offers when no smiling. No other interaction effect was found.

Keywords: smiling, openness, perceived intention, fairness, ultimatum game

OPEN ACCESS

Edited by:
Guoying Zhao,
University of Oulu, Finland

Reviewed by:
Chao Liu,
Beijing Normal University, China
Xiaobai Li,
University of Oulu, Finland

****Correspondence:***
Jianping Liu
liujianping@jxnu.edu.cn

Specialty section:
This article was submitted to
Emotion Science,
a section of the journal
Frontiers in Psychology

Received: *01 July 2017*
Accepted: *01 May 2018*
Published: *23 May 2018*

Citation:
He Z, Liu J, Rao Z and Wan L (2018)
Fairness and Smiling Mediate
the Effects of Openness on Perceived
Fairness: Beside Perceived Intention.
Front. Psychol. 9:772.
doi: 10.3389/fpsyg.2018.00772

INTRODUCTION

Fairness is of utmost importance in social life, as well as in political and economic life (Carl et al., 2006). Defined as the phenomenon of inequity aversion, violation of the social norm of fairness can elicit negative emotions (Stouten et al., 2011) and give rise to subsequent strong reactions, including punishment (Mendoza et al., 2014) or even personal revenge for unequal distributions of resources. Different theories were developed to explain why some people feel more fairness than others as they are facing the same distribution. Utility theory was first proposed with the rationality hypothesis, suggesting that when faced with resource distribution, people tend to make choices with greater utility (Fishburn, 1967). Later, implicit expected utility theory was proposed with an implicit economic cognition hypothesis, which takes effects into account in the decision process model (Raaij and Ye, 2002). However, utility is not the only thing that people consider when making decisions. Studies on belief in a just world (Lerner, 1965), defensive attributions (Shaver, 1970), retributive justice (Darley and Pittman, 2003), criminal responsibility

(Gebotys and Dasgupta, 1987), and moral psychology (Gray and Wegner, 2010; Haidt and Kesebir, 2010; Knobe et al., 2012) all converge to show that when people detect harm, they become motivated to blame someone for that harm. It has been found that a receiver's perception of the intention of a distributor affects the receiver's sense of fairness (Güroglu et al., 2011) and that perceived good intention alleviates the sense of unfairness (Ma et al., 2015b). Numerous behavioral and neuroscientific experiments have demonstrated that intentional harms make people want to blame, condemn, and punish more than unintentional harms do (Alicke, 1992; Darley and Pittman, 2003; Young and Saxe, 2009). People are notoriously sensitive to harmful intentions (Gollwitzer et al., 2009), and even exposure to fictional characters with harmful intentions can change subsequent trust behavior in real life (Rothmund et al., 2011). Intention plays an important role and might lead to sequential reciprocity (Dufwenberg and Kirchsteiger, 1998). However, because people cannot observe others' intention, intention is only perceived. Here in this paper, we use the term "perceived intention." The conclusions of perceived intention are diverse. One study showed that perceived intention was consistent with the reciprocity hypothesis (McCabe et al., 2003), which overthrew the previous conclusion that perceived intention was closely related to the experimental results, that is, the sum of money gained by the subjects. Another experiment showed that certain outcomes, along with intentions and motivations, account for reciprocity (Stanca et al., 2009).

Facial expressions are informative and expressive in social interactions, and they help the receivers reason, judge and make decisions during social interactions, and have a function in social interaction. Smiling expressions were found to reduce the perceived anger (Bugental, 1974), and different smiling models might lead to different reactions (Krumhuber et al., 2009). Smiling offers were more likely to be accepted (Mussel et al., 2013). As for fairness, facial expressions impact the decision making concerning fairness (Mussel et al., 2014). In real life, the emotional state of a distributor may affect the allocation of resources, and the perceived emotions of a distributor will also have an impact on the fairness perceived by the recipient. In face-to-face communication, the recognition of facial expressions is an important way to judge the emotion of the two sides, and it is also an important social cue that affects the psychological process of the communicator. A smiling expression might facilitate trust (van't Wout and Sanfey, 2008) and lead to cooperation. One's emotion may play a part in perceived fairness (Heussler et al., 2009); therefore, the reason that why one's partner's emotion might influence one's own emotion and thus affects perceived fairness seems logical.

Fairness evolving during resource distribution is linked to reputation, which concerns proposer's knowledge of responder's deal (Nowak et al., 2000). When it comes to social affairs, or public goods, information is of the most importance. If the proposers will be notified of what responders have done and the responders know it, an education will happen to teach the proposers a lesson (Abbink et al., 2004), and in time fairness will finally be done. Though people could deduce others' intention and emotion from their expression, they can't predict the corresponding behaviors. So during UG time, if proposers know these behaviors, the following distributions may be different. And if the responders know what they have done will get to the proposers, they may act another way. But whether or how the effect of openness will be affected by perceived intention, fairness, and expression on perceived fairness remains unknown. We regard that openness may urge the responders to show their moral courage and to make decisions more for public goods, and Chinese traditional culture such as "be wordly wise and play safely" may also take its place. The study of openness in perceived fairness is relatively fewer compared with intention, fairness or emotion. Whether the openness in resource distribution would be counterbalanced by the Chinese traditional culture remains unknown, so our principle concern in this paper is openness.

We also wonder that if openness meets obviously unfair in resource distribution, what would happen? And still, what it would be if openness meets a smiling face? Does the Chinese saying "Don't be angry to the person in smiles" still works in a resource distribution experiment? And as "Don't lose face" has extraordinary personal meaning and "Do boldly what is righteous" is of important social meaning in China, we wonder if an unfair offer together with openness and a smiling face would affect the responders' decisions.

The ultimatum game (UG) (Güth et al., 1982), measures decision-making in a resource distribution context. A classic UG has two roles, a proposer, a responder, and a certain amount of stake. The proposer receives the stake and has to make an arbitrary offer to share with the responder. The responder decides to accept or to reject the offer. If the offer is accepted, both of them receive payment as the offer requires, if it is rejected, both receive nothing. For example, a proposer divides $10 among himself and a responder, then the responder decides whether to reject the proposal so that neither player receives anything, or to accept the offer, so that each player gets her/his money according to the division. During the UG, the proposer decides the distribution of the stake, and the responder decides whether the offer works. UG concerns about resource distribution, social comparison and people's decision making, so we can say that the experimental paradigm is logically suit for the purpose of perceived fairness study.

Widely used to examine people's responses to unfairness, the UG is often modified for the purpose of different experiments. In this study, the variation in the ultimatum game was used to investigate the effects, especially the interact effects of fairness, perceived intention, smiling and the openness of a responder's responses on perceived fairness. For each participant, a certain amount of money was divided between a proposer and a responder (Güth et al., 1983). We made our experiment different from the common paradigm of the ultimatum game in that each time, two possible divisions were present. The proposer decided how to divide the money, and the responder decided whether to accept or reject the offer. We aimed to test whether the effect of the openness of the responder's decision on perceived fairness was moderated by the facial expression of the distributor (the proposer) and/or the fairness of the distribution. Perceived fairness was measured by participants' rejection rates of the distribution. The hypotheses are: (1) Fairness promotes the perceived fairness. This would be manifested by the lower rejection rate for fairness. (2) Perceived good intention reduces the perceived fairness. This would supported by a corresponding lower rejection rate for perceived good intention. (3) Fairness moderates the effect of openness. Evidence would come from that the difference between the rejection rates for fair informed vs. fair uninformed distributions is different from that of the rejection rates for unfair informed vs. unfair uninformed distributions. (4) Smiling moderates the effect of openness, it will be proved by that the difference between the rejection rates for smiling informed vs. smiling uninformed distributions is different from that of the rejection rates for no-smiling informed vs. no-smiling uninformed distributions.

MATERIALS AND METHODS

Participants

To get adequate power of statistics (above 0.8), we used G*Power 3 software (Blue et al., 2016) and it suggested a size of no less of 199 for this study to get a medium-size effect (f = 0.20). 260 healthy volunteers (undergraduates) were recruited from two universities in Nanchang, none of whom were from psychology or social disciplines. We made it clear during the recruiting that only those who had never taken part in experiments involving UG were qualified. We excluded 40 participants' data after the UG experiment because they failed the trust check for their disbelief in the truth of the experiment. Thus, the final sample included 220 students (109 females) aged 18–25 (mean age = 21.5, SD = 1.6). The experiment was conducted in accordance with the Declaration of Xiaoman Yan and was approved by the Ethics Committee of Jiangxi University of Traditional Chinese Medicine. We collected informed written consent from every participant prior to the experiment.

MATERIALS

Experimental Design and Procedure

Participants were divided into groups of 10–15 persons. On arrival, participants were told that they would play a money distribution game with partners online. They were also told that all players would be anonymous and that a blurry facial expression image would be assigned to the player. Each time, an assistant guided a group of participants to the psychological laboratory, and they were notified that they were specified randomly as the recipients. Every subject was seated in front of a screen, which was 100 cm in front of him. The stimulus was presented at the center of the screen, and the visual angle was about 8° × 7°. Half of the subjects were instructed to use "J" for "agree" and "F" for "reject," and the rest were the versus. When one finished her/his task, reward would be paid.

Design

The experiment had a 2 × 2 × 2 × 2 mixed design. The first factor, facial expression, had two levels, smiling vs. no-smiling, which was conveyed by a facial image on the screen. The number of images was balanced in terms of the sex and emotion of the proposers. No image was repeated during one participant's experiment. The second factor was the fairness of the distribution, fair vs. unfair, which was determined by the distribution rate. For example, the rate could be 6:4 (the proposer took six yuan out of 10 yuan), a relatively fair distribution, or 8:2, a rather unfair one. Other rates are shown in table one. The third factor was the proposer's perceived intention, good intention vs. bad intention, which was conveyed through the proposer's choice. The proposer made a choice between two rates, and if the proposer chose the option to maximize his/her own profit, the subject sensed bad intentions. For example, for 5:5 vs. 6:4, if the proposer chose 6:4 (thus receiving 6 out of 10 yuan), the recipient perceived bad intentions because the proposer did not choose a less selfish distribution. If the proposer chose 5:5, then the proposer received less and the recipient received more than if the proposer chose 6:4. Thus, the recipient perceived good intentions. The fourth, the only between factor, was the openness of the responder's decision, informed vs. uniformed. The subjects were randomly assigned to an informed group or an uninformed group. For more details on the stimulus design, see **Table 1**.

The present experiment was a modified mini UG paradigm (Falk et al., 2003). We made it different from the mini UG that different unfair distributions were present, and the unfair alternative rates of 9:1 vs. 10:0 were more extreme. The computer presented the distribution rates randomly. Each distribution was a pair of rates in **Table 1** and was presented the same number of times. In each pair, the rates were chosen an equal number of times. Therefore, the target rates were presented twice for the corresponding masking ones. The whole experiment consisted of 16 blocks, and each block included 10 trials. One hundred sixty emotional images (80 of them are smiling, half of them are female) were used, and no faces were of the same person. As the proposer's facial attractiveness matters during the UG (Ma et al., 2015a), we balanced the attractiveness by a procedure that let the attractiveness assessed on LAN scoring from 0 to 10 before the experiment. The assessors were freshmen and no one would join the experiment. Each picture was scored by a group of assessors, ten males and ten females. Each group assessed twenty pictures (the number of smiling females was 5, for the sake of

TABLE 1 | The stimulus design.

Facial expression	Fairness:Paired rates	perceived intention	trials
Smiling	5:5 vs. 6:4 fair	Bad	20
Smiling	6:4 vs.7:3 fair	Good	20
Smiling	8:2 vs. 9:1 unfair	Bad	20
Smiling	9:1 vs 10:0 unfair	Good	20
No-smiling	5:5 vs. 6:4 fair	Bad	20
No-smiling	6:4 vs.7:3 fair	Good	20
No-smiling	8:2 vs. 9:1 unfair	Bad	20
No-smiling	9:1 vs. 10:0 unfair	Good	20

If chosen, the underlined numbers are targets, which produce the data for analysis, and the corresponding ones, if chosen, are masks. In x:y, x is the money given to the provider, and y is the subject.

balance). We hand-picked 160 pictures scored between 5 to 8 out of 500 pictures. AOV of the scores showed no difference, $F(3,159) = 1.745$, $p > 0.05$.

PROCEDURE

The participants were told the rules of the UG. As shown in **Figure 1**, the fixation point (a red "+") shown for 500 ms at the center of the black screen indicated that the stimulus would soon be presented. Then, the proposer's facial image was displayed on the screen for 1500 ms. Next, the distribution was shown on the screen for 1500 ms. A blank screen was shown for 800 ms, meaning that the proposer was thinking, and the distribution showed up again, with the numbers colored in the bold frame as the proposer's choice. The participant pressed "F" to reject or press "J" to accept the offer. If the distribution was rejected, both the proposer and the receipt received nothing, and if it was accepted, each received the money, distributed as the proposer decided. The feedback was on the screen for 800 ms. Every participant performed four practice trials to become familiar with the experimental procedure before the formal experiment began. When the experiment was over, each participant completed a form to check whether he/she believed it was a real bargain. Each participant received 30 yuan (about 4.4 USD) for attendance, and extra decision-based payment was decided by two randomly selected of the participant's trials. On leaving, the amount was calculated and the participant was paid on the spot. The whole process was programmed with E-prime 2.0 software.

RESULTS

The rejection rates under different conditions are shown in **Figure 2**.

We performed a 2 (facial expression:smiling vs. non-smiling) × 2 (fairness:fair vs. unfair) × 2 (openness:informed vs. uninformed) × 2 (perceived intention:good vs. bad) repeated ANOVA on subjects' rejection rates for different offers in UG. The analysis revealed a significant main effect of fairness, $F(1,218) = 118.771$, $p < 0.001$, partial $\eta^2 = 0.144$, with the rejection rate for fair offers (0.2470 ± 0.0162, CI = [0.2788, 0.2152]) lower than the one for unfair offers (0.5011 ± 0.0159,CI = [0.5323, 0.4699]). The main effect of perceived intention was also significant, $F(1,218) = 107.846$, $p < 0.001$, partial $\eta^2 = 0.133$, with the rejection rate was lower for perceived good intention (0.2532 ± 0.0160, CI = [0.2846, 0.2218]) than for unfair offers(0.4953 ± 0.0160, CI = [0.5267, 0.4639]). No significant main effect of other factors was found. There was a significant interaction between fairness and openness, $F(1,218) = 4.663$, $p < 0.05$, partial $\eta^2 = 0.007$. Simple effects tests showed that if the offers were fair, the rejection rate for uninformed offers(0.2886 ± 0.0269, CI = [0.3413, 0.2359]) was significant higher than that of informed ones(0.2056 ± 0.0190, CI = [0.1682, 0.2430]), $p < 0.05$, $F(1,218) = 6.335$, partial $\eta^2 = 0.009$); when the offers were unfair, the corresponding average rejection rates were not significantly different, $p = 0.5922$. The interaction effect of smiling and openness on average rejection rate was also significant, $F(1,218) = 6.396$, $p < 0.05$, partial $\eta^2 = 0.009$. The proceeding simple effects test showed the average rejection rate for uninformed offers was higher

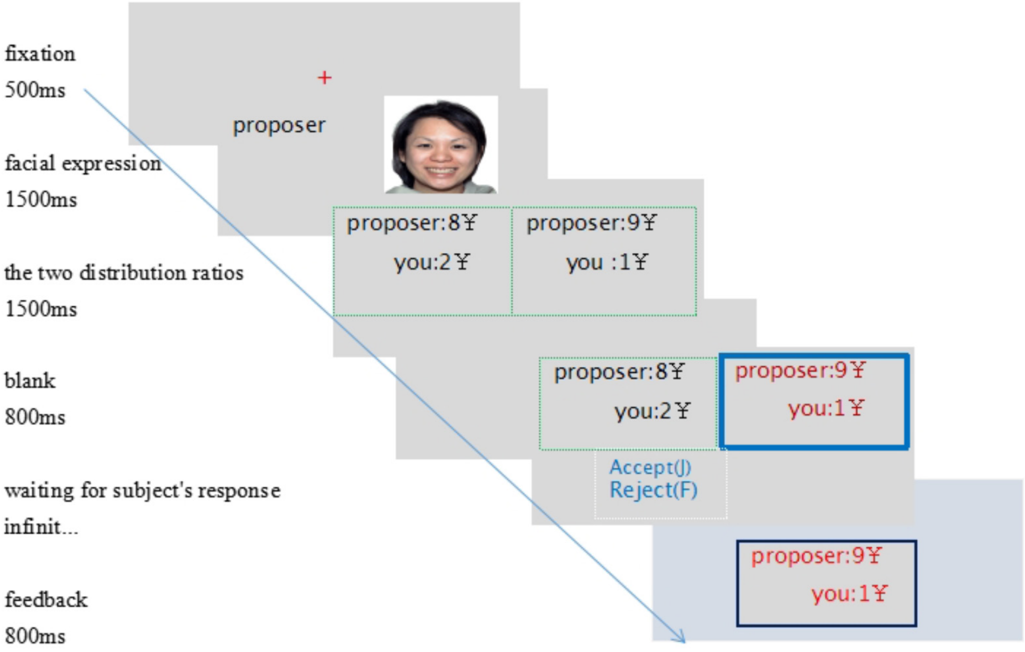

FIGURE 1 | Experimental task. (1) Fixation; (2) the proposer's facial expression; (3) the alternative division; (4) the proposer's thinking; (5) the subject's decision; (6) feedback.

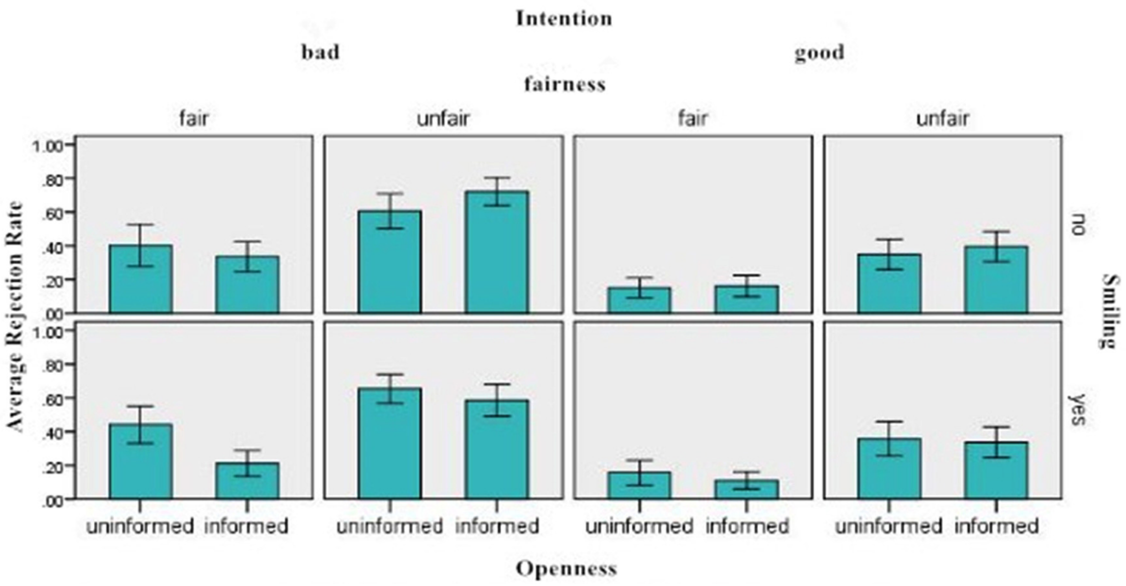

FIGURE 2 | The average rejection rates as a function of intention, fairness, expression, and openness.

(0.4031 ± 0.0269, CI = [0.3502, 0.4559]) than that for informed ones(0.3115 ± 0.0190, CI = [0.2741, 0.3489]) when the proposer was smiling, $p < 0.01$, partial $\eta^2 = 0.011$; there was no difference between the average rejection rate of uninformed offers and that of informed offers when the proposer didn't smile, $p = 0.5436$. No other two way interaction effect was found, and no any three way effect or four way effect was found either.

DISCUSSION

The data showed that fairness and perceived intention had significant effects on perceived fairness. So Hypothesis I and Hypothesis II were proved. In the distribution of resources, a fairer distribution led to a lower average rejection rate, which can be explained by utility theory (French, 2006) or unfair aversion model (Fehr and Schmidt, 1999). Unfair monetary UG offers elicit anger and might result in rejection (Gilam et al., 2015). Utility theory assumes the preferences of utility when people decide among alternatives. In our experiment, fairness meant more favorable outcomes (or more utility) for the receiver, so it is natural that fairness led to a low average rejection rate.

Perceived good intention tends to increase perceived fairness. Researchers have shown that procedural fairness has a considerable influence on employees' attitudes toward their organization and its members (Brockner et al., 2003). We deduced that perceived intention in our experiment might partly refer to procedural unfairness, which was uncontrollable for the receiver but controllable for the proposer. Perceived bad intention also induced angry and retaliatory behavior, so when the proposer made an unfair decision, the bad intention perceived by the receiver may have resulted in a relatively high average rejection rate. Or, as someone puts it (Rabin, 1993): Fairness means that if you are kind to me, I will be kind to you, but if you mean bad to me, then I will do the same to you. So the concept of perceived intention directly penetrates the meaning of fairness.

Interaction between fairness and openness was significant, as the data showed, with the average rejection rate for fair, uninformed offers higher than fair, informed ones, Hypothesis III was manifested. This may because the fair distributions are "should be taken ones" and reject them may be viewed as either wicked or unwise, so more offers were rejected if anonymous. To some extent, it may also be attributed to Chinese culture: Chinese people refuse relatively less in public. The mentality of 'Don't lose face' or 'worldly wise' was severe in China, so informed fair offers were accepted more easily: accepting the offers under the openness condition meant saving the proposer's face, that would finally help the responder himself/herself. As for fair and uninformed offers, it was always safe, so the responders might feel freer to act as what they are pleased. We reasoned that when unfair distribution appeared, an anchoring effect (Strack and Mussweiler, 1997) might occur and the informed or uninformed offers were indistinguishably treated. That is, unfairness was the most important working information for judgement. This might mean other factors had little effect when unfair distribution occurred, the final decision tended to favor the effect of unfairness. One might anticipate logically that when it's unfair, the spirit of "Do boldly what is righteous for public good" should work and lead to more average rejections of informed offers, as other researchers had described (Abbink et al., 2004). But this didn't happen. However, it didn't mean that more financial considerations than moral ones prevailed in the decision making. The unfair offers did take the form of an anchoring effect, but "safely play" counteracted the moral concern under informed condition was another possible additional reason.

This might explain why our responders didn't teach a lesson more often when unfair, informed offers provided than when unfair, uninformed ones.

The interaction effect of smiling and openness was also significant, with the smiling average rejection rate for uninformed offers was higher than that for informed ones. Hypothesis IV was manifested. According to the spreading-activation theory (Loftus, 1975), the awakening of a semantic concept will activate related concepts in the neural network simultaneously. Fairness perception relates to emotions (Namkung and Jang, 2010), upon observing a smiling face, the anchoring effect bias (Bennett, 2014) took place, concepts such as "good person," "pleasure to see" and "like" might be activated. We reasoned that smiling stirred good feelings, and when the acceptances were open, and the reponders were more likely to convey a kind repay. When the smiling expression appeared the responder might take it as the intrinsic nature of the proposer (Chee and Murachver, 2012), and if anonymous was available, to teach a lesson was a natural and safe action, and also, a noble decision. This anchoring effect was different from the traditional Chinese culture of "Don't be angry to the person in smiles," which means people tend to forgive those who apologize honestly. Our outcome might partly attribute to the traditional Chinese culture: when the decision would be sent to the proposer, declining an offer from a smiling face would easily get into an embarrassed situation that most Chinese people would try to avoid. So the corresponding average rejection rate was lower than that of smiling but uniformed offers. The anchoring effect of smiling was thus revised. According to attribution theory (Kelley, 1967), a no-smiling expression might show that the proposer does not have control, so openness didn't make difference. It was like in price-fairness experiments, price increases were perceived as less fair when the causality was directly attributable to the seller's controllable actions (Vaidyanathan and Aggarwal, 2003). Chinese people are easy to forgive, especially when the wrongdoers are forced to do, this we attributed to a strong Chinese traditional culture of "forgive wherever you can." That was a possible reason for equally rejected no-smiling, informed.

CONCLUSION

We found the fairness of a distribution itself affects perceived fairness. The fairer the distribution, the lower the average rejection rate. The distributor's perceived good intention leads to a lower average rejection rate, as the results show. We also found that smiling facial expressions moderate the effect of openness: smiling and openness lead to a lower average rejection rate, and fairness moderate the effect of openness: it is beneficial for the proposer to smile when he/she could get the information of the responder's decision for the sake of the offer to be accepted.

AUTHOR CONTRIBUTIONS

ZH and JL designed the experiments. ZH and ZR collected the data. ZH, JL, ZR, and LW wrote the manuscript.

FUNDING

This study was supported by the National Social Science Foundation of China (14BSH071) and by grants from the Social Science Foundation of Jiangxi Province (15JY14 and XL1519).

REFERENCES

Abbink, K., Sadrieh, A., and Zamir, S. (2004). Fairness, public good, and emotional aspects of punishment behavior. *Theory Decis.* 57, 25–57. doi: 10.1007/s11238-004-3672-8

Alicke, M. D. (1992). Culpable causation. *J. Pers. Soc. Psychol.* 63, 368–378. doi: 10.1037/0022-3514.63.3.368

Bennett, M. W. (2014). Confronting cognitive 'anchoring effect' and 'blind spot' biases in federal sentencing: a modest solution for reforming a fundamental flaw. *J. Crim. Law Criminol.* 104, 489–534.

Blue, P. R., Hu, J., Wang, X., Van, D. E., and Zhou, X. (2016). When do low status individuals accept less? The interaction between self- and other-status during resource distribution. *Front. Psychol.* 7:1667. doi: 10.3389/fpsyg.2016.01667

Brockner, J., Heuer, L., Magner, N., Folger, R., Umphress, E., van den Bos, K., et al. (2003). High procedural fairness heightens the effect of outcome favorability on self-evaluations: an attributional analysis. *Organ. Behav. Hum. Decis. Process.* 91, 51–68. doi: 10.1016/S0749-5978(02)00531-9

Bugental, D. E. (1974). Interpretations of naturally occurring discrepancies between words and intonation: modes of inconsistency resolution. *J. Pers. Soc. Psychol.* 30, 125–133. doi: 10.1037/h0036654

Carl, D., Steve, M., and Doug, N. (2006). Fairness and the political economy of trade. *World Econ.* 29, 989–1004. doi: 10.1111/j.1467-9701.2006.00832.x

Chee, C. S., and Murachver, T. (2012). Intention attribution in theory of mind and moral judgment. *Psychol. Stud.* 57, 40–45. doi: 10.1007/s12646-011-0133-7

Darley, J. M., and Pittman, T. S. (2003). The psychology of compensatory and retributive justice. *Pers. Soc. Psychol. Rev.* 7, 324–336. doi: 10.1207/S15327957PSPR0704_05

Dufwenberg, M., and Kirchsteiger, G. (1998). A theory of sequential reciprocity. *Games Econ. Behav.* 47, 268–298. doi: 10.1016/j.geb.2003.06.003

Falk, A., Fehr, E., and Fischbacher, U. (2003). On the nature of fair behavior. *Econ. Inq.* 41, 20–26. doi: 10.1093/ei/41.1.20

Fehr, E., and Schmidt, K. M. (1999). *A Theory of Fairness, Competition, and Cooperation*. Munich: University of Munich.

Fishburn, P. C. (1967). Utility theory. *Manage. Sci.* 14, 335–378. doi: 10.1287/mnsc.14.5.335

French, S. (2006). *Utility Theory. Encyclopedia of Environmetrics*. Hoboken, NJ: John Wiley & Sons, Ltd.

Gebotys, R. J., and Dasgupta, B. (1987). Attribution of responsibility and crime seriousness. *J. Psychol.* 121, 607–613. doi: 10.1080/00223980.1987.9712690

Gilam, G., Lin, T., Raz, G., Azrielant, S., Fruchter, E., Ariely, D., et al. (2015). Neural substrates underlying the tendency to accept anger-infused ultimatum offers during dynamic social interactions. *Neuroimage* 120, 400–411. doi: 10.1016/j.neuroimage.2015.07.003

Gollwitzer, M., Rothmund, T., and Cremer, D. D. (2009). "When the need to trust results in unethical behavior: the sensitivity to mean intentions (SeMI) model," in *Psychological Perspectives on Ethical Behavior and Decision Making*, ed. D. De Cremer (Charlotte, NC: Information Age Publishing), 135–152.

Gray, K., and Wegner, D. M. (2010). Blaming god for our pain: human suffering and the divine mind. *Pers. Soc. Psychol. Rev.* 14, 7–16. doi: 10.1177/1088868309350299

Güroglu, B., Van den Bos, W., Van Dijk, E., Rombouts, S. A., and Crone, E. A. (2011). Dissociable brain networks involved in development of fairness considerations: understanding intentionality behind unfairness. *Neuroimage* 57, 634–641. doi: 10.1016/j.neuroimage.2011.04.032

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* 3, 367–388. doi: 10.1016/0167-2681(82)90011-7

Güth, W., Schmittberger, R., and Schwarze, B. (1983). A theoretical and experimental analysis of bidding behavior in Vickrey-auction games. *Z. Gesamte Staatswiss.* 139, 269–288.

Haidt, J., and Kesebir, S. (2010). "Morality," in *Handbook of Social Psychology*, 5th Edn, eds S. Fiske, D. Gilbert, and G. Lindzey (Hoboken, NJ: Wiley), 797–832.

Heussler, T., Huber, F., Meyer, F., Vollhardt, K., and Ahlert, D. (2009). "Moderating effects of emotion on the perceived fairness of price increases," in *Advances in Consumer Research*, Vol. 36, eds A. L. McGill and S. Shavitt (Duluth, MN: Association for Consumer Research), 332–338.

Kelley, H. H. (1967). Attribution theory in social psychology. *Nebr. Symp. Motiv.* 15, 192–238.

Knobe, J., Buckwalter, W., Nichols, S., Robbins, P., Sarkissian, H., and Sommers, T. (2012). *Experimental Philosophy.* Oxford: Oxford University Press.

Krumhuber, E., Manstead, A. S. R., Cosker, D., Marshall, D., and Rosin, P. L. (2009). Effects of dynamic attributes of smiles in human and synthetic faces: a simulated job interview setting. *J. Nonverbal Behav.* 33, 1–15. doi: 10.1007/s10919-008-0056-8

Lerner, M. J. (1965). Evaluation of performance as a function of performer's reward and attractiveness. *J. Pers. Soc. Psychol.* 95, 355–360. doi: 10.1037/h0021806

Loftus, E. F. (1975). Spreading activation within semantic categories: comments on Rosch's "cognitive representation of semantic categories." *J. Exp. Psychol. Gen.* 104, 234–240. doi: 10.1037/0096-3445.104.3.234

Ma, Q., Hu, Y., Jiang, S., and Meng, L. (2015a). The undermining effect of facial attractiveness on brain responses to fairness in the ultimatum game: an ERP study. *Front. Neurosci.* 9:77. doi: 10.3389/fnins.2015.00077

Ma, Q., Meng, L., Zhang, Z., Xu, Q., Wang, Y., and Shen, Q. (2015b). You did not mean it: perceived good intentions alleviate sense of unfairness. *Int. J. Psychophysiol.* 96, 183–190. doi: 10.1016/j.ijpsycho.2015.03.011

McCabe, K. A., Rigdon, M. L., and Smith, V. L. (2003). Positive reciprocity and intentions in trust games. *J. Econ. Behav. Organ.* 52, 267–275. doi: 10.1016/S0167-2681(03)00003-9

Mendoza, S. A., Lane, S. P., and Amodio, D. M. (2014). For members only: ingroup punishment of fairness norm violations in the ultimatum game. *Soc. Psychol. Pers. Sci.* 5, 662–670. doi: 10.1177/1948550614527115

Mussel, P., Göritz, A. S., and Hewig, J. (2013). The value of a smile: facial expression affects ultimatum-game responses. *Judgm. Decis. Mak.* 8, 1–5.

Mussel, P., Hewig, J., Allen, J. J., Coles, M. G., and Miltner, W. (2014). Smiling faces, sometimes they don't tell the truth: facial expression in the ultimatum game impacts decision making and event-related potentials. *Psychophysiology* 51, 358–363. doi: 10.1111/psyp.12184

Namkung, Y., and Jang, S. C. (2010). Effects of perceived service fairness on emotions, and behavioral intentions in restaurants. *Eur. J. Mark.* 44, 1233–1259. doi: 10.1108/03090561011062826

Nowak, M. A., Page, K. M., and Sigmund, K. (2000). Fairness versus reason in the ultimatum game. *Science* 289, 1773–1775. doi: 10.1126/science.289.5485.1773

Raaij, W. F. V., and Ye, G. W. (2002). "Implicit expected utility theory for decision making and choice," in *Asia Pacific Advances in Consumer Research*, Vol. 5, eds Ramizwick and T. Ping (Valdosta, GA: Association for Consumer Research), 343–348.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *Am. Econ. Rev.* 83, 1281–1302.

Rothmund, T., Gollwitzer, M., and Klimmt, C. (2011). Of virtual victims and victimized virtues: differential effects of experienced aggression in video games on social cooperation. *Pers. Soc. Psychol. Bull.* 37, 107–119. doi: 10.1177/0146167210391103

Shaver, K. G. (1970). Defensive attribution: effects of severity and relevance on the responsibility assigned for an accident. *J. Pers. Soc. Psychol.* 14, 101–113. doi: 10.1037/h0028777

Stanca, L., Bruni, L., and Corazzini, L. (2009). Testing theories of reciprocity: do motivations matter? *J. Econ. Behav. Organ.* 71, 233–245. doi: 10.1016/j.jebo.2009.04.009

Stouten, J., Ceulemans, E., Timmerman, M. E., and Hiel, A. V. (2011). Tolerance of justice violations: the effects of need on emotional reactions after violating equality in social dilemmas 1. *J. Appl. Soc. Psychol.* 41, 357–380. doi: 10.1111/j.1559-1816.2010.00717.x

Strack, F., and Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: mechanisms of selective accessibility. *J. Pers. Soc. Psychol.* 73, 437–446. doi: 10.1037/0022-3514.73.3.437

Vaidyanathan, R., and Aggarwal, P. (2003). Who is the fairest of them all? An attributional approach to price fairness perceptions. *J. Bus. Res.* 56, 453–463. doi: 10.1016/S0148-2963(01)00231-4

van't Wout, M., and Sanfey, A. G. (2008). Friend or foe: the effect of implicit trustworthiness judgments in social decision-making. *Cognition* 108, 796–803. doi: 10.1016/j.cognition.2008.07.002

Young, L., and Saxe, R. (2009). Innocent intentions: a correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia* 47, 2065–2072. doi: 10.1016/j.neuropsychologia.2009.03.020

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer XL and handling Editor declared their shared affiliation.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership