DISCRIMINATION OF GENUINE AND POSED FACIAL EXPRESSIONS OF EMOTION

EDITED BY: Huiyu Zhou, Ling Li, Shiguang Shan, Shuo Wang and Jian K. Liu
PUBLISHED IN: Frontiers in Psychology, Frontiers in Neuroscience and Frontiers in Computational Neuroscience

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-789-7
DOI 10.3389/978-2-88971-789-7

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: frontiersin.org/about/contact

DISCRIMINATION OF GENUINE AND POSED FACIAL EXPRESSIONS OF EMOTION

Topic Editors:
Huiyu Zhou, University of Leicester, United Kingdom
Ling Li, University of Kent, United Kingdom
Shiguang Shan, University of Chinese Academy of Sciences, China
Shuo Wang, West Virginia University, United States
Jian K. Liu, University of Leeds, United Kingdom

Citation: Zhou, H., Li, L., Shan, S., Wang, S., Liu, J. K., eds. (2021). Discrimination of Genuine and Posed Facial Expressions of Emotion. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-789-7

Table of Contents

frontiers in Neuroscience

EDITORIAL
published: 07 October 2021
doi: 10.3389/fnins.2021.771529

Editorial: Discrimination of Genuine and Posed Facial Expressions of Emotion

Huiyu Zhou[1]*, *Ling Li*[2], *Shiguang Shan*[3], *Shuo Wang*[4] and *Jian K. Liu*[5]

[1] School of Computing and Mathematical Sciences, University of Leicester, Leicester, United Kingdom, [2] School of Computing, University of Kent, Canterbury, United Kingdom, [3] Institute of Computing Technology, University of Chinese Academy of Sciences, Beijing, China, [4] Department of Chemical and Biomedical Engineering and Rockefeller Neuroscience Institute, West Virginia University, Morgantown, WV, United States, [5] School of Computing, University of Leeds, Leeds, United Kingdom

Keywords: discrimination, genuine, posed facial expressions, emotion, visual presentation

Editorial on the Research Topic

Discrimination of Genuine and Posed Facial Expressions of Emotion

Facial expressions demonstrate emotional states in interpersonal situations. Evidence shows that part of the facial display reflects the emotional experience that is literally felt by the expresser. Interestingly, human beings are capable of identifying facial expressions of the sensed emotions as a form of intentional deceit to conduct social interaction and to present displays that have the support of others. Staged or posed facial expressions implement an emotion that an expresser intends to convey, where genuine expressions are considered as the companion of spontaneous emotional expressions. The ability to differentiate genuine displays of emotional experience from posed ones is very important for dealing with day-to-day social interactions.

Recent work has been conducted on whether or not people can distinguish between posed and genuine displays of emotion. In spite of few studies to investigate this ability, most prior research suggests that people have the ability to judge genuine and posed facial displays. Unfortunately, previous research has suffered from two major shortcomings: (1) the mixture of staged and genuine displays due to the lack of accounting for possible effects of intentional manipulation, and (2) struggling to consider dynamic aspects when people launch facial stimuli for experimental investigation.

This Research Topic consists of the submission of theoretical and experimental perspectives to broaden understanding of the importance of the discrimination of genuine and posed facial expressions of emotion. Some of them report new theoretical approaches, those from other disciplines of psychology not usually utilized within the discrimination of genuine and staged emotion identification or new theories and designs.

In the article entitled "The role of low-spatial frequency components in the processing of deceptive faces: A study using artificial face models," Kihara and Takeda investigated how spatial frequency information can be used to interpret true emotion. "A call for the empirical investigation of tear stimuli" authored by Krivan and Thomas presents a study on the necessity for empirical investigations of the differences (or similarities) in response to posed and genuine tearful expressions. Zhang et al. conducted research on "Brain activation in contrasts of micro-expression following emotional contexts" with the prediction that the effect of emotional contexts may be dependent on neural activities. Lander and Butcher's article, entitled "Recognizing genuine from posed facial expressions: Exploring the role of dynamic information and face familiarity," reported the importance of motion for the recognition of face identity before critically outlining the role of dynamic information in determining facial expression and distinguishing between genuine and posed expression of emotion.

OPEN ACCESS

Edited and reviewed by:
*Rufin VanRullen,
Centre National de la Recherche Scientifique (CNRS), France*

****Correspondence:***
*Huiyu Zhou
hz143@leicester.ac.uk*

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

***Received:** 06 September 2021*
***Accepted:** 22 September 2021*
***Published:** 07 October 2021*

Citation:
Zhou H, Li L, Shan S, Wang S and Liu JK (2021) Editorial: Discrimination of Genuine and Posed Facial Expressions of Emotion. Front. Neurosci. 15:771529. doi: 10.3389/fnins.2021.771529

Jia et al. introduced a review of the relevant research including spontaneous vs. posed (SVP) facial expression databases and computer vision based detection methods in their article entitled "Detection of genuine and posed facial expressions of emotion: Databases and methods." In the article authored by Ron-Angevin et al., "Performance analysis with different types of visual stimuli in a BCI-based speller under an RSVP paradigm," three different sets of stimuli were assessed under rapid serial visual presentation with the following communication features: white letters, famous faces and neutral pictures. In the article "Identifying emotional expressions: Children's reasoning about pretend emotions of sadness and anger," Serrat et al. attempted to understand children's capacity of identifying pretend emotions by analyzing different sources of information when interpreting emotions simulated in pretend play contexts. In the research work "Deep neural networks for depression recognition based on 2D and 3D facial expressions under emotional stimulus tasks," Guo et al. created a large scale dataset with subjects of performing five mood elicitation tasks. They also proposed a novel approach for depression recognition using two different deep belief network models. Finally, this thematic topic includes a survey authored by Webster et al., namely "Review: Posed vs. genuine facial emotion recognition and expression in Autism and implications for intervention," where the literature in studying the deficits of facial emotion recognition in those with autism spectrum disorder is comprehensively discussed.

The studies presented above have set up a landmark to the research on discrimination of genuine and posed facial expressions of emotion. Moving forward, we anticipate more and more research work on deep and thorough analysis of emotion using emerging artificial intelligence techniques.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 26 June 2019
doi: 10.3389/fpsyg.2019.01468

The Role of Low-Spatial Frequency Components in the Processing of Deceptive Faces: A Study Using Artificial Face Models

Ken Kihara and Yuji Takeda*

Automotive Human Factors Research Center, National Institute of Advanced Industrial, Science and Technology (AIST), Tsukuba, Japan

Interpreting another's true emotion is important for social communication, even in the face of deceptive facial cues. Because spatial frequency components provide important clues for recognizing facial expressions, we investigated how we use spatial frequency information from deceptive faces to interpret true emotion. We conducted two different tasks: a face-generating experiment in which participants were asked to generate deceptive and genuine faces by tuning the intensity of happy and angry expressions (Experiment 1) and a face-classification task in which participants had to classify presented faces as either deceptive or genuine (Experiment 2). Low- and high-spatial frequency (LSF and HSF) components were varied independently. The results showed that deceptive happiness (i.e., anger is the hidden expression) involved different intensities for LSF and HSF. These results suggest that we can identify hidden anger by perceiving unbalanced intensities of emotional expression between LSF and HSF information contained in deceptive faces.

Keywords: facial expression, deceptive face, spatial frequency, face-generating task, face-classification task

OPEN ACCESS

Edited by:
Jian K. Liu,
University of Leicester,
United Kingdom

Reviewed by:
Giulia Prete,
Università degli Studi G. d'Annunzio
Chieti e Pescara, Italy
Elena Tsankova,
Jacobs University Bremen,
Germany

****Correspondence:***
Ken Kihara
ken-kihara@aist.go.jp

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

Received: *07 December 2018*
Accepted: *11 June 2019*
Published: *26 June 2019*

Citation:
Kihara K and Takeda Y (2019) The Role of Low-Spatial Frequency Components in the Processing of Deceptive Faces: A Study Using Artificial Face Models.
Front. Psychol. 10:1468.
doi: 10.3389/fpsyg.2019.01468

INTRODUCTION

In our daily communication, facial expressions are one of the main cues used to understand other people's emotions or internal states. People often try to conceal their emotions (i.e., what they are truly feeling), instead presenting an opposing or different expression (Porter et al., 2011a). Nevertheless, we do depend on understanding true emotions in order to establish good personal relationships. Thus, interpreting true emotion is important for favorable communication (King, 1998; Butler and Gross, 2004). Generally speaking, it is difficult to generate expressions that appear the same as spontaneous ones. For example, deceptive happiness expressions are distinguishable from genuine happiness expressions by observing the movements of the zygomatic major and orbicularis oculi muscles (Ekman and Friesen, 1982). In fact, observers can discriminate between genuine and deceptive facial expressions rather rapidly (Porter and Ten Brinke, 2008). The interpretation of facial expressions depends on the observer. This is because one observer might judge a face as showing genuine anger, whereas another observer might judge the same face as showing deceptive anger. However, the type of facial information that is used for interpreting another person's hidden emotions or recognizing deceptive faces is unclear.

In this study, we focused on the spatial frequency components of faces, which are important for interpreting facial expressions (Ruiz-Soler and Beltran, 2006). Low-spatial frequencies (LSFs) carry information about the configural properties of facial parts, such as the eyes, the nose, and the mouth, whereas high-spatial frequencies (HSFs) contain finer, edge-based information supporting the processing of these features, resulting in detailed image representations and object boundaries (Goffaux et al., 2005). In studies examining the perception of static natural scenes, the parallel processing of LSF and HSF information extracted and integrated from scene images has led to the rapid interpretation of scenes presented for a short duration (i.e., 100 ms) and perceived ambiguously due to the short presentation (e.g., Kihara and Takeda, 2010, 2012). Consequently, we expected that combining different information provided by LSF and HSF would contribute to interpreting not only natural scenes presented for a short duration but also ambiguous facial expressions. Related to the above, it has been suggested that LSF and HSF components play different roles in the perception of facial expressions. Schyns and Oliva (1997, 1999) found that if face stimuli are hybrid images composed of one expression in LSF and another expression in HSF, categorizing facial expression (e.g., happiness versus anger) is dependent mainly on LSF, whereas identifying the presence of emotional expression (e.g., emotional versus neutral) is based on HSF information. Although previous findings are based on genuine-face stimuli, LSF and HSF may also contain different types of emotional cues that are used to interpret deceptive facial expressions. Indeed, characteristics of deceptive facial expressions are shown in both upper and lower face (Porter et al., 2011b), i.e., deceptive facial expression does not depend on characteristics of specific facial parts. This implies that LSF information carrying the global shape and structure of a face may play an important role in identifying deceptive facial expressions.

Several studies have identified the differential contributions of LSF and HSF to the interpretation of facial expressions. For example, Laeng et al. (2010) used hybrid images composed of emotional expressions in LSF and neutral expression in HSF and demonstrated that LSF plays an essential role in the implicit detection of emotional expressions. The participants in their study rated the images as friendly or unfriendly based on the LSF component, whereas they explicitly judged the images as neutral regardless of the LSF component. Importantly, Prete et al. (2014, 2015c) reported hemispheric asymmetry in neural processing for implicit detection of emotional expressions because hybrid images tend to be rated as less friendly when they are presented in the left visual field than in the center or the right visual fields (see also Prete et al., 2018b, for a transcranial stimulation study). This tendency is also shown when unfiltered, intact images are used as the to-be-rated expressions (Prete et al., 2015b). Also, both hybrid and intact images cause emotional aftereffects in that presenting a negative expression causes the perception of subsequent neutral expressions to be judged more positively and vice versa (Prete et al., 2018a). Furthermore, an event-related potentials study has indicated that facial and emotional processing-related P1, N170, and P2 components are evoked by hybrid, as well as intact images (Prete et al., 2015a). Such evidence suggests that hemispheric asymmetry for implicit emotional processing is a robust phenomenon that is not limited to the specific use of hybrid images. Interestingly, sensitivity for the implicit detection of emotional expressions is enhanced after oxytocin treatment as reflected by pupilar dilation because of the allocation of attention to socially relevant information (Leknes et al., 2013). These findings clearly suggest that LSF but not HSF component contributes to the implicit perception of emotional faces, implying that the perception of deceptive facial expressions, which might be processed intuitively and implicitly, is affected by LSF rather than HSF component.

It has also been demonstrated that LSF and HSF do not equally contribute to identifying negative expressions (Vuilleumier, 2005). Understanding negative emotion from facial expressions is potentially important for survival (relative to positive emotion), and this may be why the visual system is biased toward the processing of negative expressions (Taylor, 1991). In fact, negative expressions attract and hold attention more frequently and for longer than positive expressions (Mathews et al., 1997). Importantly, the prioritized processing of negative expressions is specifically attributed to the neural pathway tuned to LSF components (Vuilleumier, 2005). Low-spatial frequency preserves coarse information associated with an object's shape and layout, which is transmitted to the cortex and subcortical structures through the rapid magnocellular pathways (Bar, 2004). A functional magnetic resonance imaging (fMRI) study showed that the human amygdala, which allocates attention to negative stimuli (LeDoux, 1995), selectively responds to the LSF, but not the HSF, component of fearful expressions (Vuilleumier et al., 2003). This result suggests that LSF information projected to amygdala *via* the magnocellular pathways plays an important role in processing negative expressions (Winston et al., 2003; Pourtois et al., 2005; but see also Holmes et al., 2005; Morawetz et al., 2011, for different results). It is therefore possible that the involvement of LSF components differs when viewing deceptive faces hiding negative versus positive emotional states.

The current study investigated whether LSF or HSF components are more important when interpreting deceptive faces. To address this issue, we examined the intensity of the emotional expressions contained in LSF and HSF components of deceptive faces using two different tasks. It is known that dynamic elements of faces are critical for interpreting facial expressions (Krumhuber et al., 2016) because perception of dynamic elements is asymmetrically processed in LSF and HSF (Kauffmann et al., 2014). However, in this study, we focused on static situations for investigating the basic role of spatial frequency information on interpreting deceptive expressions. In Experiment 1, we used a face-generating task where participants were asked to generate genuine and deceptive faces by tuning the intensity of specific expressions in both LSF and HSF. There were two genuine faces: genuine happiness (positive) and genuine anger (negative). There were also two deceptive faces: deceptive happiness (concealing genuine anger) and deceptive anger (concealing genuine happiness). In Experiment 2, we used a face-classification task where participants were asked to classify presented faces composed of LSF and HSF expressions as either genuine happiness, genuine anger, deceptive happiness, or deceptive anger. Because the LSF component is important for discriminating positive versus negative facial expressions (Schyns and Oliva, 1997, 1999), we assumed that identifying the hidden expression would depend on the intensity of the expression represented in LSF. We predicted that interpreting hidden negative emotion would depend heavily on the LSF component of the deceptive positive faces, because LSF information plays a critical role in the preferential processing of negative expressions (Vuilleumier, 2005). It is important to note that we used artificial facial models because we had to control the intensity of facial expressions step by step. In addition, the artificial facial models allow us to make different facial expressions with minimal changes of each individual faces, which should be critical to superimposing two images consisting of different spatial frequencies as an integrated one.

EXPERIMENT 1

Materials and Methods

Participants

Twenty-seven adult males (mean age 23, range 19–31) from the subject pool at the National Institute of Advanced Industrial Science and Technology participated in this experiment. All participants received payment for their participation. All had normal or corrected-to-normal vision and were right-handed. This study was carried out in accordance with the recommendations of Guidelines for handling ergonomic experiments, Committee on Ergonomic Experiments, Bioethics and Biosafety Management Office, Safety Management Division, National Institute of Advanced Industrial Science and Technology and approved by the Committee on Ergonomic Experiments, Bioethics and Biosafety Management Office, Safety Management Division, National Institute of Advanced Industrial Science and Technology. All participants gave written informed consent in accordance with the Declaration of Helsinki.

Stimuli and Apparatus

Examples of facial images are given in **Figure 1A**. Eighty individual faces in frontal view were randomly generated using FaceGen Modeller 3.5 (Singular Inversions Inc.). FaceGen Modeller software allows us to manipulate realistic facial expressions, which are available for a wide range of facial expression studies (e.g., Corneille et al., 2007; Schulte-Rüther et al., 2007; Oosterhof and Todorov, 2008; Xiao et al., 2015; Hass et al., 2016). Faces were randomized for gender, age, race, and features (brow ridge, cheekbones, etc.). Each face was morphed from neutral to happy (positive) and angry (negative) in 10 steps of increasing intensity. Thus, there were 21 variations in expression, including the neutral expression, for each individual face. These expressions were given values of from −10 (the most angry) to 10 (the most happy). The neutral face was given a value of zero. The resolution of the images was 400 × 400 pixels, which subtended 6° of visual angle at a viewing distance of about 57 cm. The width of each face was about half the size of the image width. All images were converted into grayscale LSF and HSF images. The ranges of band-pass frequencies for LSF and HSF were selected based on previous studies (Schyns and Oliva, 1999; Vuilleumier et al., 2003). The LSF images were filtered in Fourier space, using a fourth-order Butterworth filter, set to filter low band-pass frequencies (1.33–2.67 cycle/degree). The HSF images were filtered with a fourth-order Butterworth high band-pass filter (5.33–10.67 cycle/degree).

Procedure

There were four conditions, such as genuine happiness, genuine anger, deceptive happiness, and deceptive anger, presented in separate blocks of trials. The 80 individual faces were randomly assigned to each block, 20 faces in each. Block order was randomized. Each participant completed a total of 80 trials (4 conditions in separate blocks × 20 individual faces). Before the experiment began, participants completed eight practice trials, using different faces that were not part of the experimental trials. The experiment was conducted in a darkened room and took about 30 min to complete.

Each block began with instructions as to which face type was to be generated: “Please generate the following expression” and then “Genuine happiness,” “Genuine anger,” “Happiness but actually anger,” or “Anger but actually happiness” in Japanese. The instruction remained displayed until the central key on the game controller (designated as the decision key) was pressed. Subsequently, a randomly selected face was presented at the center of the display with the face type to-be-generated displayed below. Each face was created by averaging LSF and HSF images, both of which were randomly selected from the 21 variations of each individual face. Participants were asked to generate the instructed expression by pressing the designated keys on the game controller. For example, the up and down keys on the left side were designated to change the LSF image, and the keys on the right side changed the HSF image. The up and down keys changed the expression value of the image in opposite directions by one step. The changes were continuous; when a maximum value was reached and the same key was pressed, the value of the expression began to decrease. Similarly, when a minimum value was reached, the value of the expression began to increase. See **Figure 1B** for a schematic illustration. The assignment of the keys (i.e., left- and right-hand side of the game controller) was counterbalanced across participants. Participants were allowed to generate each face at their own pace. They pressed the decision key when they were finished, after which the next face appeared.

Results and Discussion

Figure 2 shows the mean expression value of the images developed for each spatial frequency in the happiness and anger blocks. In this study, although a three-way analysis of variance (ANOVA) with emotion (happiness or anger), face to-be-generated (deceptive or genuine), and spatial frequency

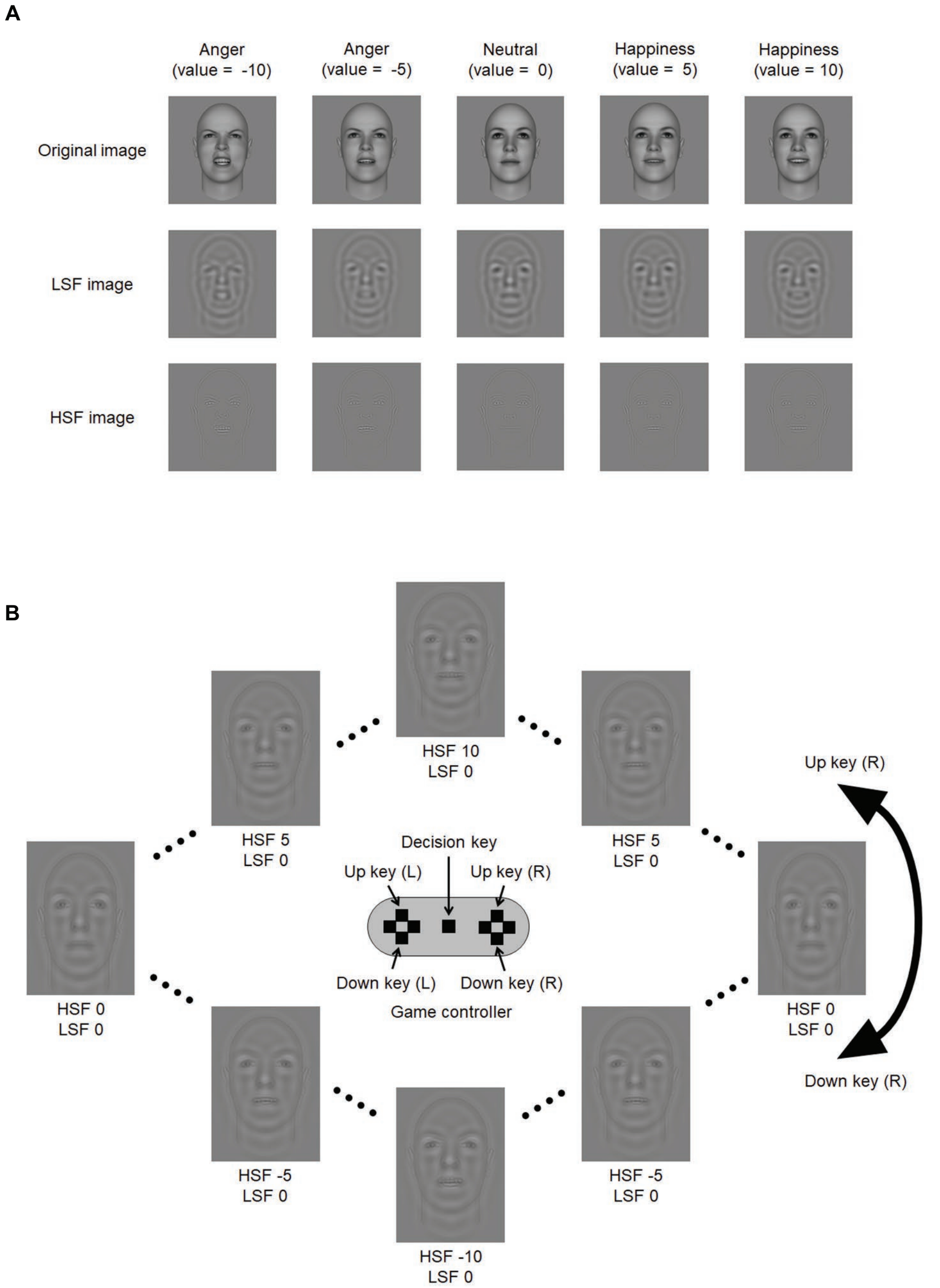


FIGURE 1 | Example of a randomly generated face image by FaceGen Modeller 3.5 software and schematic illustrations of the procedure. **(A)** Example images in the two frequency conditions and the original image. In the experiment, 21 variations in expression from the angriest (expression value of –10) to the most happiness (expression value of 10) were used. LSF images were filtered with low band-pass frequencies (4–8 c/f). HSF images were filtered with a high band-pass filter (16–32 c/f). **(B)** Schematic illustration of the game controller and the relationship between the up and down keys used to change the expression value of the image. In this schematic, the up key on the right side of the controller changes the HSF component of the image, ranging between –10 and 10 (in single steps) in a counterclockwise direction. The down key changes the HSF component of the image in the opposite direction. The up and down keys on the left side of the controller change the LSF component of the image.

as independent variables could be preferred to prevent a possible loss of effects, we conducted two-way ANOVAs separately for the happiness and anger blocks because the meaning of values in the happiness and anger blocks could be in opposite direction. That is, a lower value in the happiness block indicates the facial expressions close to neutral, whereas a lower value in the anger block indicates more negative facial expressions. On the other hand, the biases toward positive (negative) facial expressions can result in higher (lower) values both in the happiness and anger blocks. In

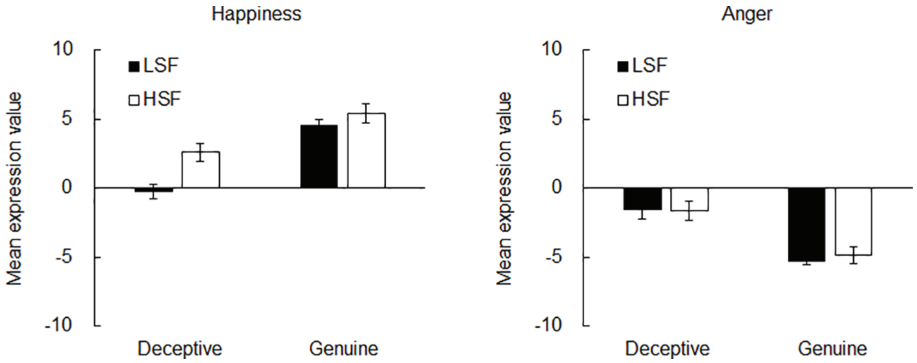


FIGURE 2 | Results of Experiment 1. Mean expression value of the deceptive and genuine images created for each frequency in the happiness (left panel) and anger (right panel) conditions. Error bars indicate standard errors of the mean.

this case, the interpretation of the three-way ANOVA can be very complex. Therefore, we decided to use two-way ANOVAs separately for the happiness and anger blocks. The independent variables (within-subject factors) were face to-be-generated (deceptive or genuine) and spatial frequency. The dependent variable was the mean expression value. The ANOVA revealed that there was a significant main effect of the face to-be-generated when the expression was happiness, $F(1, 26) = 74.77$, $p < 0.001$, $\eta_p^2 = 0.74$. The power of the *post hoc* analysis calculated by G-power 3.1.9 (Faul et al., 2007, 2009) = 1.00. The mean values (±SD) of the deceptive and genuine conditions were 1.16 (±3.29) and 4.97 (±3.04). There was also a significant main effect of the spatial frequency, $F(1, 26) = 7.77$, $p < 0.01$, $\eta_p^2 = 0.23$, power = 1.00. The mean values (±SD) were 2.14 (±3.43) for LSF and 3.98 (±3.73) for HSF. Importantly, there was a significant interaction between the face to-be-generated and the spatial frequency, $F(1, 26) = 5.74$, $p < 0.03$, $\eta_p^2 = 0.18$, power = 1.00. *Post hoc* analysis using the Duncan test ($p < 0.05$) revealed that there were significant differences between all the conditions except between the LSF and HSF conditions in the genuine face condition. The mean values (±SD) were −0.27 (±2.61) for LSF-deceptive, 2.59 (±3.31) for HSF-deceptive, 4.55 (±2.27) for LSF-genuine, and 5.38 (±3.65) for HSF-genuine. These results suggest that genuine happiness contains equally high intensity LSF and HSF components (i.e., approximately five points of mean expression values each), whereas deceptive happiness consists of differential intensities of expression in terms of LSF (i.e., approximately zero points) and HSF (i.e., approximately three points). Conversely, the ANOVA for the angry faces revealed a significant main effect of the face to-be-generated, $F(1, 26) = 61.90$, $p < 0.001$, $\eta_p^2 = 0.70$, power = 1.00 (deceptive: −1.62 ± 3.69; genuine: −5.08 ± 2.43). However, there was no significant main effect of the spatial frequency, $F(1, 26) = 0.04$, $p > 0.84$, $\eta_p^2 = 0.01$, power = 0.07. The mean values (± SD) were −3.42 (±3.40) for LSF and −3.27 (±3.74) for HSF. Also, there was no significant interaction, $F(1, 26) = 0.15$, $p > 0.69$, $\eta_p^2 = 0.01$, power = 0.17. The mean values (± SD) were −1.58 (± 3.72) for LSF-deceptive, −1.66 (± 3.72) for HSF-deceptive, −5.27 (± 1.63) for LSF-genuine, and −4.89 (± 3.05) for HSF-genuine. These results suggest that deceptive anger is different from genuine anger only in terms of the intensity of anger expressed, regardless of spatial frequency.

EXPERIMENT 2

The results of Experiment 1 demonstrated that only deceptive happiness consisted of differential expression intensities for LSF and HSF. These findings were provided by a face-generation task in which participants generated the instructed facial expressions. To validate these results independently of task demands, we next examined whether the findings from the face-generation task could be replicated using another task. In Experiment 2, we used a face-classification task in which participants were asked to classify presented faces that depicted certain LSF and HSF expression values as either genuine happiness, genuine anger, deceptive happiness, or deceptive anger. We predicted that Experiment 2 would produce a similar pattern of results to Experiment 1, if indeed differential intensities of expression between LSF and HSF are an important cue for interpreting deceptive happiness. That is, the faces showing lower LSF expression as compared to HSF would tend to be classified as deceptive happiness.

Materials and Methods

Participants

Thirty-three adult males (mean age 22.4, range 18–34) from the subject pool at National Institute of Advanced Industrial Science and Technology participated in this experiment. All participants received payment for their participation. All had normal or corrected-to-normal vision and two were left-handed. This study was carried out in accordance with the recommendations of Guidelines for handling ergonomic experiments, Committee on Ergonomic Experiments, Bioethics and Biosafety Management Office, Safety Management Division, National Institute of Advanced Industrial Science and Technology and approved by the Committee on Ergonomic Experiments, Bioethics and Biosafety Management Office, Safety Management Division, National Institute of Advanced Industrial Science and Technology.

All participants gave written informed consent in accordance with the Declaration of Helsinki.

Stimuli, Apparatus, and Procedures

Stimuli, apparatus, and procedures were the same as those used in Experiment 1, except for the changes described here. Twenty individual faces were randomly chosen from the pool of 80 individual faces used in Experiment 1. There were five variations of expression value for both LSF and HSF images for each individual face (i.e., expression values are −10, −5, 0, 5, and 10). To-be-classified faces were created by averaging LSF and HSF images, both of which were selected from the five variations of each individual face. All possible combinations of LSF and HSF images were used. Thus, 500 faces (20 individuals × 5 values in LSF × 5 values in HSF) were used for the classification task.

At the start of the experiment, a randomly selected face was presented at the center of the display. After 2,000 ms, participants were asked to classify the presented face as “Genuine happiness,” “Genuine anger,” “Happiness but actually anger,” or “Anger but actually happiness” by pressing the designated keys on the game controller, without time pressure. After pressing the key, the next face appeared. Face order was randomized. Each participant completed a total of 500 trials. Before the experiment began, participants completed eight practice trials, using different faces that were not used during the experimental trials. The experiment was conducted in a darkened room and took about 40 min to complete.

Results and Discussion

Figure 3 shows the mean classification percentages for the four types of face across participants. Obviously, participants tended to classify the faces comprising higher expression values for both LSF and HSF as genuine happiness. Conversely, they classified the faces with lower expression values for both LSF and HSF as genuine anger. On the contrary, zero or near zero values for both LSF and HSF seem to be preferred as the deceptive faces.

We estimated the mode of the data of each participant to clarify the combination of LSF and HSF expression values that were subject to be classified as each face type. If the mode was more than one (i.e., there were two or more peaks in the frequency histogram), they were averaged. We decided to use the mode rather than the mean of the classification proportion because the mean would not reflect the typical values in each category. For example, typical values of genuine anger expression should be near −10 for LSF and HSF. However, the mean values of the classification proportion increase close to zero because of the edge effect (i.e., stimuli more negative than −10 cannot be made). Therefore, the mode was considered appropriate to estimate the typical values in each category. **Figure 4** shows the mean expression value of the mode for each spatial frequency for happiness and anger expressions across the participants. A two-way ANOVA with the mean expression value of the mode as the dependent variable indicated that there was a significant main effect of the face for the expression of happiness, $F(1, 32) = 52.95$, $p < 0.001$, $\eta_p^2 = 0.62$, power = 1.00. The mean values (±SD) were 3.72 (± 4.96) for deceptive and 9.17 (± 2.53) for genuine.

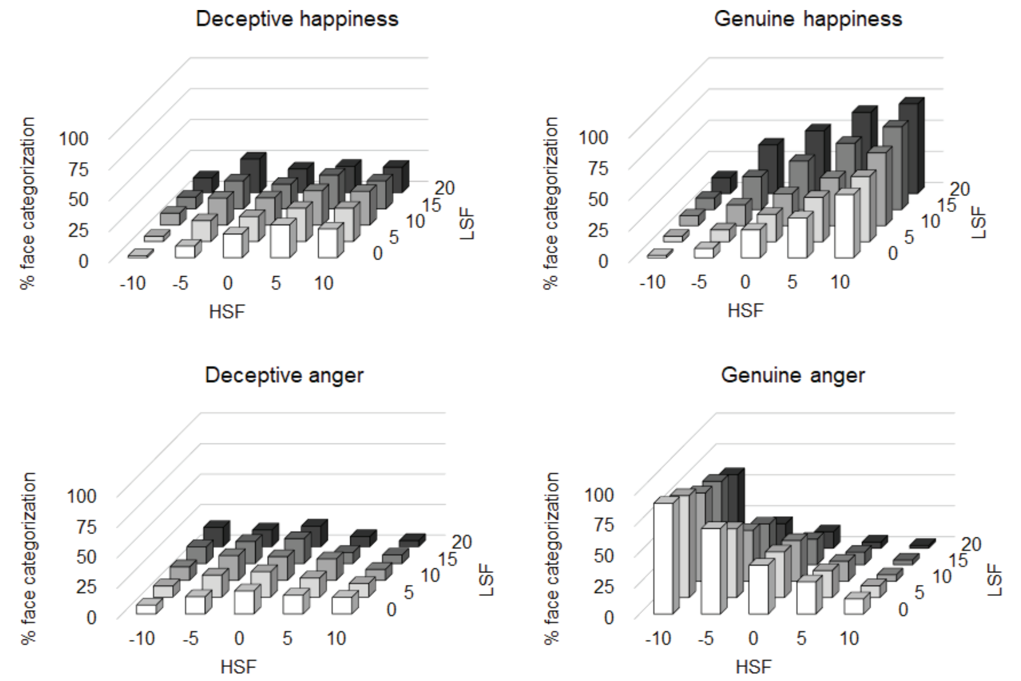


FIGURE 3 | Results of Experiment 2. Mean percentage of classification as the deceptive happiness (top-left panel), the genuine happiness (top-right panel), deceptive anger (bottom-left panel), and genuine anger (bottom-right panel).

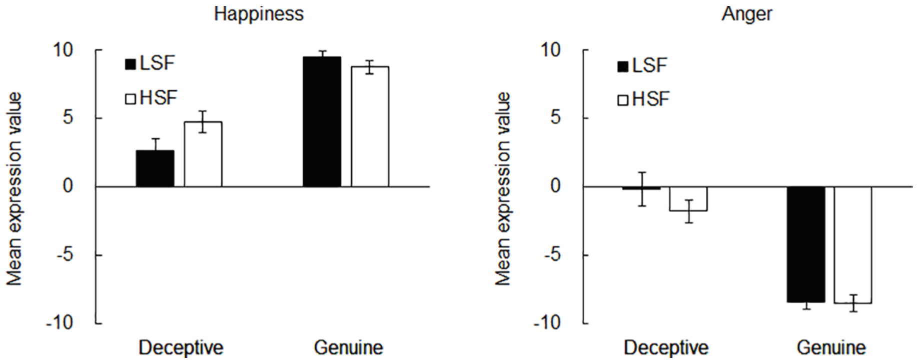


FIGURE 4 | Results of Experiment 2. Mean expression values of the mode for the deceptive and genuine images created by each frequency in the happiness (left panel) and anger (right panel) expressions. Error bars indicate standard errors of the mean.

However, there was no significant main effect of the spatial frequency, $F(1, 32) = 2.38$, $p > 0.13$, $\eta_p^2 = 0.07$, power = 0.87. The mean values (±SD) were 6.11 (±5.22) for LSF and 6.78 (±4.31) for HSF. Importantly, there was a significant interaction between the face and the spatial frequency, $F(1, 32) = 5.42$, $p < 0.03$, $\eta_p^2 = 0.14$, power = 1.00. *Post hoc* analysis using the Duncan test ($p < 0.05$) revealed that there were significant differences between all the conditions except between LSF and HSF in the genuine face condition. The mean values (± SD) were 2.68 (± 5.11) for LSF-deceptive, 4.77 (± 4.65) for HSF-deceptive, 9.55 (± 2.21) for LSF-genuine, and 8.79 (± 2.80) for HSF-genuine. The ANOVA for the angry faces revealed a significant main effect of the face, $F(1, 32) = 55.58$, $p < 0.001$, $\eta_p^2 = 0.63$, power = 1.00. The mean values (±SD) were −0.95 (±6.01) for deceptive and −8.45 (±3.30) for genuine. However, there was no significant main effect of the spatial frequency, $F(1, 32) = 1.04$, $p > 0.31$, $\eta_p^2 = 0.03$, power = 0.54. The mean values (±SD) were −4.28 (±6.78) for LSF and −5.11 (±5.41) for HSF. Also, there was no significant interaction, $F(1, 32) = 0.96$, $p > 0.33$, $\eta_p^2 = 0.03$, power = 0.64. The mean values (±SD) were −0.15 (±7.01) for LSF-deceptive, −1.74 (±4.78) for HSF-deceptive, −8.41 (±2.99) for LSF-genuine, and −8.48 (±3.64) for HSF-genuine. These results are consistent with those of Experiment 1. The finding suggests that deceptive happiness consisted of differential expression intensities for LSF and HSF does not depend on the task demands.

DISCUSSION

The spatial frequency components of faces provide critical clues for recognizing facial expressions (Farah et al., 1998; Ruiz-Soler and Beltran, 2006). However, it is not clear how we use such spatial frequency information in deceptive faces to interpret true emotion, and whether the contribution of LSF and HSF differs between deceptive happiness and anger facial expressions. To address these issues, we asked participants to generate deceptive and genuine faces by tuning the intensities of happiness and anger, which were contained in both LSF and HSF components (Experiment 1), and to classify presented faces composed of LSF and HSF images as either genuine happiness, genuine anger, deceptive happiness, or deceptive anger (Experiment 2). The results of the experiments show that deceptive happiness consists of differential intensities of expression between LSF and HSF, while deceptive anger consists of low LSF and HSF intensities. These results suggest that contribution of the LSF and HSF components are not equal when interpreting happy and angry deceptive faces.

The present study suggests that it is possible to discriminate deceptive happiness from a genuine one. This is because a deceptive happiness consists of unbalanced amounts of LSF and HSF expression, whereas a genuine happiness consists of approximately equal LSF and HSF intensities. In other words, detecting the unbalanced intensities of happiness expression between LSF and HSF allows us to be sensitive to hidden anger. Conversely, it must be difficult to distinguish between deceptive and genuine anger because both are represented by approximately equal LSF and HSF intensities. Although deceptive anger has lower intensities of both LSF and HSF expressions, there is no way to distinguish this from slight anger. In this case, other clues, such as facial movement, tone of voice, and/or contextual information, must be used when trying to interpret true emotion from anger facial expressions. Considering the fact that a high sensitivity for negative expressions has an adaptive function that promotes survival (Mathews et al., 1997), the visual system is likely biased toward hidden, as well as genuine, negative emotion. Based on this notion, LSF components play a key role in the sensitivity of interpreting hidden anger in deceptive happiness. Low-spatial frequency information about genuine anger facial expressions conveyed through the rapid magnocellular pathway reaches and activates the amygdala, a specific brain region for processing bias toward negative stimuli (Vuilleumier, 2005), which is essential for an adaptive function of quick risk aversion (Taylor, 1991). It is possible that the sensitivity to hidden anger in deceptive happiness found in this study is governed by the same visual pathway, although there is not yet empirical evidence for a relationship between amygdala activation and the processing of hidden anger facial expressions.

We adopted a face-generating task in Experiment 1 and a face-classification task in Experiment 2 and asked the participants to encode and decode facial expressions. Although these tasks examined different processes (i.e., encoding/decoding deceptive facial expressions), both tasks showed similar results with a trend for only deceptive happiness to show differential intensities in the expressions between LSF and HSF. These consistent results supported the notion that processing deceptive happiness depends on the balance between LSF and HSF components.

Note that we used only anger as a negative emotional expression in this study, although there are a variety of negative expressions, such as fear, disgust, and sadness. Regarding this, many studies have provided strong support for the idea that LSF components convey important information for processing of fear expressions (Vuilleumier et al., 2003; Winston et al., 2003; Pourtois et al., 2005; Vlamings et al., 2009; Bannerman et al., 2012; but see Holmes et al., 2005; Morawetz et al., 2011). It has also been suggested that LSF components of fear and disgusted expressions are related to non-conscious processing of negative expressions (Willenbockel et al., 2012). However, there are a few studies that demonstrate a relationship between HSF components and identification of grimacing (Deruelle et al., 2008) and sadness (Kumar and Srinivasan, 2011). Thus, we do not claim that the LSF component of deceptive faces is important for interpreting all hidden negative expressions. It is also possible that spatial frequency components higher than those used in this study could contain clues to identifying negative expressions. Obviously, further studies are required to investigate whether interpreting all types of hidden negative expressions is dependent on LSF components and that higher spatial frequency components contribute to discriminating between deceptive and genuine happiness.

Another limitation of the present study using artificial face models is that perceptual sensitivity to spatial frequency may differ between artificial and real faces. It has been reported that artificial facial models could give us different impression comparing with photos of real faces, although general tendencies to evaluate face images are similar (Crookes et al., 2015; Balas and Pacella, 2017; Balas et al., 2018; González-Álvarez and Cervera-Crespo, 2019). It is also unclear whether LSF and HSF components contain different facial expression when deceptive faces are made in real situations. Although our data clearly demonstrate that human observers have an ability to categorize deceptive happiness of artificial face models depending on the mismatch between LSF and HSF components, it is the first step to understand how spatial frequency information is used to identify real deceptive faces.

The results of this study are based only on male participants because of limitations in the subject pool that was available to us. Several studies have suggested that women are more sensitive to emotional faces than men (e.g., Kato and Takeda, 2017). However, many other studies have indicated that the gender of the participants does not affect the detection of emotions in hybrid facial images composed of emotional expressions in LSF and neutral expressions in HSF. For instance, Prete et al. (2014) demonstrated that female faces tend to be evaluated as more friendly than male faces regardless of LSF expression, whereas the friendliness ratings were not significantly different between male and female participants (see also Prete et al., 2015c, 2018a). Consequently, it is possible that female participants would also produce similar trends to those shown by the male participants in this study. Nevertheless, further work is needed to clarify the relationship between the gender of participants and the perception of hybrid facial expressions.

In conclusion, this study provides evidence that the LSF components of a deceptive happiness may allow us to interpret the true emotional state of anger. This finding indicates that we can understand another's hidden anger facial expression rapidly simply by using visual information from a static face, such as the unbalanced intensities of emotional expression between LSF and HSF. On the other hand, it is difficult to distinguish between genuine and deceptive anger from faces alone, suggesting that other clues need to be used to determine the true emotion. A high sensitivity for hidden anger facial expression could contribute to an adaptive function of risk aversion.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of "Guidelines for handling ergonomic experiments, Committee on Ergonomic Experiments, Bioethics and Biosafety Management Office, Safety Management Division, National Institute of Advanced Industrial Science and Technology"; with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the "Committee on Ergonomic Experiments, Bioethics and Biosafety Management Office, Safety Management Division, National Institute of Advanced Industrial Science and Technology."

AUTHOR CONTRIBUTIONS

KK and YT contributed to the conception and the design of the study. YT collected the data. KK wrote the first draft of the manuscript. KK and YT read and approved the final manuscript.

FUNDING

This work was supported by Grant-in-aid from the Japan Society for the Promotion of Science to KK (17K04494).

ACKNOWLEDGMENTS

We thank Linda Lillakas for many helpful comments on an earlier version of this article.

REFERENCES

Balas, B., and Pacella, J. (2017). Trustworthiness perception is disrupted in artificial faces. *Comput. Hum. Behav.* 77, 240–248. doi: 10.1016/j.chb.2017.08.045

Balas, B., Tupa, L., and Pacella, J. (2018). Measuring social variables in real and artificial faces. *Comput. Hum. Behav.* 88, 236–243. doi: 10.1016/j.chb.2018.07.013

Bannerman, R. L., Hibbard, P. B., Chalmers, K., and Sahraie, A. (2012). Saccadic latency is modulated by emotional content of spatially filtered face stimuli. *Emotion* 12, 1384–1392. doi: 10.1037/a0028677

Bar, M. (2004). Visual objects in context. *Nat. Rev. Neurosci.* 5, 617–629. doi: 10.1038/nrn1476

Butler, E. A., and Gross, J. J. (2004). "Hiding feelings in social contexts: out of sight is not out of mind" in *The regulation of emotion*. eds. P. Philippot, and R. S. Feldman (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 101–126.

Corneille, O., Hugenberg, K., and Potter, T. (2007). Applying the attractor field model to social cognition: perceptual discrimination is facilitated, but memory is impaired for faces displaying evaluatively congruent expressions. *J. Pers. Soc. Psychol.* 93, 335–352. doi: 10.1037/0022-3514.93.3.335

Crookes, K., Ewing, L., Gildenhuys, J.-D., Kloth, N., Hayward, W. G., Oxner, M., et al. (2015). How well do computer-generated faces tap face expertise? *PLoS One* 10:e0141353. doi: 10.1371/journal.pone.0141353

Deruelle, C., Rondan, C., Salle-Collemiche, X., Bastard-Rosset, D., and Da Fonséca, D. (2008). Attention to low- and high-spatial frequencies in categorizing facial identities, emotions and gender in children with autism. *Brain Cogn.* 66, 115–123. doi: 10.1016/j.bandc.2007.06.001

Ekman, P., and Friesen, W. V. (1982). Felt, false, and miserable smiles. *J. Nonverbal Behav.* 6, 238–252. doi: 10.1007/Bf00987191

Farah, M. J., Wilson, K. D., Drain, M., and Tanaka, J. N. (1998). What is "special" about face perception? *Psychol. Rev.* 105, 482–498. doi: 10.1037/0033-295X.105.3.482

Faul, F., Erdfelder, E., Buchner, A., and Lang, A. G. (2009). Statistical power analyses using G*power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149

Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G*power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., and Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception* 34, 77–86. doi: 10.1068/p5370

González-Álvarez, J., and Cervera-Crespo, T. (2019). Gender differences in sexual attraction and moral judgment: research with artificial face models. *Psychol. Rep.* 122, 525–535. doi: 10.1177/0033294118756891

Hass, N. C., Weston, T. D., and Lim, S.-L. (2016). Be happy not sad for your youth: the effect of emotional expression on age perception. *PLoS One* 11:e0152093. doi: 10.1371/journal.pone.0152093

Holmes, A., Winston, J. S., and Eimer, M. (2005). The role of spatial frequency information for ERP components sensitive to faces and emotional facial expression. *Cogn. Brain Res.* 25, 508–520. doi: 10.1016/j.cogbrainres.2005.08.003

Kato, R., and Takeda, Y. (2017). Females are sensitive to unpleasant human emotions regardless of the emotional context of photographs. *Neurosci. Lett.* 651, 177–181. doi: 10.1016/j.neulet.2017.05.013

Kauffmann, L., Ramanoel, S., and Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Front. Integr. Neurosci.* 8:37. doi: 10.3389/fnint.2014.00037

Kihara, K., and Takeda, Y. (2010). Time course of the integration of spatial frequency-based information in natural scenes. *Vis. Res.* 50, 2158–2162. doi: 10.1016/j.visres.2010.08.012

Kihara, K., and Takeda, Y. (2012). Attention-free integration of spatial frequency-based information in natural scenes. *Vis. Res.* 65, 38–44. doi: 10.1016/j.visres.2012.06.008

King, L. A. (1998). Ambivalence over emotional expression and reading emotions in situations and faces. *J. Pers. Soc. Psychol.* 74, 753–762. doi: 10.1037/0022-3514.74.3.753

Krumhuber, E. G., Skora, L., Küster, D., and Fou, L. (2016). A review of dynamic datasets for facial expression research. *Emot. Rev.* 9, 280–292. doi: 10.1177/1754073916670022

Kumar, D., and Srinivasan, N. (2011). Emotion perception is mediated by spatial frequency content. *Emotion* 11, 1144–1151. doi: 10.1037/a0025453

Laeng, B., Profeti, I., Saether, L., Adolfsdottir, S., Lundervold, A. J., Vangberg, T., et al. (2010). Invisible expressions evoke core impressions. *Emotion* 10, 573–586. doi: 10.1037/a0018689

Ledoux, J. E. (1995). Emotion: clues from the brain. *Annu. Rev. Psychol.* 46, 209–235. doi: 10.1146/annurev.ps.46.020195.001233

Leknes, S., Wessberg, J., Ellingsen, D. M., Chelnokova, O., Olausson, H., and Laeng, B. (2013). Oxytocin enhances pupil dilation and sensitivity to 'hidden' emotional expressions. *Soc. Cogn. Affect. Neurosci.* 8, 741–749. doi: 10.1093/scan/nss062

Mathews, A., Mackintosh, B., and Fulcher, E. P. (1997). Cognitive biases in anxiety and attention to threat. *Trends Cogn. Sci.* 1, 340–345. doi: 10.1016/S1364-6613(97)01092-9

Morawetz, C., Baudewig, J., Treue, S., and Dechent, P. (2011). Effects of spatial frequency and location of fearful faces on human amygdala activity. *Brain Res.* 1371, 87–99. doi: 10.1016/j.brainres.2010.10.110

Oosterhof, N. N., and Todorov, A. (2008). The functional basis of face evaluation. *Proc. Natl. Acad. Sci. USA* 105, 11087–11092. doi: 10.1073/pnas.0805664105

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

Porter, S., Ten Brinke, L., Baker, A., and Wallace, B. (2011a). Would I lie to you? "leakage" in deceptive facial expressions relates to psychopathy and emotional intelligence. *Personal. Individ. Differ.* 51, 133–137. doi: 10.1016/j.paid.2011.03.031

Porter, S., Ten Brinke, L., and Wallace, B. (2011b). Secrets and lies: involuntary leakage in deceptive facial expressions as a function of emotional intensity. *J. Nonverbal Behav.* 36, 23–37. doi: 10.1007/s10919-011-0120-7

Pourtois, G., Dan, E. S., Grandjean, D., Sander, D., and Vuilleumier, P. (2005). Enhanced extrastriate visual response to bandpass spatial frequency filtered fearful faces: time course and topographic evoked-potentials mapping. *Hum. Brain Mapp.* 26, 65–79. doi: 10.1002/hbm.20130

Prete, G., Capotosto, P., Zappasodi, F., Laeng, B., and Tommasi, L. (2015a). The cerebral correlates of subliminal emotions: an eleoencephalographic study with emotional hybrid faces. *Eur. J. Neurosci.* 42, 2952–2962. doi: 10.1111/ejn.13078

Prete, G., D'ascenzo, S., Laeng, B., Fabri, M., Foschi, N., and Tommasi, L. (2015b). Conscious and unconscious processing of facial expressions: evidence from two split-brain patients. *J. Neuropsychol.* 9, 45–63. doi: 10.1111/jnp.12034

Prete, G., Laeng, B., Fabri, M., Foschi, N., and Tommasi, L. (2015c). Right hemisphere or valence hypothesis, or both? The processing of hybrid faces in the intact and callosotomized brain. *Neuropsychologia* 68, 94–106. doi: 10.1016/j.neuropsychologia.2015.01.002

Prete, G., Laeng, B., and Tommasi, L. (2014). Lateralized hybrid faces: evidence of a valence-specific bias in the processing of implicit emotions. *Laterality* 19, 439–454. doi: 10.1080/1357650X.2013.862255

Prete, G., Laeng, B., and Tommasi, L. (2018a). Modulating adaptation to emotional faces by spatial frequency filtering. *Psychol. Res.* 82, 310–323. doi: 10.1007/s00426-016-0830-x

Prete, G., Laeng, B., and Tommasi, L. (2018b). Transcranial random noise stimulation (tRNS) over prefrontal cortex does not influence the evaluation of facial emotions. *Soc. Neurosci.*, 1–5. doi: 10.1080/17470919.2018.1546226

Ruiz-Soler, M., and Beltran, F. S. (2006). Face perception: an integrative review of the role of spatial frequencies. *Psychol. Res.* 70, 273–292. doi: 10.1007/s00426-005-0215-z

Schulte-Rüther, M., Markowitsch, H. J., Fink, G. R., and Piefke, M. (2007). Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: a functional magnetic resonance imaging approach to empathy. *J. Cogn. Neurosci.* 19, 1354–1372. doi: 10.1162/jocn.2007.19.8.1354

Schyns, P. G., and Oliva, A. (1997). Flexible, diagnosticity-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. *Perception* 26, 1027–1038. doi: 10.1068/p261027

Schyns, P. G., and Oliva, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition* 69, 243–265. doi: 10.1016/S0010-0277(98)00069-9

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: the mobilization-minimization hypothesis. *Psychol. Bull.* 110, 67–85. doi: 10.1037/0033-2909.110.1.67

Vlamings, P. H. J. M., Goffaux, V., and Kemner, C. (2009). Is the early modulation of brain activity by fearful facial expressions primarily mediated by coarse low spatial frequency information? *J. Vis.* 9, 12–12. doi: 10.1167/9.5.12

Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. *Trends Cogn. Sci.* 9, 585–594. doi: 10.1016/j.tics.2005.10.011

Vuilleumier, P., Armony, J. L., Driver, J., and Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nat. Neurosci.* 6, 624–631. doi: 10.1038/nn1057

Willenbockel, V., Lepore, F., Nguyen, D. K., Bouthillier, A., and Gosselin, F. (2012). Spatial frequency tuning during the conscious and non-conscious perception of emotional facial expressions – an intracranial ERP study. *Front. Psychol.* 3:237, 1–12. doi: 10.3389/fpsyg.2012.00237

Winston, J. S., Vuilleumier, P., and Dolan, R. J. (2003). Effects of low-spatial frequency components of fearful faces on fusiform cortex activity. *Curr. Biol.* 13, 1824–1829. doi: 10.1016/j.cub.2003.09.038

Xiao, W. S., Fu, G., Quinn, P. C., Qin, J., Tanaka, J. W., Pascalis, O., et al. (2015). Individuation training with other-race faces reduces preschoolers' implicit racial bias: a link between perceptual and social representation of faces in children. *Dev. Sci.* 18, 655–663. doi: 10.1111/desc.12241

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

PERSPECTIVE
published: 31 January 2020
doi: 10.3389/fpsyg.2020.00052

A Call for the Empirical Investigation of Tear Stimuli

Sarah J. Krivan[1,2]* and *Nicole A. Thomas*[2]

[1]*Department of Psychology, Applied Attention and Perceptual Processing Laboratory, College of Healthcare Sciences, James Cook University, Cairns, QLD, Australia,* [2]*Applied Attention and Perceptual Processing Laboratory, School of Psychological Sciences, Turner Institute for Brain and Mental Health, Monash University, Melbourne, VIC, Australia*

Emotional crying is a uniquely human behavior, which typically elicits helping and empathic responses from observers. However, tears can also be used to deceive. "Crocodile tears" are insincere tears used to manipulate the observer and foster prosocial responses. The ability to discriminate between genuine and fabricated emotional displays is critical to social functioning. When insincere emotional displays are detected, they are most often met with backlash. Conversely, genuine displays foster prosocial responses. However, the majority of crying research conducted to date has used posed stimuli featuring artificial tears. As such it is yet to be determined how the artificial nature of these displays impacts person perception. Throughout this article, we discuss the necessity for empirical investigation of the differences (or similarities) in responses to posed and genuine tearful expressions. We will explore the recent adoption of genuine stimuli in emotion research and review the existing research using tear stimuli. We conclude by offering suggestions and considerations for future advancement of the emotional crying field through investigation of both posed and genuine tear stimuli.

Keywords: tear effect, face perception, adult crying, emotion, interpersonal communication, crocodile tears

OPEN ACCESS

Edited by:
Shuo Wang,
West Virginia University,
United States

Reviewed by:
Chujun Lin,
California Institute of Technology,
United States
DongWon Oh,
New York University, United States
Ad Vingerhoets,
Tilburg University, Netherlands

****Correspondence:***
Sarah J. Krivan
sarah.krivan@my.jcu.edu.au

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

Received: *30 August 2019*
Accepted: *09 January 2020*
Published: *31 January 2020*

Citation:
Krivan SJ and Thomas NA (2020)
A Call for the Empirical
Investigation of Tear Stimuli.
Front. Psychol. 11:52.
doi: 10.3389/fpsyg.2020.00052

INTRODUCTION

Why do we cry? Emotional crying is a uniquely human display that has fascinated both scientists and lay people alike; this interest stems from an attempt to determine the functions of adult emotional tearing. A popular theory is that emotional tears serve a communicative function (Hendriks et al., 2008; Reed et al., 2015; Vingerhoets et al., 2016). Although tears have been readily touted as an honest signal of emotion (Trimble, 2012; Vingerhoets, 2013), there is a lack of empirical evidence to justify this perception. Furthermore, tears reliably elicit empathic responses (Lockwood et al., 2013) and social support from observers (Vingerhoets et al., 2016), while also signaling appeasement, which serves to reduce aggression (Hasson, 2009). While the presence of tears on a face can signal the need for social support, tears are also used to manipulate and deceive.

The accurate detection of emotional deception is critical to social functioning. Fake tears or "crocodile tears" are an insincere tearing display that evokers use to elicit sympathy and support from observers. How evokers produce these insincere tears is not yet known. Crocodile tears are typically associated with the disingenuous tears of celebrities and politicians (Manusov and Harvey, 2011), and conveying fabricated remorse during criminal court proceedings (ten Brinke et al., 2012). Alexander (2003) found that insincere narcissistic crying appears empty and orchestrated, and that witnessing this tearful display results in feeling uneasy and unmoved in a therapeutic environment. As such, insincere emotional displays elicit negative responses (Hideg and van Kleef, 2017) and reduced trust (Krumhuber et al., 2007). However, crocodile tears could also be elicited *via* deep acting where the evoker draws on previous experience in an effort to feel the emotion they are displaying (Lu et al., 2019). As such, tears elicited in this manner are driven by genuine feeling, however, are acted and physically effortful by nature. Given that tears are known to increase perceptions of remorse during apologies (Hornsey et al., 2019), and remorse is an important factor in sentencing and parole hearings (Bandes, 2016), further research exploring how we distinguish between sincere and crocodile tears is needed.

Despite the negative connotations associated with insincere emotion, most crying research has used standardized or posed faces featuring artificial tears. Although these studies have demonstrated that tears are responded to favorably (Balsters et al., 2013; Lockwood et al., 2013), how the artificial nature of these displays impacts person perception is yet to be determined. We call for the empirical investigation of the perception of both posed and genuine emotional tear displays. We first discuss the movement toward the adoption of genuine and ecologically valid stimuli in emotion research. Then, we explore research utilizing images of crying faces and highlight the advancements achieved and potential implications for the posed face methodology. Furthermore, we discuss recommendations for future research that highlight the perceptual differences between posed and genuine tearful displays. Finally, we conclude it is necessary to explore perceptions of genuine and disingenuous crying and believe posed and genuine stimuli can aid in this investigation.

GENUINE EMOTIONAL DISPLAYS

The recent movement toward using genuine expressions has predominantly stemmed from human ability to determine the genuineness of emotional displays (McLellan et al., 2010). Primarily, genuine expression research has investigated the difference between Duchenne and non-Duchenne smiles (Duchenne, 1862/1990; Ekman et al., 1990). Smiles are characterized by the activation of the zygomaticus major (i.e., the muscle responsible for drawing the corners of the mouth upward), while Duchenne smiles feature both zygomaticus and orbicularis oculi activation (i.e., the muscle associated with the crinkling of the eyes). Duchenne smiles are reliably judged as more intensely happy (Leppänen and Hietanen, 2007), and are mimicked more than non-Duchenne smiles (Krumhuber et al., 2014). Additionally, when mimicry is constrained, people are less accurate at recognizing emotional expressions (Oberman et al., 2007) and they display a reduced ability to discriminate between posed and genuine smiles (Rychlowska et al., 2014).

Compared to happiness, the literature exploring genuine displays of sadness is limited. Despite this reduced inquiry, findings are similar to smiling research. McLellan et al. (2010) confirmed that participants can discriminate between posed and genuine sadness equally as well as happiness. In a follow-up study, genuine happy and sad displays resulted in greater neural activation in brain regions associated with emotion recognition relative to posed expressions (McLellan et al., 2012). Applied research by Hackett et al. (2008) revealed that participants who expected rape victims to be emotionally expressive perceived crying victims to be more credible than non-criers. Given that Hornsey et al. (2019) found tearful apologies were more remorseful, it seems that viewer expectations about tears in negative displays are particularly important. Moving forward, research will need to encompass a wider variety of tearing stimuli to afford an understanding of how insincere crocodile tears are distinguished from genuine emotion.

Caveats associated with the use of genuine emotional stimuli stem from the time-consuming, labor-intensive demands of creating these displays, as well as less experimental control. For these reasons, some researchers have employed blended emotional displays where smiles are paired with eye-displays that feature expressions other than happiness (Gutiérrez-García and Calvo, 2015). Although this research has offered useful information about facial markers, these expressions are not authentic. As such, future investigations should explore whether people rely on facial markers to determine authenticity, or if they discriminate between shown and felt emotions. An interesting alternative to the caveats associated with the generation of genuine stimuli stems from a normative study by Dawel et al. (2017). While several posed facial databases, most notably the Pictures of Facial Affect database (Ekman and Friesen, 1976), were not perceived as showing genuine emotion, other posed facial expressions were perceived as genuine. Thus, posed perceived-as-genuine expressions offer a compromise to the difficulties associated with generating authentic stimuli, while allowing additional control. This advancement is particularly important for tear research, as it is currently unknown whether posed-tearful displays are perceived as perceptually genuine. As such, it is important that future investigations explore whether there are differences (or similarities) between the posed expressions typically used in crying research and genuine tearful stimuli.

THE ARTIFICIAL TEAR

Most existing research investigating the perception of emotional tearing uses posed facial expressions that feature artificial tears, added using eyedrops or digital enhancement (Reed et al., 2015; Ito et al., 2019). These artificial images have been used to explore how the presence of tears influences the perception of sadness (Hendriks et al., 2007; Ito et al., 2019), and the degree of helping behaviors elicited (Hendriks and Vingerhoets, 2006; Balsters et al., 2013; Lockwood et al., 2013). When images with visible tears are perceived as significantly sadder than the same image without tears, it is referred to as the *tear effect* (Provine et al., 2009).

In exploring perceptions of sadness, tears are typically added to sad and neutral faces, and various measures (e.g., reaction time, rating scales, and electroencephalography) are employed to examine how tears are perceived (Hendriks et al., 2007; Balsters et al., 2013). Balsters et al. (2013) demonstrated that even when brief presentations of tearful sad and neutral faces are shown, participants correctly categorize perceived sadness faster for sad expressions with tears, relative to sad and neutral tear-free expressions. However, contradictory evidence has been demonstrated when exploring the affective ratings of Duchenne smiles featuring tears. Reed et al. (2015) demonstrated that a tearful Duchenne smile was perceived as more intense than the tear-free counterpart. Furthermore, a trend toward increased happiness ratings was observed for the tearful Duchenne face. Thus, it is possible that Duchenne smiles signify genuine joy when they are accompanied by tears, akin to a dimorphous expression. Research exploring dimorphous event-elicited expressions of tearful-joy has identified that context is essential to the perception of tearful-joy as positive; as in the absence of context, the emotions were perceived as negative (Aragón, 2017). Thus, further work investigating whether posed happy-tear displays and genuine happy-tear displays are perceptually distinct is a worthy avenue of future research.

Recently, researchers have investigated whether the *tear effect* extends beyond sad, happy, and neutral expressions, as tears are elicited in response to a variety of emotions (Vingerhoets, 2013). Ito et al. (2019) concluded that the presence of tears on all negative emotions rendered them more perceptually similar to sadness, when examined in multidimensional space. Reed et al. (2015) further explored the *tear effect* using dynamic prototypical displays of anger, fear, disgust, sadness, and neutral expressions. An actress posed these expressions twice, once as traditional expressions, and once after using eyedrops to simulate tears. Importantly, no differences in the perceived authenticity of the displays were observed between tearful and non-tearful expressions. When examining intensity, valence, and emotion-specific ratings, further generalized support was demonstrated for the *tear effect* and the role of tears as a marker of sadness. Although Reed et al. (2015) found no perceptual differences in authenticity between tearful and non-tearful expressions, no other study has considered the influence of perceived genuineness. However, people are able to distinguish between posed and genuine sadness (McLellan et al., 2010). Thus, further research is needed to determine whether people are able to distinguish between posed and genuine tearful displays.

In the context of our everyday lives, it is of interest to understand the relationship between tears, emotional support, and empathy. There is consensus that tears elicit greater emotional support and empathy compared to tear-free expressions (Hendriks and Vingerhoets, 2006; Balsters et al., 2013; Lockwood et al., 2013). Hendriks and Vingerhoets (2006) concluded that tearful expressions elicit greater support and reduced avoidance behaviors relative to other emotional displays. Furthermore, tears elicited greater perceived personal distress. Thus, despite participants' belief that encountering a tearful person would increase their own distress, they still reported greater helping responses to tears. Lockwood et al. (2013) further explored the role of empathy in response to emotional crying using reaction time. Participants responded to neutral and caregiving words after witnessing subliminally presented emotional face primes of happy, sad, and crying faces. Individuals high in cognitive empathy showed no differences in response time. However, individuals low in cognitive empathy were slower to respond to caregiving words after being primed with a crying face, but not after sad or happy expressions. Thus, the level of empathy experienced by the observer also influences how individuals respond to crying persons. Collectively, these studies demonstrate that posed facial expressions elicit empathic responses; however, they neglect to explore the role of empathy in responding to genuine versus posed displays.

To adopt more ecologically valid crying stimuli, researchers have used crying photographs from the image-sharing site Flickr. Selecting crying photographs allows for the investigation of the *tear effect* using the inverse of the artificial tear addition technique. Provine et al. (2009) were the first to demonstrate the *tear effect* using Flickr images that included tears, which were digitally removed to create tear-free duplicates. Takahashi et al. (2015) also used the tear removal paradigm in an fMRI study investigating the perception of tears on sad and neutral expressions. The *tear effect* for sad expressions featuring tears was replicated, and they further concluded that the *tear effect* was larger for neutral faces than sad faces. As tears serve as a salient marker of sadness, their presence resolved the ambiguity of the neutral faces.

Although these studies used stimuli with greater ecological validity, it is impossible to tell whether their images were perceived as authentic expressions of emotion by the participants. As Flickr is a website where people primarily upload their own images to share with friends and followers, images shared to the platform are likely posed and self-selected by the individuals to present themselves in a positive manner (Angus and Thelwall, 2010; Malinen, 2010). Thus, posed datasets allowed for the investigation of the perception of tears with rigorous control (Balsters et al., 2013; Lockwood et al., 2013), and stimuli with greater ecological validity have replicated these effects (Provine et al., 2009; Takahashi et al., 2015); however, the need for research using genuine tearful expressions remains.

THE GENUINE TEAR

Recently, researchers have begun to use photographic stimuli featuring emotional tearing, which were captured in a moment of genuine emotional experience. These images were captured during the Museum of Modern Art, Artist is Present exhibit, where nearly 1,000 people sat with Marina Abramović and cried during the experience. As these tears were elicited in a moment of genuine emotion, these studies have investigated the perceived warmth and competence of the crying persons (van de Ven et al., 2017; Zickfeld et al., 2018; Zickfeld and Schubert, 2018), as well as the perceived social-connectedness and willingness to provide help to crying persons (Vingerhoets et al., 2016; Stadel et al., 2019). The original study by van de Ven et al. (2017) concluded that tearful individuals were perceived as warmer, though less competent, than tear-free individuals. Two replications of this study also determined that tearful individuals were perceived as warmer; however, neither study replicated the reduced competence effect when using a larger sample of target crying faces (Zickfeld et al., 2018; Zickfeld and Schubert, 2018). Zickfeld et al. (2018) concluded that the competence effect from the original study was likely target specific, and thus the presence of tears is unlikely to alter perceptions of competence.

Importantly, the work conducted using genuine tear stimuli has also replicated the findings that emotional tears elicit support and willingness to help. Vingerhoets et al. (2016) concluded that participants attribute greater helping behaviors to individuals with tears, than without tears. Furthermore, through mediation analysis it was determined that helping behaviors stemmed from a perception of closeness to the individuals in the crying images. Similarly, tearful stimuli facilitate approach behaviors relative to avoidance (Riem et al., 2017; Gračanin et al., 2018). Furthermore, Stadel et al. (2019) identified that participants show increased willingness to help individuals with tears, and concluded that this willingness was the strongest between female and mixed dyads, compared to male dyads. Therefore, it seems that tears are a signal that elicits helping responses from observers; however, both the gender of the participant and the expressor might mediate the degree of assistance offered. Future research should expand upon these findings, which stem from self-report willingness to help measures, to better encompass whether perception is aligned with actual helping behavior. Additionally, while these stimuli were captured during a moment of genuine experience, it is unknown what the individuals were feeling. Aragón and Clark (2018) explored responses to genuine dimorphous happy tears. Participants reported a greater likelihood of down-regulation responses to tearful-joy, than joy expressed with smiles. Thus, future research needs to consider the role that emotional state plays in establishing the way that we respond to tears.

TABLE 1 | A comparison of the effect sizes reported in published studies examining tears.

Authors	Stimulus type	Tear method	Effect size
Faster reaction time to tearful images			
Balsters et al. (2013)	KDEF	Digitally added	$\eta^2 = 0.284^{\dagger}$
Gračanin et al. (2018)	MoMA	Digitally removed	$\eta_p^2 = 0.26^{\dagger}$
Riem et al. (2017)	MoMA	Digitally removed	$\eta_p^2 = 0.69^{\dagger}$
Greater perceived sadness for tearful images			
Provine et al. (2009)	Flickr tear images	Digitally removed	$\eta^2 = 0.26^{\dagger}$
Takahashi et al. (2015)	Flickr tear images	Digitally removed	$\eta_p^2 = 0.793^*$
Reed et al. (2015)	Female actress using FACS	Eyedrops	$d = 0.22$
Ito et al. (2019)	TFEID	Digitally added	$\eta_p^2 = 0.073$
van de Ven et al. (2017)	MoMA	Digitally removed	$\eta_p^2 = 0.15^{\dagger}$
Zickfeld et al. (2018)	MoMA	Digitally removed	$d = 0.86$
Greater willingness to help/greater perceived support for tearful images			
Balsters et al. (2013)	KDEF	Digitally added	$\eta^2 = 0.375^{\dagger}$
Vingerhoets et al. (2016)	MoMA	Digitally removed	$d = 0.85–1.32$
Zickfeld and Schubert (2018)	MoMA	Digitally removed	$d_S = 0.70–0.82$

*KDEF, Karolinska Directed Emotional Faces; MoMA, Genuine tear expressions captured during Museum of Modern Art Performance; TFEID, Taiwanese Facial Expression Image Database; Flickr tear images, images of tearful individuals found on Flickr (unknown if genuine or posed). Effect sizes are reported as in the published papers.*Denotes that original paper did not report effect size, and thus it was estimated from main effect of tears; †Denotes effect size from main effect.*

DISCUSSION

To date, research using images of teary expressions has focused on expressions of sadness and the anticipated perception and response of individuals. Although crying research has recently adopted the use of genuine expressions, there is no empirical evidence exploring differences in perceived authenticity between posed and genuine displays of emotion featuring tears. **Table 1** provides a collation of the studies examining the tear effect, and the influence that tears have on empathic responses. This table highlights the type of stimuli used in each experiment, the method of tear addition or removal, and the effect sizes reported in the published literature. It must be noted that the type of task, the number of identities used, and the gender of the stimuli varied widely across these studies. This variability further highlights the need for empirical studies using both posed and genuine tearful expressions. This empirical investigation will assist with better understanding the perceptual differences between tear stimuli and aid in our understanding of how we discriminate genuine and posed emotion.

Furthermore, as the majority of the work conducted to date has used posed expressions, there has been limited focus on the other facial responses that accompany emotional tears, including blotchy faces and bloodshot eyes (Provine et al., 2011, 2013). Küster (2018) explored the influence of tears and pupil size on the perception of sadness using digital avatars. While both the presence of tears and smaller pupil sizes increased perceived sadness, there was no interaction effect between tears and pupil size. The inverse consideration of the extreme features accompanying emotional crying is the perceptual and affective differences between tearing up and crying uncontrollably (i.e., ugly crying). Research using vignettes has demonstrated that the intensity of tears moderates observer reactions, where in some scenarios just tearing up may elicit more positive responses than weeping (Wong et al., 2011). Thus, further work in this field should explore the relationship between the intensity of the tears and observer responses. It may be that assistance for emotional crying is curvilinear, where there is an optimum level of tearing that elicits helping responses from observers.

Finally, the adoption of investigative techniques like psychophysiology may offer insight into the perceptions of tears to further corroborate the results from self-report studies. Recently, mirror neurons have been proposed as a mechanism for sharing others' emotional states, with "feeling" and "perceiving" emotion sharing neural substrates (Wicker et al., 2003; Singer et al., 2004). Similarly, facial mimicry studies have identified that when participants' ability to mimic is impaired, they show reduced emotion recognition abilities (Oberman et al., 2007;

Rychlowska et al., 2014). In addition, examination of other physiological techniques, such as eye-tracking and galvanic skin response, may yield fruitful information about the features that individuals attend to in decoding an emotional face, and the degree of arousal that tearful expressions elicit. Analysis of the arousal response may assist in determining the motivation for the helping behaviors as a metric of personal distress. Furthermore, the inclusion of psychophysiological metrics allows for greater certainty in the true nature of the self-report responses.

In this paper, we have reviewed recent work using facial expressions as a means of investigating inter-individual functions of crying. Reviewing these studies has revealed that the use of posed expressions has afforded an understanding of the communicative functions of emotional tears by employing rigorously controlled stimuli between conditions. In addition, the use of genuine expressions of emotion in more recent crying research has replicated findings that both posed and genuine expressions of emotion are effective at eliciting support and attention. However, whether posed tearful expressions are being treated as perceptually authentic, or if their staged nature is impacting upon person perception is yet to be determined. Thus, to continue advancing the understanding about the interpersonal functions of human emotional tearing, we need to adopt an approach that better explores how we perceive both genuine and non-genuine crying expressions. This advancement needs to encompass a greater range of tearing stimuli to allow for the exploration of the physiological effects that accompany emotional tearing. This research will provide a basis for understanding the type of emotional tears we respond to. People are able to distinguish between posed and genuine emotions, yet tears have not received this same inquiry. Determining how we distinguish between posed and genuine tearful expressions will aid in further understanding the functions of this uniquely human phenomenon.

AUTHOR CONTRIBUTIONS

SK conceptualized and designed the article and wrote the first draft of the manuscript. SK and NT contributed to manuscript revision, read and approved the submitted version.

FUNDING

This research is supported by an Australian Government Research Training Program (RTP) Scholarship awarded to SK, and an Australian Research Council (ARC) Discovery Early Career Research Award (DECRA) awarded to NT (DE150101108).

REFERENCES

Alexander, T. (2003). Narcissism and the experience of crying. *Br. J. Psychother.* 20, 27–38. doi: 10.1111/j.1752-0118.2003.tb00112.x

Angus, E., and Thelwall, M. (2010). "Motivations for image publishing and tagging on Flickr" in *Proceedings of the 14th International Conference on Electronic Publishing*. eds. T. Hedlund and Y. Tonta (Helsinki, Finland: Hanken School of Economics), 189–204.

Aragón, O. R. (2017). "Tears of joy" and "tears and joy?" personal accounts of dimorphous and mixed expressions of emotion. *Motiv. Emot.* 41, 370–392. doi: 10.1007/s11031-017-9606-x

Aragón, O. R., and Clark, M. S. (2018). "Tears of joy" & "smiles of joy" prompt distinct patterns of interpersonal emotion regulation. *Cognit. Emot.* 32, 913–940. doi: 10.1080/02699931.2017.1360253

Balsters, M. J. H., Krahmer, E. J., Swerts, M. G. J., and Vingerhoets, A. J. J. M. (2013). Emotional tears facilitate the recognition of sadness and the perceived need for social support. *Evol. Psychol.* 11, 148–158. doi: 10.1177/147470491301100114

Bandes, S. A. (2016). Remorse and criminal justice. *Emot. Rev.* 8, 14–19. doi: 10.1177/1754073915601222

Dawel, A., Wright, L., Irons, J., Dumbleton, R., Palermo, R., O'Kearney, R., et al. (2017). Perceived emotion genuineness: normative ratings for popular facial expression stimuli and the development of perceived-as-genuine and perceived-as-fake sets. *Behav. Res. Methods* 49, 1539–1562. doi: 10.3758/s13428-016-0813-2

Duchenne, B. (1862/1990). *The mechanism of human facial expression*. New York, NY: Cambridge University Press (Original work published 1862).

Ekman, P., Davidson, R., and Friesen, W. V. (1990). The Duchenne smile: emotional expression and brain physiology: II. *J. Pers. Soc. Psychol.* 58, 342–353. doi: 10.1037/0022-3514.58.2.342

Ekman, P., and Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto: Consulting Psychologists Press.

Gračanin, A., Krahmer, E., Rinck, M., and Vingerhoets, A. J. J. M. (2018). The effects of tears on approach–avoidance tendencies in observers. *Evol. Psychol.* 16:1474704918791058. doi: 10.1177/1474704918791058

Gutiérrez-García, 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

Hackett, L., Day, A., and Mohr, P. (2008). Expectancy violation and perceptions of rape victim credibility. *Leg. Criminol. Psychol.* 13, 323–334. doi: 10.1348/135532507X228458

Hasson, O. (2009). Emotional tears as biological signals. *Evol. Psychol.* 7, 363–370. doi: 10.1177/147470490900700302

Hendriks, M., Croon, M. A., and Vingerhoets, A. J. J. M. (2008). Social reactions to adult crying: the help-soliciting function of tears. *J. Soc. Psychol.* 148, 22–42. doi: 10.3200/SOCP.148.1.22-42

Hendriks, M. C., van Boxtel, G. J., and Vingerhoets, A. J. J. M. (2007). An event-related potential study on the early processing of crying faces. *NeuroReport* 18, 631–634. doi: 10.1097/WNR.0b013e3280bad8c7

Hendriks, M., and Vingerhoets, A. J. J. M. (2006). Social messages of crying faces: their influence on anticipated person perception, emotions and behavioural responses. *Cognit. Emot.* 20, 878–886. doi: 10.1080/02699930500450218

Hideg, I., and van Kleef, G. A. (2017). When expressions of fake emotions elicit negative reactions: the role of observers' dialectical thinking. *J. Organ. Behav.* 38, 1196–1212. doi: 10.1002/job.2196

Hornsey, M. J., Wohl, M. J. A., Harris, E. A., Okimoto, T. G., Thai, M., and Wenzel, M. (2019). Embodied remorse: physical displays of remorse increase positive responses to public apologies, but have negligible effects on forgiveness. *J. Pers. Soc. Psychol.* doi: 10.1037/pspi0000208 [Epub ahead of print].

Ito, K., Ong, C. W., and Kitada, R. (2019). Emotional tears communicate sadness but not excessive emotions without other contextual knowledge. *Front. Psychol.* 10:878. doi: 10.3389/fpsyg.2019.00878

Krumhuber, E. G., Likowski, K. U., and Weyers, P. (2014). Facial mimicry of spontaneous and deliberate Duchenne and non-Duchenne smiles. *J. Nonverbal Behav.* 38, 1–11. doi: 10.1007/s10919-013-0167-8

Krumhuber, E., Manstead, A. S. R., Cosker, D., Marshal, D., Rosin, P. L., and Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion* 7, 730–735. doi: 10.1037/1528-3542.7.4.730

Küster, D. (2018). Social effects of tears and small pupils are mediated by felt sadness: an evolutionary view. *Evol. Psychol.* 16:1474704918761104. doi: 10.1177/1474704918761104

Leppänen, J. M., and Hietanen, J. K. (2007). Is there more in a happy face than just a big smile? *Vis. Cogn.* 15, 468–490. doi: 10.1080/13506280600765333

Lockwood, P., Millings, A., Hepper, E., and Rowe, A. C. (2013). If I cry, do you care? Individual differences in empathy moderate the facilitation of caregiving words after exposure to crying faces. *J. Individ. Differ.* 34, 41–47. doi: 10.1027/1614-0001/a000098

Lu, Y., Wu, W., Mei, G., Zhao, S., Zhou, H., Li, D., et al. (2019). Surface acting or deep acting, who need more effortful? A study on emotional labor using functional near-infrared spectroscopy. *Front. Hum. Neurosci.* 13:151. doi: 10.3389/fnhum.2019.00151

Malinen, S. (2010). "Photo exhibition or online community? The role of social interaction in Flickr" in *Paper presented at the 2010 Fifth International Conference on Internet and Web Applications and Services* (Barcelona: IEEE).

Manusov, V., and Harvey, J. (2011). Bumps and tears on the road to the presidency: media framing of key nonverbal events in the 2008 democratic election. *West. J. Commun.* 75, 282–303. doi: 10.1080/10570314.2011.571650

McLellan, T., Johnston, L., Dalrymple-Alford, J., and Porter, R. (2010). Sensitivity to genuine versus posed emotion specified in facial displays. *Cognit. Emot.* 24, 1277–1292. doi: 10.1080/02699930903306181

McLellan, T. L., Wilcke, J. C., Johnston, L., Watts, R., and Miles, L. K. (2012). Sensitivity to posed and genuine displays of happiness and sadness: a fMRI study. *Neurosci. Lett.* 531, 149–154. doi: 10.1016/j.neulet.2012.10.039

Oberman, L. M., Winkielman, P., and Ramachandran, V. S. (2007). Face to face: blocking facial mimicry can selectively impair recognition of emotional expressions. *Soc. Neurosci.* 2, 167–178. doi: 10.1080/17470910701391943

Provine, R. R., Cabrera, M. O., Brocato, N. W., and Krosnowski, K. A. (2011). When the whites of the eyes are red: a uniquely human cue. *Ethology* 117, 395–399. doi: 10.1111/j.1439-0310.2011.01888.x

Provine, R. R., Krosnowski, K. A., and Brocato, N. W. (2009). Tearing: breakthrough in human emotional signaling. *Evol. Psychol.* 7, 52–56. doi: 10.1177/147470490900700107

Provine, R. R., Nave-Blodgett, J., and Cabrera, M. O. (2013). The emotional eye: red sclera as a uniquely human cue of emotion. *Ethology* 119, 993–998. doi: 10.1111/eth.12144

Reed, L. I., Deutchman, P., and Schmidt, K. L. (2015). Effects of tearing on the perception of facial expressions of emotion. *Evol. Psychol.* 13, 1–5. doi: 10.1177/1474704915613915

Riem, M. M. E., van Ijzendoorn, M. H., De Carli, P., Vingerhoets, A. J. J. M., and Bakermans-Kranenburg, M. J. (2017). Behavioral and neural responses to infant and adult tears: the impact of maternal love withdrawal. *Emotion* 17, 1021–1029. doi: 10.1037/emo0000288

Rychlowska, M., Cañadas, E., Wood, A., Krumhuber, E. G., Fischer, A., and Niedenthal, P. M. (2014). Blocking mimicry makes true and false smiles look the same. *PLoS One* 9:e90876. doi: 10.1371/journal.pone.0090876

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., and Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science* 303, 1157–1162. doi: 10.1126/science.1093535

Stadel, M., Daniels, J. K., Warrens, M. J., and Jeronimus, B. F. (2019). The gender-specific impact of emotional tears. *Motiv. Emot.* 43, 696–704. doi: 10.1007/s11031-019-09771-z

Takahashi, H. K., Kitada, R., Sasaki, A. T., Kawamichi, H., Okazaki, S., Kochiyama, T., et al. (2015). Brain networks of affective mentalizing revealed by the tear effect: the integrative role of the medial prefrontal cortex and precuneus. *Neurosci. Res.* 101, 32–43. doi: 10.1016/j.neures.2015.07.005

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

Trimble, M. (2012). *Why humans like to cry: Tragedy, evolution, and the brain*. Oxford, UK: Oxford University Press.

van de Ven, N., Meijs, M. H. J., and Vingerhoets, A. J. J. M. (2017). What emotional tears convey: tearful individuals are seen as warmer, but also as less competent. *Br. J. Soc. Psychol.* 56, 146–160. doi: 10.1111/bjso.12162

Vingerhoets, A. J. J. M. (2013). *Why only humans weep: Unravelling the mysteries of tears*. Oxford, England: Oxford University Press.

Vingerhoets, A. J. J. M., van de Ven, N., and Velden, Y. (2016). The social impact of emotional tears. *Motiv. Emot.* 40, 455–463. doi: 10.1007/s11031-016-9543-0

Wicker, B., Keysers, C., Plailly, J., Royet, J.-P., Gallese, V., and Rizzolatti, G. (2003). Both of us disgusted in my insula: the common neural basis of seeing and feeling disgust. *Neuron* 40, 655–664. doi: 10.1016/S0896-6273(03)00679-2

Wong, Y. J., Steinfeldt, J. A., LaFollette, J. R., and Tsao, S. C. (2011). Men's tears: football players' evaluations of crying behavior. *Psychol. Men Masculinity* 12, 297–310. doi: 10.1037/a0020576

Zickfeld, J. H., and Schubert, T. W. (2018). Warm and touching tears: tearful individuals are perceived as warmer because we assume they feel moved and touched. *Cognit. Emot.* 32, 1691–1699. doi: 10.1080/02699931.2018.1430556

Zickfeld, J. H., van de Ven, N., Schubert, T. W., and Vingerhoets, A. J. J. M. (2018). Are tearful individuals perceived as less competent? Probably not. *Compr. Results Soc. Psychol.* 3, 119–139. doi: 10.1080/23743603.2018.1514254

Conflict of Interest: 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: 29 April 2020
doi: 10.3389/fnins.2020.00329

Brain Activation in Contrasts of Microexpression Following Emotional Contexts

*Ming Zhang[1], Ke Zhao[1], Fangbing Qu[1,2], Kaiyun Li[1,3] and Xiaolan Fu[1,4]**

[1] Institute of Psychology, Chinese Academy of Sciences, Beijing, China, [2] College of Preschool Education, Capital Normal University, Beijing, China, [3] School of Education and Psychology, University of Jinan, Jinan, China, [4] Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

The recognition of microexpressions may be influenced by emotional contexts. The microexpression is recognized poorly when it follows a negative context in contrast to a neutral context. Based on the behavioral evidence, we predicted that the effect of emotional contexts might be dependent on neural activities. Using the synthesized microexpressions task modified from the Micro-Expression Training Tool (METT), we performed an functional MRI (fMRI) study to compare brain response in contrasts of the same targets following different contexts. Behaviorally, we observed that the accuracies of target microexpressions following neutral contexts were significantly higher than those following negative or positive contexts. At the neural level, we found increased brain activations in contrasts of the same targets following different contexts, which reflected the discrepancy in the processing of emotional contexts. The increased activations implied that different emotional contexts might differently influence the processing of subsequent target microexpressions and further suggested interactions between the processing of emotional contexts and of microexpressions.

Keywords: emotion context, microexpression, recognition, activation, fMRI

OPEN ACCESS

Edited by:
Shuo Wang,
West Virginia University, United States

Reviewed by:
Zhongzheng Fu,
California Institute of Technology,
United States
Zhihao Zhang,
University of California, Berkeley,
United States

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

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

Received: *10 September 2019*
Accepted: *20 March 2020*
Published: *29 April 2020*

Citation:
Zhang M, Zhao K, Qu F, Li K and Fu X (2020) Brain Activation in Contrasts of Microexpression Following Emotional Contexts. Front. Neurosci. 14:329. doi: 10.3389/fnins.2020.00329

INTRODUCTION

As we know, emotional information always affects the recognition of subsequent facial expression and then exerts an important context effect (Wieser and Brosch, 2012). It will facilitate the recognition of subsequent facial expressions if they convey the same emotional components (Werheid et al., 2005). For example, anger is recognized more accurately following a negative context, whereas happiness is recognized better following a positive context (Hietanen and Astikainen, 2013). The microexpression, as a quick facial expression, generally lasts for 1/25 to 1/5 s and occurs in the flow of facial expressions, especially when individuals try to repress or conceal their true emotions (Ekman, 2009). Its recognition is influenced by emotional stimuli (e.g., facial expression) appearing before and after the microexpressions (i.e., the emotional contexts). Microexpressions are recognized poorly when they followed a negative context, regardless of the duration of the microexpressions (Zhang et al., 2014). However, existing studies provide limited behavioral evidence for the presence of an effect of emotional context in microexpression recognition (Zhang et al., 2014, 2018). In order to recognize microexpressions more accurately in realistic emotional contexts, further evidence for the neural basis of the effect deriving from the emotional context is necessary.

The effect of emotional contexts on the perception of facial expressions could be reflected in neural activations (Schwarz et al., 2013). The facial expression alone generally activated visual processing regions, yet the facial expression with a context was more associated with social and emotional processing regions (Lee and Siegle, 2014). Emotional contexts including some affective stimuli could influence cerebral cortex reactions, altering activation regions or activation levels. Facial expressions conveying specific emotions engage specific brain areas, such as the medial prefrontal cortex, the fusiform gyrus, the superior temporal gyrus, the parahippocampal gyrus, the insula, the precuneus, the inferior parietal, and the amygdala (Haxby et al., 2000; Heberlein et al., 2008). Moreover, brain activities related to facial expressions are not always clear cut and are strongly influenced by the emotional context. Facial expressions will be interpreted differently in various emotional contexts (Schwarz et al., 2013). By presenting target (fearful/neutral) faces against the background of threatening or neutral scenes, Van den Stock et al. (2014) found that the emotional valence of contexts modulates the processing of faces in the right anterior parahippocampal area and subgenual anterior cingulate cortex, which showed higher activations for targets in neutral contexts compared to those in threatening contexts. In addition, response inhibitions coming from the interaction of facial expressions and preceding contexts were observed in the left insula cortex and right inferior frontal gyrus (Schulz et al., 2009). Consistent with these accounts, brain responses to ambiguous facial expressions (surprise) were found to be modified by contextual conditions—that is, activations (especially in the amygdala) were stronger for surprised faces embedded in negative contexts compared to those in positive contexts (Kim et al., 2004). These findings showed that the perception of facial expression is modulated by contextual information, reflecting context-dependent neural processing of facial expressions.

In view of the effect of emotional context on the brain's responses to facial expressions, microexpressions should be influenced by emotional contexts. Behavioral evidence for the effect of emotional contexts on microexpression recognition leads us to believe that the effect of emotional contexts should depend on neural activities. The present fMRI study focused on brain activation in contrasts of the microexpression following different emotional contexts and aimed to provide neural evidence for the potential effect of emotional context on microexpressions. The previous study showed that emotion recognition is modulated by a distributed neural system (Zhao et al., 2017). The process of emotion recognition involves increased activity in visual areas (e.g., fusiform gyrus), limbic areas (e.g., parahippocampal gyrus and amygdala), temporal areas (e.g., superior temporal gyrus and middle temporal gyrus), and prefrontal areas (e.g., medial frontal gyrus and middle frontal gyrus) (Haxby et al., 2000; Heberlein et al., 2008). Based on these findings, it is reasonable to predict that contrasts of the same targets following different contexts will elicit different patterns of increased brain activity. This study adopted a synthesized task modified from the Micro-Expression Training Tool (METT) to simulate a microexpression (Ekman, 2002; Shen et al., 2012) and compared the brain activations of contrasts of the same targets following different contexts.

MATERIALS AND METHODS

Participants

Twenty-one healthy, right-handed undergraduates (age ranged from 18 to 23, $M = 20.90$, $SD = 1.37$; 11 females) with normal or corrected-to-normal vision participated in our fMRI study and were compensated for their participation. Before entering the MRI scanner, they completed a questionnaire provided by the Southwest University MRI Centre that required all individuals to report honestly their current health status and medical records, including physical injuries and mental disorders. No participant reported a neurological or psychiatric history. Written informed consent to participate was obtained, and participants were informed of their right to discontinue participation at any time. The experimental protocol was approved by the Institutional Review Board of the Institute of Psychology at the Chinese Academy of Sciences. All procedures were conducted according to the Declaration of Helsinki.

Stimuli

The materials, including 120 images (20 models, 10 females), were adapted from a previous study (Zhang et al., 2018). The visual stimuli were presented via a video projector (frequency, 60 Hz; resolution, 1,024 × 768; frame-rate, ~16.7 ms) onto a rear-projection screen mounted at the head of the scanner bore (see stimuli samples in **Figure 1**). Participants viewed the stimuli through a mirror on the head coil positioned over their eyes. All the stimuli (visual angle, 11.8° × 15.1°) were displayed on a uniform silver background.

Procedure

The task was adopted from the previous study (Zhang et al., 2014) and was modified for the present fMRI experiment. Both stimulus presentation and behavioral response collection were controlled by E-Prime 2.0 (Psychological Software Tools, Inc., Pittsburgh, PA, United States). Participants performed a practice experiment outside the MRI, using the same procedure as the real experiment. There were four sessions in total, each lasting 8 min. Each session included nine experimental conditions, in which three emotional contexts (negative, neutral, and positive) and three target microexpressions (anger, neutral, and happiness) were randomly combined, as well as a blank condition. All of these conditions were repeated eight times—that is, each of these nine conditions was repeated 32 times in total in four sessions. The trial sequence in each session was randomized with a trial time of 6 s.

Each trial proceeded as follows (see **Figure 1**). First, a black fixation cross was presented for 500 ms, followed by either an angry, neutral, or happy expression context (all with closed mouth) for 2,000 ms (119 frames). Subsequently, one of the three target microexpressions (anger, happiness, or neutral, all with open mouth, the same model as in the forward context) was presented for 60 ms (four frames). Then, the same expression context was presented for 2,000 ms (119 frames) again. After that, the task instructions were presented, and participants were asked to recognize the fleeting expression by pressing one of the

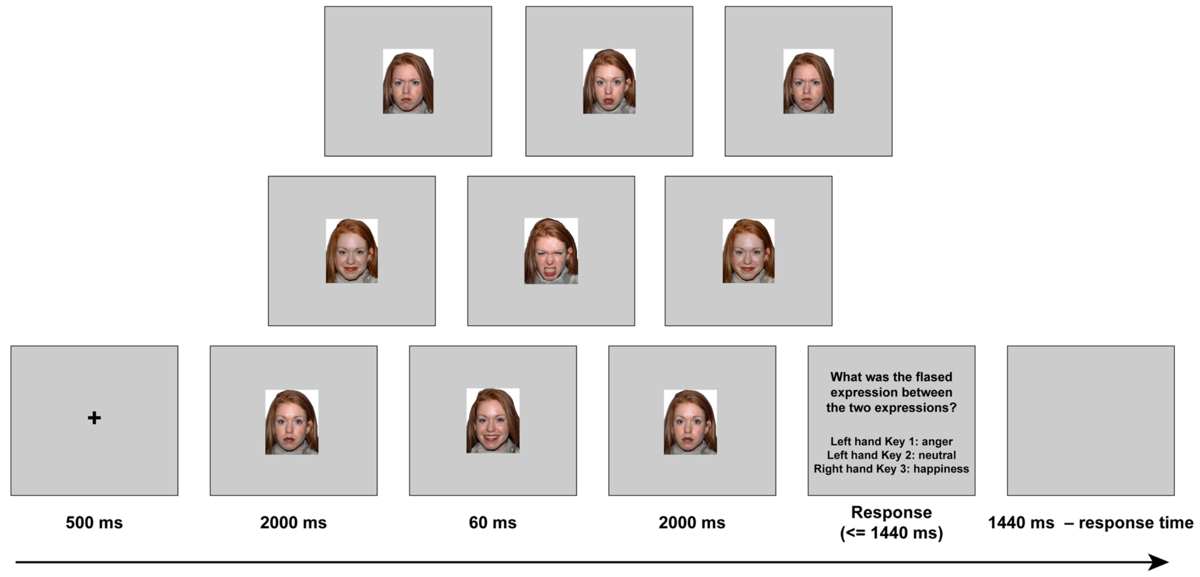


FIGURE 1 | The experimental setup of each trial.

three buttons (half of the participants were told to press 1 or 2 with the right hand, and 3 with the left hand, while the other half of the participants were told to press 1 or 2 with the left hand, and 3 with the right hand). If there was no reaction, the task instructions would disappear after 1,440 ms. Finally, a blank screen was presented for 1,440 ms minus reaction time to ensure that the total duration of each trial was the same.

Data Acquisition

A Siemens 3.0-T scanner (Siemens Magnetom Trio Tim, Erlangen, Germany), equipped with a 12-channel head matrix coil, was used for functional brain imaging in the present study. The participant's head was securely but comfortably stabilized with firm foam padding. Scan sessions began with shimming and transverse localization. Functional data were acquired using an echo-planar imaging (EPI) sequence using an axial slice orientation [33 slices, repetition time (TR)/echo time (TE) = 2,000/30 ms, slice thickness = 3 mm, field of view (FOV) = 200 mm, flip angle = 90°; matrix size, 64 × 64] covering the whole brain. A high-resolution T1-weighted 3D MRI sequence was acquired between the second and third sessions of fMRI (ascending slices, 128 slices, TR/TE = 2,530/2.5 ms, FOV = 256 mm × 256 mm, flip angle = 7°, voxel size = 1.0 mm × 1.0 mm × 1.3 mm).

Data Analysis

The data were preprocessed and analyzed using Statistical Parametric Mapping software SPM8 (Wellcome Department of Cognitive Neurology, London, United Kingdom). Standard fMRI preprocessing was performed including slice timing, realignment (data with translation of more than 3 mm or rotation angle greater than 2.5° were removed), spatial normalization [EPI template; Montreal Neurologic Institute (MNI)], reslicing (3 mm × 3 mm × 3 mm voxels), and smoothing with a 6-mm full-width at half-maximum (FWHM) Gaussian kernel. The conventional two-level approach using SPM8 was adopted for event-related fMRI data. The variance in blood-oxygen-level-dependent (BOLD) signal was decomposed in a general linear model separately for each run (Friston et al., 1994). The time course of activity of each voxel was modeled as a sustained response during each trial, convolved with a standard estimate of the hemodynamic impulse response function (Boynton et al., 1996). Low-frequency BOLD signal noise was removed by high-pass filtering of 128 s. For the whole-brain analysis, cluster-level familywise error (FWE) corrected at $p < 0.05$ and cluster size ≥ 13 voxels were applied. Considering the number of missed trials without a response was minor (56 trails out of the total of 6,048), we kept all the trials for the next data processing.

The whole-brain analysis was conducted to reveal the brain activation using context and target as explanatory variables. The initial comparisons of task-related events[1] time-locked to the front context onset (duration = 2.06 s) and baseline were performed by a single-sample *t*-test in the first-level analysis (Fusar-Poli et al., 2009; Sabatinelli et al., 2011). In the second-level analysis, using the context-by-target interaction term (e.g., negative context–anger target), we analyzed the brain activation related to task-related conditions. The emotional reactivity contrasts[2] were obtained by group analysis in second-level analysis using paired *t*-test ($p < 0.001$).

[1] The nine task-related events (negative context–anger target, neutral context–anger target, positive context–anger target, negative context–neutral target, neutral context–neutral target, positive context–neutral target, negative context–happiness target, neutral context–happiness target, and positive context–happiness target).

[2] Negative context–anger target > neutral context–anger target, positive context–anger target > neutral context–anger target, negative context–neutral target > neutral context–neutral target, positive context–neutral target > neutral

RESULTS

Behavioral Performance

The effect of emotional context on behavioral measures was assessed by applying a two-way repeated ANOVA to the recognition accuracies, with the context and the target as within-participant variables. It revealed a significant main effect of context, $F(2,40) = 33.76$, $p < 0.001$, $\eta_p^2 = 0.628$. The accuracies of targets following negative and positive conditions were significantly lower than that following neutral condition (see **Table 1**), $t(19) = -5.88$, $p < 0.001$, $d = 0.27$; $t(19) = -7.71$, $p < 0.001$, $d = 0.35$. The main effect of target microexpression was not significant, $F(2,40) = 0.54$, $p = 0.587$. The interaction of context and target reached significance, $F(4,80) = 4.58$, $p = 0.002$, $\eta_p^2 = 0.186$. Further analysis revealed that the accuracy rate for anger was significantly higher following neutral context than that following positive context, $t(19) = 3.35$, $p = 0.009$, $d = 0.69$; the accuracy rate for neutral was significantly higher following neutral context than that following negative or positive context, $t(19) = 4.90$, $p < 0.001$, $d = 1.29$; $t(19) = 4.87$, $p < 0.001$, $d = 1.27$; the accuracy rate for happiness was significantly higher following neutral context than that following negative context, $t(19) = 3.43$, $p = 0.008$, $d = 0.66$ (see **Figure 2A**).

We also analyzed the response time to examine the effect of emotional context on target. The two-way repeated ANOVA showed a significant main effect of context, $F(2,40) = 3.988$, $p = 0.026$, $\eta_p^2 = 0.166$. The response time of targets following negative condition (476.40 ± 213.35) were marginally significantly longer than that following neutral condition (458.11 ± 195.37), $t(19) = 2.45$, $p = 0.07$ (see **Figure 2B**). The main effects of target microexpression and the interaction were not significant, $F(2,40) = 2.36$, $p = 0.108$; $F(4,80) = 0.94$, $p = 0.446$.

fMRI Results

The whole-brain analysis based on paired t-test model for contrast conditions revealed that brain activations to target microexpressions varied across emotional contexts. Several areas exhibited significant increases in BOLD signals for contrasts of the same targets following different contexts (see **Table 2**). When the target was anger, there were increased BOLD signals mainly in the right intraparietal sulcus and extranuclear for the contrast of negative context against neutral context (negative context–anger target > neutral context–anger target, **Figure 3A**) whereas in the right precuneus and subgyral for the contrast of positive context against neutral context (positive context–anger target > neutral context–anger target, **Figure 3B**). When the target was neutral, there were increased BOLD signals mainly in the right inferior parietal lobule for the contrast of negative context against neutral context (negative context–neutral target > neutral context–neutral target, **Figure 3C**) whereas in the right precuneus and left dorsal posterior cingulate cortex for the contrast of positive context against neutral context (positive context–neutral target > neutral context–neutral target, **Figure 3D**; see **Supplementary Table S1** for more results).

TABLE 1 | Means and standard deviations of accuracies in all conditions.

	Context			Mean accuracies
	Negative	Neutral	Positive	
	M ± SD	*M* ± SD	*M* ± SD	
Target				
Anger	0.84 ± 0.13	0.85 ± 0.13	0.71 ± 0.24	0.80 ± 0.40
Neutral	0.73 ± 0.20	0.93 ± 0.09	0.71 ± 0.23	0.79 ± 0.41
Happiness	0.77 ± 0.17	0.87 ± 0.14	0.83 ± 0.15	0.83 ± 0.38
Mean accuracies	0.78 ± 0. 41	0.88 ± 0.32	0.75 ± 0.43	

DISCUSSION

We verified that emotional contexts influence microexpression recognition, which is consistent with previous findings (Zhang et al., 2014). Target microexpressions were recognized better following neutral contexts than those following positive or negative contexts. Emotional stimuli affect how we process and respond to targets (Siciliano et al., 2017). Compared to the neutral stimuli, the emotional ones can be highly salient, and these emotionally salient events can disrupt the recognition of targets (Siciliano et al., 2017). Attention allocation was reported to be related to and modulated by the emotional valences of stimuli—that is, emotional stimuli could capture more attention (Wilson and Hugenberg, 2013). Increasing attentional load decreases the processing resources available for the subsequent task (Kurth et al., 2016). In our study, it seemed that there was not enough attention directed to the subsequent target microexpressions because of emotional contexts, and poor performance for recognition was therefore observed. Our fMRI results also supported this: there were increased activities in some attention-related functional regions when microexpressions followed negative or positive contexts.

Emotional stimuli, either pleasant or unpleasant, prompted significantly more activity than did neutral pictures (Lang et al., 1998). Accordingly, we found that brain activities associated with the same target microexpression following various emotional contexts differed in functional regions. Anger microexpressions followed negative context (negative context–anger target) compared to neutral context (neutral context–anger target) in that they activated the right intraparietal sulcus and extranuclear, whereas they followed positive context (positive context–anger target) compared to neutral context (neutral context–anger target), activating the right precuneus. Furthermore, it was observed that different regions responded to the neutral-target-related emotional contrast, for instance, the right precuneus. As in previous studies, these regions played a role in facial expression recognition (Mourao-Miranda et al., 2003). In our study, positive context compared to neutral context with the same target activated more brain regions, including the right precuneus. The precuneus participated in

context–neutral target, negative context–happiness target > neutral context–happiness target, and positive context–happiness target > neutral context–happiness target.

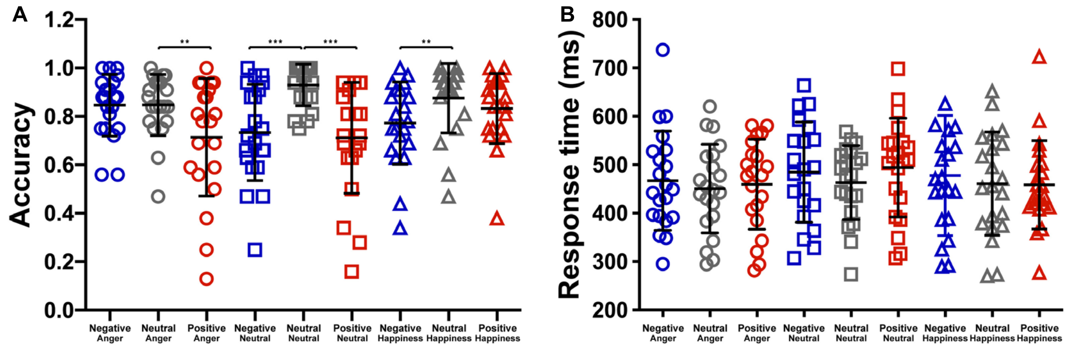


FIGURE 2 | The effect of context on the **(A)** accuracy and the **(B)** response time (***$p < 0.001$, **$p < 0.01$).

positive stimuli assessment (Paradiso et al., 1999), memory (Berthoz, 1997), and attention (Goldin and Gross, 2010). Unlike previous findings that negative stimuli (fear expressions) could also significantly activate the emotion-related areas (Sprengelmeyer et al., 1998), here, we only found that positive context activated the right precuneus, meaning that positive emotions could cause strong emotional arousal during this microexpression task. The left extranuclear was also reported to be significantly activated by happy faces compared with neutral faces (Trautmann et al., 2009) and was related to emotional regulation (McRae et al., 2008). Here, we found that the right extranuclear was activated in the negative context compared to the neutral context when they were followed by anger microexpressions, implying that the activities of facial expressions including context and target could be complicated. The brain responses in these contrasts reflected the discrepancy in the processing of emotional contexts, which could suggest interactions between the processing of emotional contexts and of microexpressions. These discrepancies in emotional contexts implied that different emotional contexts might differently influence the processing of subsequent target microexpressions.

Based on the findings on behavioral performance and brain activities, emotional contexts lead to a decrease in recognition accuracies and an increase in context-related activations in some emotional and attentional regions. The increased perceptual load of negative and positive contexts yields increased brain activations along with decreased behavioral performance, due to the additional monitoring and attention necessary for inhibition of emotional contexts (Siciliano et al., 2017). Thus, the recognition of microexpression would be affected by the emotional contexts, which has been proven on behavioral performance. These activities in attention-related regions indicated that attention being occupied by negative and positive contexts might be a source of the effect of emotional contexts on the processing of microexpressions.

Limitations

Considering that a microexpression is very fast and is always submerged in other microexpressions, we did not leave a long break between context and target in order to simulate the real situation in which the microexpression happened. This led to our being unable to extract the exact BOLD response to the target and instead having combined the forward context and target and examined the whole duration. Here, our findings only showed that there were discrepancies in brain response between contrasts of the same targets following different contexts and suggested a limited potential effect of emotional context on subsequent target microexpressions, but not a very exact effect on microexpressions. Taking these issues into account, future work could focus on exploring the processing of different functional areas' responses to microexpression with more ecological validity and suitable experimental design in order to explore the neural mechanism for the effect of emotional context on microexpression.

TABLE 2 | Coordinates in Montreal Neurologic Institute (MNI) space and associated *t* scores showing the BOLD differences for the contrast of emotional contexts followed by the same microexpressions.

Brain regions	BA	Cluster size	*t*	*Z*	MNI		
					x	*y*	*z*
Target anger: negative > neutral							
Intraparietal sulcus (R)	7	100	5.26	4.00	18	−84	36
Extranuclear (R)		55	5.79	4.25	27	−39	15
Target anger: positive > neutral							
Precuneus (R)		144	4.78	3.76	24	−75	39
Subgyral (R)		49	5.95	4.31	33	−78	−6
Target neutral: negative > neutral							
Inferior parietal lobule (R)		111	6.00	4.34	42	−42	54
Target neutral: positive > neutral							
Precuneus (R)		1,108	8.31	5.19	12	−72	48
Dorsal posterior cingulate cortex (L)	31	772	8.47	5.23	−18	−84	36
Cerebellum_10 (L)		281	5.57	4.15	−24	−30	−42
Inferior semilunar lobule (L)		151	5.45	4.09	−30	−78	−45
Declive (R)		105	5.17	3.95	39	−60	−21

x, y, and z are coordinates in the MNI space. All p-values (p < 0.05) passed familywise error (FWE) correction at cluster level.

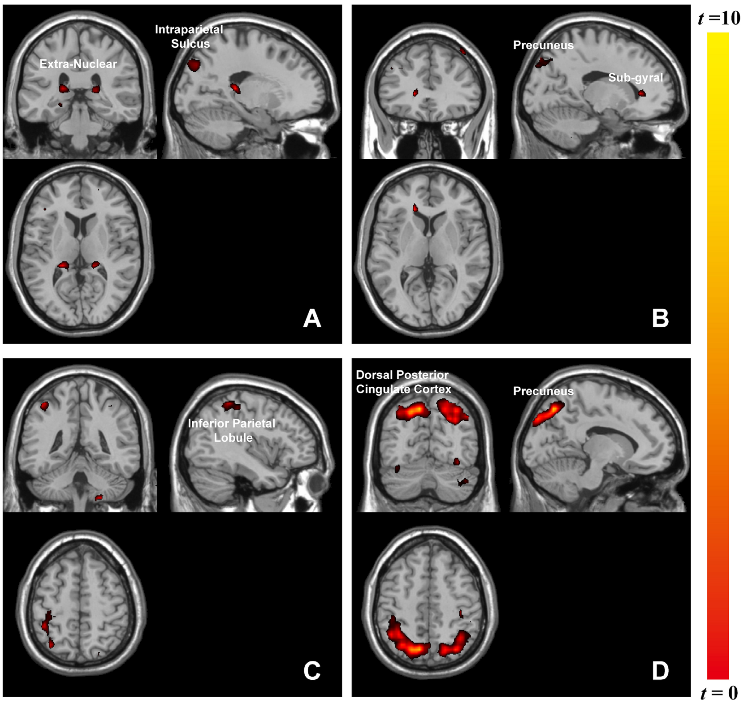


FIGURE 3 | Brain activation in contrasts of microexpressions following emotional contexts. When target was anger: **(A)** negative context > neutral context, extranuclear ($x = 27$, $y = -39$, $z = 15$), intraparietal sulcus ($x = 18$, $y = -84$, $z = 36$), **(B)** positive context > neutral context, precuneus ($x = 24$, $y = -75$, $z = 39$), subgyral ($x = 33$, $y = -78$, $z = -6$); when target was neutral: **(C)** negative context > neutral context, inferior parietal lobule ($x = 42$, $y = -42$, $z = 54$), **(D)** positive context > neutral context, precuneus ($x = 12$, $y = -72$, $z = 48$), dorsal posterior cingulate cortex ($x = -18$, $y = -84$, $z = 36$).

CONCLUSION

Compared with previous studies on emotional processing, our study made a bold attempt to explore the context effect on microexpression using the unconventional fMRI paradigm. The study showed that there were discrepancies between contrasts of the same targets following different contexts and suggested interactions between the processing of emotional contexts and of microexpressions. That is, brain responses in these contrasts reflected discrepancy in the processing of emotional contexts, meaning that different emotional contexts might differently interfere with the processing of subsequent target microexpressions.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of the Institute of Psychology at the Chinese Academy of Sciences. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MZ, KZ, and XF contributed to designing the experiment, analyzing the data, and writing the manuscript. FQ contributed to analyzing the data and writing the manuscript. MZ and KL contributed in collecting the data.

FUNDING

This research was supported by the National Natural Science Foundation of China, the grant of the German Research Foundation (DFG) in Project Crossmodel Learning (NSFC61621136008/DFG TRR-169), and the Scientific Foundation of Institute of Psychology, Chinese Academy of Sciences (Y9CX432005). Funds received open access publication fees from our grants.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2020.00329/full#supplementary-material

REFERENCES

Berthoz, A. (1997). Parietal and hippocampal contribution to topokinetic and topographic memory. *Biol. Sci.* 352, 1437–1448. doi: 10.1098/rstb.1997.0130

Boynton, G. M., Engel, S. A., Glover, G. H., and Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.* 16, 4207–4221. doi: 10.1523/JNEUROSCI.16-13-04207.1996

Ekman, P. (2002). *MicroExpression Training Tool (METT)*. San Francisco: University of California.

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

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., and Frackowiak, R. S. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210. doi: 10.1002/hbm.460020402

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., et al. (2009). Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *J. Psychiatry Neurosci.* 34, 418–432.

Goldin, P., and Gross, J. (2010). Effect of mindfulness meditation training on the neural bases of emotion regulation in social anxiety disorder. *Emotion* 10, 83–84.

Haxby, J. V., Hoffman, E., 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

Heberlein, A. S., Padon, A. A., Gillihan, S. J., Farah, M. J., and Fellows, L. K. (2008). Ventromedial frontal lobe plays a critical role in facial emotion recognition. *J. Cogn. Neurosci.* 20, 721–733. doi: 10.1162/jocn.2008.20049

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. Psychiatry* 92, 114–124. doi: 10.1016/j.biopsycho.2012.10.005

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

Kurth, S., Majerus, S., Bastin, C., Collette, F., Jaspar, M., Bahri, M. A., et al. (2016). Effects of aging on task- and stimulus-related cerebral attention networks. *Neurobiol. Aging* 44, 85–95. doi: 10.1016/j.neurobiolaging.2016.04.015

Lang, P. J., Bradley, M. M., Fitzsimmons, J. R., Cuthbert, B. N., Scott, J. D., Moulder, B., et al. (1998). Emotional arousal and activation of the visual cortex: an fMRI analysis. *Psychophysiology* 35, 199–210. doi: 10.1111/1469-8986.3520199

Lee, K. H., and Siegle, G. J. (2014). Different brain activity in response to emotional faces alone and augmented by contextual information. *Psychophysiology* 51, 1147–1157. doi: 10.1111/psyp.12254

McRae, K., Ochsner, K. N., Mauss, I. B., Gabrieli, J. J., and Gross, J. J. (2008). Gender differences in emotion regulation: An fMRI study of cognitive reappraisal. *Group Processes Intergroup Relat.* 11, 143–162. doi: 10.1177/1368430207088035

Mourao-Miranda, J., Volchan, E., Moll, J., de Oliveira-Souza, R., Oliveira, L., Bramati, I., et al. (2003). Contributions of stimulus valence and arousal to visual activation during emotional perception. *Neuroimage* 20, 1955–1963. doi: 10.1016/j.neuroimage.2003.08.011

Paradiso, S., Johnson, D. L., Andreasen, N. C., O'Leary, D. S., Watkins, G. L., Boles Ponto, L. L., et al. (1999). Cerebral blood flow changes associated with attribution of emotional valence to pleasant, unpleasant, and neutral visual stimuli in a PET study of normal subjects. *Am. J. Psychiatr.* 156, 1618–1629. doi: 10.1176/ajp.156.10.1618

Sabatinelli, D., Fortune, E. E., Li, Q., Siddiqui, A., Krafft, C., Oliver, W. T., et al. (2011). Emotional perception: meta-analyses of face and natural scene processing. *Neuroimage* 54, 2524–2533. doi: 10.1016/j.neuroimage.2010.10.011

Schulz, K. P., Clerkin, S. M., Halperin, J. M., Newcorn, J. H., Tang, C. Y., and Fan, J. (2009). Dissociable neural effects of stimulus valence and preceding context during the inhibition of responses to emotional faces. *Hum. Brain Mapp.* 30, 2821–2833. doi: 10.1002/hbm.20706

Schwarz, K. A., Wieser, M. J., Gerdes, A. B. M., Mühlberger, A., and Pauli, P. (2013). Why are you looking like that? How the context influences evaluation and processing of human faces. *Soc. Cogn. Affect. Neurosci.* 8, 438–445. doi: 10.1093/scan/nss013

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

Siciliano, R. E., Madden, D. J., Tallman, C. W., Boylan, M. A., Kirste, I., Monge, Z. A., et al. (2017). Task difficulty modulates brain activation in the emotional oddball task. *Brain Res.* 1664, 74–86. doi: 10.1016/j.brainres.2017.03.028

Sprengelmeyer, R., Rausch, M., Eysel, U. T., and Przuntek, H. (1998). Neural structures associated with recognition of facial expressions of basic emotions. *Biol. Sci.* 265, 1927–1931. doi: 10.1098/rspb.1998.0522

Trautmann, S. A., Fehr, T., and Herrmann, M. (2009). Emotions in motion: dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain Res.* 1284, 100–115. doi: 10.1016/j.brainres.2009.05.075

Van den Stock, J., Vandenbulcke, M., Sinke, C. B., Goebel, R., and de Gelder, B. (2014). How affective information from faces and scenes interacts in the brain. *Soc. Cogn. Affect. Neurosci.* 9, 1481–1488. doi: 10.1093/scan/nst138

Werheid, K., Alpay, G., Jentzsch, I., and Sommer, W. (2005). Priming emotional facial expressions as evidenced by event-related brain potentials. *Int. J. Psychophysiol.* 55, 209–219. doi: 10.1016/j.ijpsycho.2004.07.006

Wieser, M. J., and Brosch, T. (2012). Faces in context: a review and systematization of contextual influences on affective face processing. *Front. Psychol.* 3:471. doi: 10.3389/fpsyg.2012.00471

Wilson, J. P., and Hugenberg, K. (2013). Shared signal effects occur more strongly for salient outgroups than ingroups. *Soc. Cogn.* 31, 636–648. doi: 10.1521/soco.2013.31.6.636

Zhang, M., Fu, Q., Chen, Y. H., and Fu, X. (2018). Emotional context modulates micro-expression processing as reflected in event-related potentials. *PsyCh J.* 7, 13–24. doi: 10.1002/pchj.196

Zhang, M., Fu, Q. F., Chen, Y. H., and Fu, X. L. (2014). Emotional context influences micro-expression recognition. *PLoS One* 9:e95018. doi: 10.1371/journal.pone.0095018

Zhao, K., Zhao, J., Zhang, M., Cui, Q., and Fu, X. L. (2017). Neural responses to rapid facial expressions of fear and surprise. *Front. Psychol.* 8:761. doi: 10.3389/fpsyg.2017.00761

Conflict of Interest: 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

MINI REVIEW
published: 03 July 2020
doi: 10.3389/fpsyg.2020.01378

Recognizing Genuine From Posed Facial Expressions: Exploring the Role of Dynamic Information and Face Familiarity

Karen Lander[1]* *and Natalie L. Butcher*[2]

[1] Division of Neuroscience and Experimental Psychology, University of Manchester, Manchester, United Kingdom, [2] School of Social Sciences, Humanities and Law, Teesside University, Middlesbrough, United Kingdom

The accurate recognition of emotion is important for interpersonal interaction and when navigating our social world. However, not all facial displays reflect the emotional experience currently being felt by the expresser. Indeed, faces express both genuine and posed displays of emotion. In this article, we summarize the importance of motion for the recognition of face identity before critically outlining the role of dynamic information in determining facial expressions and distinguishing between genuine and posed expressions of emotion. We propose that both dynamic information and face familiarity may modulate our ability to determine whether an expression is genuine or not. Finally, we consider the shared role for dynamic information across different face recognition tasks and the wider impact of face familiarity on determining genuine from posed expressions during real-world interactions.

Keywords: expression recognition, genuine and posed, dynamic information, face familiarity, face recognition

OPEN ACCESS

Edited by:
Shuo Wang,
West Virginia University, United States

Reviewed by:
Pedro Guerra,
University of Granada, Spain
Harold Hill,
University of Wollongong, Australia

****Correspondence:***
Karen Lander
karen.lander@manchester.ac.uk

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

Received: *05 February 2020*
Accepted: *22 May 2020*
Published: *03 July 2020*

Citation:
Lander K and Butcher NL (2020)
Recognizing Genuine From Posed
Facial Expressions: Exploring the Role
of Dynamic Information and Face
Familiarity. Front. Psychol. 11:1378.
doi: 10.3389/fpsyg.2020.01378

INTRODUCTION

Face perception is a crucial part of social cognition, and on a daily basis, we encounter many faces. Faces convey characteristics of the viewed person, like their age, gender, emotional state, and identity. Face identity recognition is particularly important for social functioning as it enables us to identify a familiar person from an unknown individual. Previous research has revealed that factors including facial attractiveness, distinctiveness (Wiese et al., 2014), race (Meissner and Brigham, 2001), and facial motion (Lander et al., 1999) influence how well a face is recognized. Similarly, the ability to accurately determine another person's emotional state is important for navigating day-to-day social interactions, for example, realizing whether a person is friendly or frightened, angry or sad. Previous research has shown that we use voice prosody (e.g., Wurm et al., 2001), body position (de Gelder, 2006), gait (Montepare et al., 1987), and facial expression (Adolphs, 1999) to determine emotional state.

Displayed facial expressions may reflect a genuinely felt emotion linked to an actual, remembered, or imagined event, for example, fear when scared or sad when remembering the death of a loved one. However, in some circumstances, facial expression may not reflect genuine emotion but instead be posed. Here, there may be no strong emotional experience, like smiling on cue or faking a surprised look. Alternatively, the expression displayed may mask the genuine emotion felt, like smiling when receiving a disappointing present. "Display rules" are rules learnt early in life that help determine the appropriate expression of emotion in different social contexts (Ekman and Friesen, 1969) and cultures (Matsumoto et al., 2009). Emotions may be amplified or de-amplified; they may be masked, neutralized, or simulated. Masking of emotions may be one way to recruit the help of others or otherwise gain a social advantage (Krumhuber and Manstead, 2009).

Research on facial expression processing has predominantly used static facial images taken at the expression "apex." For example, Ekman and Friesen (1976) created a set of standardized static images of the "basic" facial expressions of happiness, sadness, fear, anger, disgust, surprise, and neutral. However, in the real world, facial expressions are dynamic in nature, rapidly changing over time. Interestingly, it is known that we are highly sensitive to dynamic information available from the face (Edwards, 1998; Dobs et al., 2014). Accordingly, sets of dynamic expressions have been developed (Amsterdam Dynamic Facial Expressions Set; ADFES; van der Schalk et al., 2011). It is important to consider the way in which expression sets are created. Typically, they are created by telling or showing the "actors" how to display prototypical expressions [based on facial action coding scheme (FACS) coding; Ekman and Friesen, 1978]. However, some research aims to capture genuine facial expressions that spontaneously occur as part of an emotional experience (see McLellan et al., 2010). Work on expression genuineness necessarily utilizes this method, with "genuine expressions" usually filmed in the lab. We return to consider the real-world application of such work, later in this article.

In this review, our overall aim is to explore the role of dynamic information in determining genuine from posed expressions. We start by outlining work investigating the recognition of face identity, highlighting the potential role for "characteristic motion signatures" (O'Toole et al., 2002). Next, we consider the role of dynamic information when recognizing facial expressions. Characteristic motion signatures may also be associated with emotional expressions and thus play a role in determining expression genuineness. Accordingly, we critically consider the difference between genuine and posed emotional expressions, in terms of the static- and dynamic-based cues available. Lastly, we consider the possible mediating effect of dynamic information and face familiarity when discriminating between genuine and posed expressions.

MOVEMENT AND THE RECOGNITION OF FACE IDENTITY

Research has established that dynamic information is important when determining face identity ("motion advantage"; see Schiff et al., 1986; Knight and Johnston, 1997; Lander et al., 1999). Specifically, research has found that seeing a face move aids the learning of face identity (Pike et al., 1997; Knappmeyer et al., 2003; Lander and Bruce, 2003; Pilz et al., 2006; Lander and Davies, 2007; Butcher et al., 2011), identification of familiar faces (Knight and Johnston, 1997; Lander et al., 2001), and accurate and faster face matching (Thornton and Kourtzi, 2002). Dynamic facial information seems to be a particularly useful cue to identity recognition when viewing conditions are difficult, for example, when faces are presented in photographic negative (see Knight and Johnston, 1997; in a negative image, the pattern of brightness is reversed) or blurred (Lander et al., 2001). Also, dynamic information is useful when there is perceiver impairment, such as prosopagnosia (see Steede et al., 2007; Longmore and Tree, 2013; Xiao et al., 2014; Bennetts et al., 2015).

O'Toole et al. (2002) proposed several theoretical reasons why seeing a face move may facilitate identity recognition. These theories are not mutually exclusive and the extent to which they each account for the motion advantage may depend on whether the to-be-recognized face is unfamiliar or known. For unfamiliar faces, seeing a face move may help build robust face representations via structure-from-motion processes ("representation enhancement hypothesis"). However, for familiar faces, people may learn characteristic motion patterns associated with their identity, which act as an additional cue to identity ("supplemental information hypothesis"). Finally, social cues available from the moving face may attract attention to the identity-specific areas of the face, facilitating identity processing ("social signals hypothesis"). While both the representation enhancement and supplemental information hypotheses have received empirical support (e.g., Knappmeyer et al., 2003; Butcher et al., 2011), the plausibility of the social signals hypothesis is relatively unknown, as its predictions have received little attention. To summarize, dynamic information available from a moving face may be useful for both building new face representations and accessing established ones.

MOVEMENT AND THE RECOGNITION OF FACIAL EXPRESSIONS

While the motion advantage in identity recognition appears relatively robust, the effect of dynamic information on facial expression recognition is less consistent. Some research has shown that dynamic facial expressions are recognized more accurately (Cunningham and Wallraven, 2009; Trautmann et al., 2009) and rapidly (Calvo et al., 2016) than static facial expressions (see Krumhuber et al., 2013). However, other studies have found no difference between static and dynamic expression recognition (Kätsyri et al., 2008; Fiorentini and Viviani, 2011) or have only found a dynamic recognition advantage for some expressions (Fujimura and Suzuki, 2010; Recio et al., 2011).

One potential issue when comparing dynamic and static facial expression recognition is that static performance typically approaches ceiling, leaving little "room" to demonstrate any advantage. Interestingly, the usefulness of dynamic information for expression recognition is seen in studies that make recognition more difficult, through the use of point-light stimuli (Matsuzaki and Sato, 2008), subtle expressions (Ambadar et al., 2005), or by imposing time pressures (Zhongqing et al., 2014). Furthermore, Kamachi et al. (2001) found that changing the dynamic parameters of morphed expressions affected how well different expressions were recognized. As with identity recognition, dynamic facial information may support expression recognition in a flexible way, optimizing face perception when the task demands of everyday face-to-face interactions are such that static cues alone are not sufficient (Xiao et al., 2014).

In additional work supporting the distinction between recognition of moving and static expressions, Humphreys et al. (1993) report the case of an acquired prosopagnosic patient who could make expression judgments from moving (but not static) faces, consistent with the idea of at least partially dissociable static and dynamic expression processing. A number of neuroimaging studies have also investigated neural differences when viewing dynamic and static facial expressions (Kilts et al., 2003; Sato et al., 2004; Trautmann et al., 2009; Foley et al., 2012). Trautmann et al. (2009) found that dynamic faces enhanced emotion-specific brain activation patterns in the parahippocampal gyrus, including the amygdala, fusiform gyrus, superior temporal gyrus, inferior frontal gyrus, and occipital and orbitofrontal cortex. *Post hoc* ratings of the dynamic stimuli revealed better recognizability in comparison to the static stimuli (but see Trautmann-Lengsfeld et al., 2013). To summarize, much behavioral and neural work suggests that dynamic information can be useful in face expression recognition, particularly when recognition is difficult. However, this advantage is not unequivocally shown in the existing literature.

MOVEMENT AND THE RECOGNITION OF GENUINE FROM POSED EXPRESSIONS

Increasingly, researchers have become interested in the distinction between genuine and posed facial expressions. Initially, research concentrated on static happy expressions (see Frank et al., 1993; Gunnery and Ruben, 2016). Here, genuine smiles ("Duchenne" smiles) are thought to involve crinkling around the eyes ("Crows feet") caused by activation of the orbicularis oculi muscles. Posed smiles instead involve just an upturned mouth, created by contraction of the zygomatic major muscle. More recent work has investigated expression genuineness discrimination across a range of emotions.

Accordingly, McLellan et al. (2010) found that perceivers were able to distinguish between static genuine and posed happy, sad, and fear facial expressions. They also found that participants made valence judgments to words faster after viewing a genuine valence-congruent expression (i.e., smile before a positive word) compared to a posed expression. Additional support for differences between the perception of genuine and posed expressions comes from neuroimaging work which showed different patterns of neural activation (McLellan et al., 2012). However, findings by Dawel et al. (2015) suggest that the differences between genuine and posed expressions are less apparent than previously proposed. They found that both adults and children could discriminate genuine from posed happy expressions and adults were able to discriminate sad displays. However, neither group could discriminate between genuine and posed scared facial expressions. We conclude that most research, using static pictures, suggests that people can successfully discriminate between genuine and posed facial expressions in some circumstances – but that this ability may vary by expression and individual.

It is also important to consider the role of dynamic information in determining expression genuineness. Dynamic aspects of an expression may serve as useful cues when distinguishing genuine from posed expressions (Hess and Kleck, 1994; Gunnery and Ruben, 2016). Early research proposed that genuine smiles last between 500 and 4000 ms with posed smiles being either shorter or longer than this (Ekman, 2009). In addition, genuine smiles may have a slower onset speed and longer onset duration (Schmidt et al., 2006) than posed smiles. Recent research has begun to investigate the role of dynamic information in the recognition of expression genuineness across a range of facial expressions.

Interestingly, Namba et al. (2018) asked participants to judge whether viewed facial expressions were being depicted (posed) or experienced (genuine). Expressions (amusement, surprise, disgust, and fear) were shown as dynamic or static clips. For all expressions, genuine expressions were judged more as being experienced than posed. Importantly, participants were better at differentiating between genuine and posed expressions when dynamic than static. Similarly, Zloteanu et al. (2018) found that the use of moving stimuli improved the discrimination of surprise authenticity. We note that as with static images, overall performance on dynamic expression genuineness decisions may depend on the exact task used, what emotions are considered, the participants themselves, and so on. However, cues to expression authenticity may be present in the dynamics of the facial movement.

INTERDEPENDENCE BETWEEN FACE FAMILIARITY AND FACE MOVEMENT IN THE RECOGNITION OF EXPRESSION GENUINENESS

We have already outlined research that suggests dynamic facial information is useful when determining the genuineness of facial expressions of emotion. Here, we further propose that there may be interdependence between face familiarity and face movement when determining expression genuineness.

In terms of face familiarity, it is known from neuroimaging studies that personal familiarity impacts on the response of neural systems involved in expression processing (Gobbini et al., 2004; Leibenluft et al., 2004). There is also some evidence that familiarity plays a role in the recognition of genuine emotional expressions, with performance seen to improve with familiarity (Wild-Wall et al., 2008; Huynh et al., 2010). However, other studies indicate a detrimental effect of familiarity on expression recognition in children (Herba et al., 2008) and some clinical populations (e.g., schizophrenia; Lahera et al., 2013). Thus, there is inconsistency regarding the role of familiarity on expression recognition.

Interestingly, research investigating the recognition of expression genuineness typically uses unfamiliar faces. This may be reflective of some real-life tasks, for example, in a criminal situation where the task is to determine whether an unfamiliar suspect is displaying a genuine expression or covering up a lie (Porter and ten Brinke, 2010). However, often, our interpretation of expression genuineness involves familiar people – for example, is our child genuinely happy or sarcastically smiling? Further research is needed to determine how face familiarity influences our ability to determine expression genuineness. We propose that for familiar faces, there may be additional cues that help us determine whether an expression is genuine or not, for example, a particular lop-sided smile associated with the genuine smile of a friend. Such idiosyncratic static-based cues may aid the distinction between genuine and posed smiles for this person. Thus, it is possible that face familiarity plays a mediating role in the recognition of genuine versus posed expressions, with better discrimination for familiar compared with unfamiliar faces.

It is also important to consider the possible interdependence between familiarity and dynamic information. When a face is familiar, characteristic motion patterns may act as an additional cue to identity. Indeed, the size of the motion advantage for face recognition is positively associated with face familiarity (Butcher and Lander, 2016). Such characteristic motion patterns may be linked to expressional movements. Thus, face familiarity may play a more prominent role when recognizing genuine from posed expressions using dynamic stimuli. For example, a friend may have a characteristic smile (present in the static image) but they may also have a characteristic way of smiling (dynamic characteristics). Here, cues to expression genuineness may be present in both the static- and dynamic-based parameters of a familiar person's expression. To summarize, further work is needed to determine whether expression genuineness decisions are better for familiar than unfamiliar faces and whether this advantage is exaggerated for dynamic compared with static clips. In addition, we need to consider the interdependence between face familiarity, dynamic information, and expression genuineness.

CONCLUDING COMMENTS AND FUTURE DIRECTIONS

The literature reviewed demonstrates that dynamic information is useful for face identification (Lander et al., 1999), expression recognition (Krumhuber et al., 2013), and for expression genuineness judgments (Namba et al., 2018). Further, we propose a possible facilitative effect of face familiarity and face movement when determining expression genuineness. It is interesting to consider what other issues remain in this research area.

First, we propose a shared role for dynamic information across different face tasks. Much facial motion contains both identity-specific and expression information which, on an everyday basis, are processed simultaneously. Work is needed to determine whether neural models of face processing can account for the shared importance of dynamic information across different face processing tasks. According to Haxby's neural account (Haxby et al., 2000; Haxby and Gobbini, 2011), there is one cortical pathway that processes invariant aspects of faces (identity and gender; Fusiform Face Area) and another that processes changeable aspects of faces (expression and eye gaze; posterior superior temporal sulcus face area; pSTS-FA). Pitcher et al. (2014) suggest that the dynamic motor and static components of a face are processed via dissociable cortical pathways. Alternatively, Bernstein et al. (2018) suggest an integrated neural model of face processing, with dorsal face areas (pSTS-FA) sensitive to dynamic and changeable facial aspects whereas ventral areas (Occipital Face Area and Fusiform Face Area) extract form information from both invariant and changeable facial aspects. Such neural accounts need to be integrated with behavioral work to better understand the shared role of dynamic information for the different face tasks we encounter in the real world.

Second, to fully understand the task of recognizing expression genuineness, it is necessary to know what information is required for this task. Low and high spatial frequencies play different roles in the perception of facial expressions (Vuilleumier et al., 2003). Low spatial frequencies carry global/configural information whereas high spatial frequencies convey localized/fine-grain information. Low and high spatial frequencies may also play different roles in the classification of expression genuineness (Laeng et al., 2010; Kihara and Takeda, 2019). Additional work is needed to isolate which spatial frequency aspects of faces are diagnostic of expression genuineness when shown as dynamic clips.

Finally, it is important to consider the collection and use of expressions used in recognition experiments. Genuine expressions using emotion elicitation methods in the lab may lack the spontaneity of genuine expressions in the real world (Smoski and Bachorowski, 2003). The selection of genuine expressions by the experimenter may also rely on the criteria used in posed expressions. We suggest that real world expressions may be more idiosyncratic and individualist than those collected in the lab, modulated by familiarity and context. Investigation of these issues is important so that we can further consider expression genuineness and the impact of familiarity and dynamic information.

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 funded from the University of Manchester Open Access fund.

REFERENCES

Adolphs, R. (1999). Social cognition and the human brain. *Trends Cogn. Sci.* 3, 469–479.

Ambadar, Z., Schooler, J. W., and Cohn, J. F. (2005). Deciphering the enigmatic face—the importance of facial dynamics in interpreting subtle facial expressions. *Psychol. Sci.* 16, 403–410. doi: 10.1111/j.0956-7976.2005.01548.x

Bennetts, R. J., Butcher, N., Lander, K., and Bate, S. (2015). Movement cues aid face recognition in developmental prosopagnosia. *Neuropsychology* 29, 855–860. doi: 10.1037/neu0000187

Bernstein, M., Erez, Y., Blank, I., and Yovel, G. (2018). An integrated neural framework for dynamic and static face processing. *Sci. Rep.* 8:7036.

Butcher, N., and Lander, K. (2016). Exploring the motion advantage: evaluating the contribution of familiarity and differences in facial motion. *Q. J. Exp. Psychol.* 70, 919–929. doi: 10.1080/17470218.2016.1138974

Butcher, N., Lander, K., Fang, H., and Costen, N. (2011). The effect of motion at encoding and retrieval for same and other race face recognition. *Br. J. Psychol.* 102, 931–942. doi: 10.1111/j.2044-8295.2011.02060.x

Calvo, M. G., Avero, P., Fernández-Martín, A., and Recio, G. (2016). Recognition thresholds for static and dynamic emotional faces. *Emotion* 16, 1186–1200. doi: 10.1037/emo0000192

Cunningham, D. W., and Wallraven, C. (2009). Dynamic information for the recognition of conversational expressions. *J. Vis.* 9, 7.1–7.17.

Dawel, A., Palermo, R., O'Kearney, R., and McKone, E. (2015). Children can discriminate the authenticity of happy but not sad or fearful facial expressions, and use an immature intensity-only strategy. *Front. Psychol.* 6:462. doi: 10.3389/fpsyg.2015.00462

de Gelder, B. (2006). Toward a biological theory of emotional body language. *Biol. Theory* 1, 130–132. doi: 10.1162/biot.2006.1.2.130

Dobs, K., Bulthoff, I., Breidt, M., Vuong, Q. C., Curio, C., and Schultz, J. (2014). Quantifying human sensitivity to spatio-temporal information in dynamic faces. *Vis. Res.* 100, 78–87. doi: 10.1016/j.visres.2014.04.009

Edwards, K. (1998). The face of time: temporal cues in facial expression of emotion. *Psychol. Sci.* 9, 270–276. doi: 10.1111/1467-9280.00054

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

Ekman, P., and Friesen, W. V. (1969). The repertoire of nonverbal behavior: categories, origins, usage, and coding. *Semicotica* 1, 49–98.

Ekman, P., and Friesen, W. V. (1976). *Pictures of Facial Affect*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., and Friesen, W. V. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto, CA: Consulting Psychologists Press.

Fiorentini, C., and Viviani, P. (2011). Is there a dynamic advantage for facial expressions? *J. Vis.* 11:17. doi: 10.1167/11.3.17

Foley, E., Rippon, G., Thai, N. J., Longe, O., and Senior, C. (2012). Dynamic facial expressions evoke distinct activation in the face perception network: a connectivity analysis study. *J. Cogn. Neurosci.* 24, 507–520. doi: 10.1162/jocn_a_00120

Frank, M. G., Ekman, P., and Friesen, W. V. (1993). Behavioral markers and recognizability of the smile of enjoyment. *J. Pers. Soc. Psychol.* 64, 83–93. doi: 10.1037/0022-3514.64.1.83

Fujimura, T., and Suzuki, N. (2010). Recognition of dynamic facial expressions in peripheral and central vision. *Jpn. J. Psychol.* 81, 348–355. doi: 10.4992/jjpsy.81.348

Gobbini, M. I., Leibenluft, E., Santiago, N., and Haxby, J. V. (2004). Social and emotional attachment in the neural representation of faces. *Neuroimage* 22, 1628–1635. doi: 10.1016/j.neuroimage.2004.03.049

Gunnery, S. D., and Ruben, M. A. (2016). Perceptions of Duchenne and non-Duchenne smiles: a meta-analysis. *Cogn. Emot.* 30, 501–515. doi: 10.1080/02699931.2015.1018817

Haxby, J. V., and Gobbini, M. I. (2011). Distributed neural systems for face perception. *Oxford Handb. Face Percept.* 6, 93–110.

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

Herba, C. M., Benson, P., Landau, S., Russell, T., Goodwin, C., Lemche, E., et al. (2008). Impact of familiarity upon children's developing facial expression recognition. *J. Child Psychol. Psychiatry* 49, 201–210. doi: 10.1111/j.1469-7610.2007.01835.x

Hess, U., and Kleck, R. E. (1994). The cues decoders use in attempting to differentiate emotion-elicited and posed facial expressions. *Eur. J. Soc. Psychol.* 24, 367–381. doi: 10.1002/ejsp.2420240306

Humphreys, G. W., Donnelly, N., and Riddoch, M. J. (1993). Expression is computed separately from facial identity, and it is computed separately for moving and static faces: neuropsychological evidence. *Neuropsychologia* 31, 173–181. doi: 10.1016/0028-3932(93)90045-2

Huynh, C. M., Vicente, G. I., and Peissig, J. J. (2010). The effects of familiarity on genuine emotion recognition. *J. Vis.* 10:628. doi: 10.1167/10.7.628

Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., and Akamatsu, S. (2001). Dynamic properties influence the perception of facial expressions. *Perception* 30, 875–887. doi: 10.1068/p3131

Kätsyri, J., Saalasti, S., Tiippana, K., von Wendt, L., and Sams, M. (2008). Impaired recognition of facial emotions from low-spatial frequencies in Asperger syndrome. *Neuropsychologia* 46, 1888–1897. doi: 10.1016/j.neuropsychologia.2008.01.005

Kihara, K., and Takeda, Y. (2019). The role of low-spatial frequency components in the processing of deceptive faces: a study using artificial face models. *Front. Psychol.* 10:1468. doi: 10.3389/fpsyg.2019.01468

Kilts, C. D., Egan, G., Gideon, D. A., Ely, T. D., and Hoffman, J. M. (2003). Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions. *NeuroImage* 18, 156–168. doi: 10.1006/nimg.2002.1323

Knappmeyer, B., Thornton, I., and Bülthoff, H. (2003). The use of facial motion and facial form during the processing of identity. *Vis. Res.* 43, 1921–1936. doi: 10.1016/s0042-6989(03)00236-0

Knight, B., and Johnston, A. (1997). The role of movement in face recognition. *Vis. Cogn.* 4, 265–273. doi: 10.1080/713756764

Krumhuber, E. G., Kappas, A., and Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: a review. *Emot. Rev.* 5, 41–46. doi: 10.1177/1754073912451349

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

Laeng, B., Profeti, I., Saether, L., Adolfsdottir, S., Lundervold, A. J., Vangberg, T., et al. (2010). Invisible expressions evoke core impressions. *Emotion* 10, 573–586. doi: 10.1037/a0018689

Lahera, G., Herrera, S., Fernández, C., Bardón, M., de los Ángeles, V., and Fernández-Liria, A. (2013). Familiarity and face emotion recognition in patients with schizophrenia. *Comp. Psychiatry* 55, 199–205. doi: 10.1016/j.comppsych.2013.06.006

Lander, K., and Bruce, V. (2003). The role of motion in learning new faces. *Vis. Cogn.* 10, 897–912. doi: 10.1080/13506280344000149

Lander, K., Bruce, V., and Hill, H. (2001). Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. *Appl. Cogn. Psychol.* 15, 101–116. doi: 10.1002/1099-0720(200101/02)15:1<101::aid-acp697>3.0.co;2-7

Lander, K., Christie, F., and Bruce, V. (1999). The role of movement in the recognition of famous faces. *Mem. Cogn.* 27, 974–985. doi: 10.3758/bf03201228

Lander, K., and Davies, R. (2007). Exploring the role of characteristic motion when learning new faces. *Q. J. Exp. Psychol.* 60, 519–526. doi: 10.1080/17470210601117559

Leibenluft, E., Gobbini, M. I., Harrison, T., and Haxby, J. V. (2004). Mothers' neural activation in response to pictures of their, and other, children. *Biol. Psychiatry* 56, 225–232. doi: 10.1016/j.biopsych.2004.05.017

Longmore, C., and Tree, J. (2013). Motion as a cue to face recognition: evidence from congenital prosopagnosia. *Neuropsychologia* 51, 864–875. doi: 10.1016/j.neuropsychologia.2013.01.022

Matsumoto, D., Willingham, B., and Olide, A. (2009). Sequential dynamics of culturally moderated facial expressions of emotion. *Psychol. Sci.* 20, 1269–1274. doi: 10.1111/j.1467-9280.2009.02438.x

Matsuzaki, N., and Sato, T. (2008). The perception of facial expression from two-frame apparent motion. *Perception* 37:1560. doi: 10.1068/p5769

McLellan, T., Johnston, L., Dalrymple-Alford, J., and Porter, R. (2010). Sensitivity to genuine versus posed emotion specified in facial displays. *Cogn. Emot.* 24, 1277–1292. doi: 10.1080/02699930903306181

McLellan, T. L., Wilcke, J. C., Johnston, L., Watts, R., and Miles, L. K. (2012). Sensitivity to posed and genuine displays of happiness and sadness: a fMRI study. *Neurosci. Lett.* 531, 149–154. doi: 10.1016/j.neulet.2012.10.039

Meissner, C. A., and Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: a meta-analytic review. *Psychol. Public Policy Law* 7, 3–35. doi: 10.1037//1076-8971.7.1.3

Montepare, J., Goldstein, S., and Clausen, A. (1987). The identification of emotions from gait information. *J. Nonverb. Behav.* 11, 33–42. doi: 10.1007/bf00999605

Namba, S., Kabir, R. S., Miyatani, M., and Nakao, T. (2018). Dynamic displays enhance the ability to discriminate genuine and posed facial expressions of emotion. *Front. Psychol.* 9:672. doi: 10.3389/fpsyg.2018.00672

O'Toole, A. J., Roark, D. A., and Abdi, H. (2002). Recognizing moving faces: a psychological and neural synthesis. *Trends Cogn. Sci.* 6, 261–266. doi: 10.1016/s1364-6613(02)01908-3

Pike, G. E., Kemp, R. I., Towell, N. A., and Phillips, K. C. (1997). Recognizing moving faces: the relative contribution of motion and perspective view information. *Vis. Cogn.* 4, 409–437.

Pilz, K. S., Thornton, I. M., and Bülthoff, H. H. (2006). A search advantage for faces learned in motion. *Exp. Brain Res.* 171, 436–447. doi: 10.1007/s00221-005-0283-8

Pitcher, D., Duchaine, B., and Walsh, V. (2014). Combined TMS and fMRI reveal dissociable cortical pathways for dynamic and static face perception. *Curr. Biol.* 24, 2066–2070. doi: 10.1016/j.cub.2014.07.060

Porter, S., and ten Brinke, L. (2010). The truth about lies: what works in detecting high-stakes deception? *Legal Criminol. Psychol.* 15, 57–75. doi: 10.1348/135532509x433151

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

Sato, W., Kochiyama, T., Yoshikawa, S., Naito, E., and Matsumura, M. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study. *Cogn. Brain Res.* 20, 81–91. doi: 10.1016/j.cogbrainres.2004.01.008

Schiff, W., Banka, L., and Galdi, G. D. (1986). Recognizing people seen in events via dynamic "mug shots". *Am. J. Psychol.* 99, 219–231.

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. Nonverb. Behav.* 30, 37–52. doi: 10.1007/s10919-005-0003-x

Smoski, M. J., and Bachorowski, J. A. (2003). Antiphonal laughter between friends and strangers. *Cogn. Emot.* 17, 327–340. doi: 10.1080/02699930302296

Steede, L., Tree, J., and Hole, G. (2007). Dissociating mechanisms involved in accessing identity by dynamic and static cues. *Vis. Cogn.* 15, 116–119.

Thornton, I. M., and Kourtzi, Z. (2002). A matching advantage for dynamic human faces. *Perception* 31, 113–132. doi: 10.1068/p3300

Trautmann, S. A., Fehr, T., and Herrmann, M. (2009). Emotions in motion: dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain Res.* 1284, 100–115. doi: 10.1016/j.brainres.2009.05.075

Trautmann-Lengsfeld, S. A., Dominguez-Vorras, J., Escera, C., Herrmann, M., and Fehr, T. (2013). The perception of dynamic and static facial expressions of happiness and disgust investigated by ERPs and fMRI constrained source analysis. *PLoS One* 8:e66997. doi: 10.1371/journal.pone.0066997

van der Schalk, J., Hawk, S. T., Fischer, A. H., and Doosje, B. (2011). Moving faces, looking places: validation of the Amsterdam dynamic facial expression set (ADFES). *Emotion* 11, 907–920. doi: 10.1037/a0023853

Vuilleumier, P., Armony, J., Driver, J., and Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nat. Neurosci.* 6, 624–631. doi: 10.1038/nn1057

Wiese, H., Altmann, C. S., and Schweinberger, S. R. (2014). Effects of attractiveness on face memory separated from distinctiveness: evidence from event-related brain potentials. *Neuropsychologia* 56, 26–36. doi: 10.1016/j.neuropsychologia.2013.12.023

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

Wurm, L. H., Vakoch, D. A., Strasser, M. R., Calin-Jageman, R., and Ross, S. E. (2001). Speech perception and vocal expression of emotion. *Cogn. Emot.* 15, 831–852. doi: 10.1080/02699930143000086

Xiao, N. G., Perrotta, S., Quinn, P. C., Wang, Z., Sun, Y. H. P., and Lee, K. (2014). On the facilitative effects of face motion on face recognition and its development. *Front. Psychol. Emot. Sci.* 5:633. doi: 10.3389/fpsyg.2014.00633

Zhongqing, J., Wenhui, L., Recio, G., Ying, L., Wenbo, L., Doufei, Z., et al. (2014). Pressure inhibits dynamic advantage in the classification of facial expressions of emotion. *PLoS One* 9:e100162. doi: 10.1371/journal.pone.0100162

Zloteanu, M., Krumhuber, E. G., and Richardson, D. C. (2018). Detecting genuine and deliberate displays of surprise in static and dynamic faces. *Front. Psychol.* 9:1184. doi: 10.3389/fpsyg.2018.01184

Conflict of Interest: 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: 30 November 2020
doi: 10.3389/fpsyg.2020.602385

Identifying Emotional Expressions: Children's Reasoning About Pretend Emotions of Sadness and Anger

Elisabet Serrat, Anna Amadó, Carles Rostan, Beatriz Caparrós and Francesc Sidera*

Department of Psychology, University of Girona, Girona, Spain

This study aims to further understand children's capacity to identify and reason about pretend emotions by analyzing which sources of information they take into account when interpreting emotions simulated in pretend play contexts. A total of 79 children aged 3 to 8 participated in the final sample of the study. They were divided into the young group (ages 3 to 5) and the older group (6 to 8). The children were administered a facial emotion recognition task, a pretend emotions task, and a non-verbal cognitive ability test. In the pretend emotions task, the children were asked whether the protagonist of silent videos, who was displaying pretend emotions (pretend anger and pretend sadness), was displaying a real or a pretend emotion, and to justify their answer. The results show significant differences in the children's capacity to identify and justify pretend emotions according to age and type of emotion. The data suggest that young children recognize pretend sadness, but have more difficulty detecting pretend anger. In addition, children seem to find facial information more useful for the detection of pretend sadness than pretend anger, and they more often interpret the emotional expression of the characters in terms of pretend play. The present research presents new data about the recognition of negative emotional expressions of sadness and anger and the type of information children take into account to justify their interpretation of pretend emotions, which consists not only in emotional expression but also contextual information.

Keywords: sadness, anger, pretend emotions, children, emotional expression

OPEN ACCESS

Edited by:
Jian K. Liu,
University of Leicester,
United Kingdom

Reviewed by:
Taiyong Bi,
Zunyi Medical University, China
Yangmei Luo,
Shaanxi Normal University, China
Zulay Lugo,
Institut Pere Mata, Spain

****Correspondence:***
Elisabet Serrat
elisabet.serrat@udg.edu

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

Received: *03 September 2020*
Accepted: *26 October 2020*
Published: *30 November 2020*

Citation:
Serrat E, Amadó A, Rostan C,
Caparrós B and Sidera F (2020)
Identifying Emotional Expressions:
Children's Reasoning About Pretend
Emotions of Sadness and Anger.
Front. Psychol. 11:602385.
doi: 10.3389/fpsyg.2020.602385

INTRODUCTION

This study explores children's capacity to comprehend that the emotions expressed in pretend play contexts may have playful intentions. This capacity to detect emotions simulated by other people is considered important because it helps people to identify reliable individuals and establish positive and trusting relationships with others, and to communicate effectively in social contexts (Saarni et al., 2007; Walle and Campos, 2014). Specifically, this research focuses on how the ability to discriminate facial expressions of emotion is developed, but with particular emphasis on the more specific ability of detecting pretend emotion (or emotions simulated in pretend play contexts), so as to understand how children explain their interpretations of pretend facial expressions. In addition, this emotional recognition is studied in the context of pretend play, where the simulation of emotions often occurs in childhood. Regarding this, it is assumed that contextual information is fundamental in the recognition of facial expressions and pretend emotions. We will now discuss these aspects of the study in more detail.

Children's Recognition of Emotional Expressions

Recognizing the emotions of other people through their facial expression is important in human relationships. It is an essential ability in interpersonal interactions, since it allows us to behave properly in different social contexts, is an important aspect of interpersonal communication, and is crucial in regulating the behavior of others (Saarni, 1999; Scharfe, 2000).

Throughout their development, children progress in their ability to identify and understand facial expressions associated with emotions. Babies begin to discriminate emotions in facial expressions during the first year of life (Morgan et al., 2010). Toward the end of the first year and at the beginning of the second, babies try to give meaning to situations based on the information obtained from the emotional expressions of others, a skill that has been called social reference (Sorce et al., 1985; Widen and Russell, 2008). Therefore, at this age children already understand that people's emotions have a meaning linked to external events, and after 14 months they are able to identify where the emotion is directed (Repacholi, 1998), and to match some negative emotions to specific eliciting events (Ruba et al., 2019).

Developing the recognition of facial emotional expressions can be understood as a process of increasing expertise in the ability to discriminate emotions (Widen, 2013). According to some authors the ability to recognize basic emotional expressions could begin with the distinction between two broad categories - feel good, feel bad (Widen and Russell, 2010; Widen, 2013) - and improves throughout childhood and adolescence, although there are emotions for which the level of recognition is similar between the ages of 6 and 16, as in the case of joy, sadness, and anger (Lawrence et al., 2015).

Despite there being discrepancies in specific aspects of emotion recognition depending on the method used in the study, there is a consensus that children begin to identify four basic emotions at 3 years of age: joy, fear, sadness, and anger (Pons et al., 2004; Székely et al., 2011). Pons et al. proposed that, apart from recognizing emotions from facial expressions, children up to the age of 5 also begin to understand the causes of emotion. Later, and up to 7 years, children understand the mental nature of emotions and the possibility of hiding them. And in a third period, between 9 and 11 years of age, children understand ambivalence in emotions, moral emotions, and the cognitive regulation of emotions. The above study also indicated that understanding the external aspects of emotions is a prerequisite for understanding internal psychological aspects.

Previous research has shown that 6-year-olds can recognize some emotions - joy, sadness, and anger - in a similar manner to adolescents (Lawrence et al., 2015). But little (or less) is known about identifying these emotions in pretend situations. In this respect, in the present research two emotions with negative valence - sadness and anger - were selected to study the recognition of emotional expression in pretend play contexts. In addition, these two emotions are the first two negative emotions that children usually recognize (Widen, 2013), and may therefore also be the first ones to be interpreted in terms of pretend emotions. The emotion of happiness is usually the first to be labeled by children in free labeling tasks (Widen, 2013), but Sidera et al. (2011) pointed out methodological difficulties when studying children's understanding of pretend happiness (children might interpret that pretending to be happy makes one actually happy), so we decided not to include pretend happiness in the present research.

Although many studies have been conducted on the recognition of basic emotional expressions in early childhood and in later development, few have focused on the recognition of pretend emotions. And despite this fact, emotions are often hidden or simulated for different reasons in everyday interpersonal communication and social relationships (see Zeman and Garber, 1996). Thus, the focus of this article is on this emotional simulation, and specifically, on pretend play situations where children express emotions that are different from their real ones for play purposes.

Children's ability to understand that the real emotion of a person may differ from their emotional expression has often been studied in contexts of deception (see Sidera et al., 2013). Children begin to control their emotional expressions at the age of 4; however, at that age they are not yet able to deceive other people through emotional expression. This latter ability is closely related to the ability to understand that internal emotion and external emotion may differ, which usually develops between the ages of 4 and 6 (Harris et al., 1986; Pons et al., 2004; Misailidi, 2006; Sidera et al., 2012; Kromm et al., 2015).

Some studies have shown that children aged 8 to 12 may have difficulties in discriminating genuine from non-genuine emotional expressions (Dawel et al., 2015). The aforementioned authors found that children have particular difficulty with sadness, but not joy; adults have been found to be better at detecting both. In fact, at the age of 4 children are already able to explicitly discriminate between Duchenne vs. non-Duchenne smiles, and implicitly at the age of 3 (Song et al., 2016). That said, Dawel et al. (2015) considered that the skills required to carefully determine the authenticity of emotions from facial information mature at a later stage. There is evidence, then, that discriminating between genuine and pretend sadness is difficult in childhood, although we do not have enough information to know whether the same is true of anger (see Felleman et al., 1983).

Regarding research on emotional expression in contexts of play, the study by Mizokawa (2011), where children were presented with picture stories in a play context and in a non-play context, showed that 4- and 5-year-old were better at distinguishing pretend crying from real crying in a pretend play context than in a non-play context. So Mizokawa (2011) suggested that the context of pretend play facilitates children's understanding of pretend crying. Furthermore, the study by Sidera et al. (2013) showed that at the age of 4, despite not mastering the distinction between internal and external emotion, most children understand the playful intentionality of emotions expressed in pretend play contexts. Specifically, Sidera et al. found that 4-year-old children are capable of understanding that when a character, or themselves, show a sad expression in a pretend play context, this person is just pretending, and is not really sad.

The present research aims to broaden the results of these studies by analyzing in greater depth the type of reasoning children use when it comes to detecting pretend emotions. This will allow us to identify which elements children who understand pretend emotions take into account that other children do not.

Context and Emotional Expression Recognition

In daily life, faces are not usually seen in isolation. On the contrary, they appear in a multisensory context that includes aspects such as a voice, body posture and movement, or other people, and the recognition of facial expression is influenced by this context. Contextual influences are perceived early and automatically, and information provided by the facial expression is combined with that of the context (Righart and de Gelder, 2008).

The present study analyzes in greater depth children's identification of and reasoning about simulated facial expressions with negative valence in a natural and playful context. We assume that this recognition incorporates contextual information in a natural and routine way, meaning that this is important for inferring the meaning of facial expressions. Hence, emotional perception is not only guided by the structural configuration of a person's facial actions, but also from the context in which a face is encoded (Barrett et al., 2011), while we can also state that very young children appear to use contextual congruency, among other cues, to detect the authenticity of emotions (Walle and Campos, 2014).

Following Barrett et al. (2011), we consider that, although faces carry emotional information, their emotional meaning is constructed from the context in which they are embedded, and that people infer emotional meaning from facial movement and other social information (Barrett et al., 2019). In line with this, Keltner et al. (2019) hold that people's interpretation of a target's emotional expression is influenced by factors such as the following: who expresses the emotion (e.g., their gender); the mental states attributed to that person; the context (e.g., the action being undertaken by the person expressing the emotion); and the emotional expressions of the surrounding people.

Some studies with adults have shown the positive influence contextual information has on recognizing facial expressions (for a review, see de Gelder et al., 2006). There are also studies with children that have shown how a congruent visual context increases emotional recognition in children (Theurel et al., 2016), although other studies offer less clear results (Reichenbach and Masters, 1983; Nelson and Russell, 2011). Theurel et al. (2016) pointed out that there are methodological questions to consider here, and suggested that context may help to disambiguate the meaning of emotional expressions (e.g., sadness and fear, which it takes children a long time to discriminate between).

There is a discussion in this field regarding whether facial expression is the best clue for recognizing emotions in comparison to other sources of information (Face Superiority Effect), at least in early developmental stages (Denham, 1998). For example, Balconi and Carrera (2007) found that children recognize the emotions of joy and sadness better from facial expressions than from a story. However, they also found that the opposite is true with fear and disgust (Story Superiority Effect). These results suggest that the developmental order in which certain emotions are acquired is relevant when considering which informational sources are better for recognizing them. In line with this, the study by Nelson et al. (2013) also found that explaining a story provides a better clue for later emerging emotions than static or dynamic facial expressions (in this case, for the emotion of fear), whereas in children aged 3 to 5 and for the emotions of sadness and anger, still faces or videos were better than stories. To sum up, these results show that contextual information could be more important than facial expression in the recognition of emotions for complex emotions. This might also be the case for pretend emotions. In the cases of sadness and anger, the results published by Nelson et al. (2013) showed that children basically relied on still faces to detect anger, while for sadness they relied on videos showing different emotional cues (facial expression, voice, body posture, and movement).

In sum, despite prior research showing that some children aged 4 years are capable of understanding pretend sadness, some children are not. This previous research has not analyzed the reasoning children use when interpreting emotions expressed in pretend play contexts. Doing so would help to understand how children use contextual information in interpreting pretend emotions, and why some children understand emotions expressed in pretend play contexts as real. Moreover, prior research has studied how children understand pretend sadness, but we do not know whether children understand other negative emotions in a similar way. Therefore, as the labels for sadness and anger are the first labels for negative emotions that children acquire (Widen and Russell, 2003; Maassarani et al., 2014), we decided to study both the emotions of anger and sadness.

The aim of the present study, then, is to provide a more in-depth understanding of children's recognition of and reasoning about pretend emotional expressions expressed in pretend play contexts. In this sense, we explored whether children's capacity to detect pretend sadness and pretend anger vary by age or emotion. Moreover, to explore the role played by contextual information in identifying pretend emotions, we asked children to justify their interpretation of a pretend emotion in order to study what importance they award to information gleaned from facial expression as opposed to context. Specifically, our objectives are first, studying children's recognition of simulated emotions in pretend play contexts for two emotions of negative valence; and second, exploring what kind of information children consider when performing this recognition.

MATERIALS AND METHODS

Participants

The initial sample was comprised of 91 hearing children (46 girls and 45 boys) with a mean age of 72.74 months; *SD* = 18.86; range: 39 to 107 months. An initial emotion recognition task was administered in order to avoid difficulties with the pretend emotion understanding tasks. Thus, those children who failed to recognize the facial expressions of anger and sadness from the facial emotion recognition task were not included in the final sample of the study.

The final sample was comprised of 79 children (42 girls and 37 boys; mean age = 74.14 months; *SD* = 18.00; range = 40 to 107 months), who were separated into two age groups. Participants were divided into these two groups so as to detect developmental differences; the young group included children from preschool years, while the older group contained children from the first years of primary school. The group of young children was comprised of 36 children aged 3 to 5 (20 girls and 16 boys; mean age = 57.42 months; *SD* = 9.16; range = 40 to 71 months), and the older group contained 43 children aged 6 to 8 (22 girls and 21 boys; mean age = 88.14 months; *SD* = 9.56; range = 72 to 107 months). The Chi-Square test showed that there was not a significant relation between age group and gender ($p > 0.05$), so the sex distribution was similar between both age groups.

In **Table 1**, we describe some of the demographic information related to the children, separated by age groups.

In the area where the study was conducted, the main school languages were Catalan and Spanish, so all the selected children knew at least one of the two languages, and all families but three informed us that at least one of the parents communicated to their child either in Catalan or Spanish.

The children were recruited from four state-run schools in Spain. Written informed consent was obtained from parents before administering the tasks to their children. None of the children were reported to have cognitive delays.

Materials

Five experimental tasks were administered to the children, and their teachers were asked to complete two questionnaires. However, of the five experimental tasks, only the following three tasks are considered in the present study (thus, the results of an expressive vocabulary task and a pretend actions task are not considered here):

TABLE 1 | Mean values (and *SD*) in different variables related to the participants as a function of the age group.

	Younger children	Older children	Age group comparison (Mann-Whitney)
Number of siblings	0.79 (0.74)	1.15 (0.70)	$U = 341.500$ $Z = -2.241$ $p < 0.05$
Age of schooling	1.95 (1.02)	2.60 (0.81)	$U = 129.500$ $Z = -2.333$ $p < 0.05$
Level of studies of the father	1.95 (0.85)	2.00 (0.84)	$U = 165.000$ $Z = -0.193$ $p > 0.05$
Level of studies of the mother	1.95 (0.74)	2.20 (0.83)	$U = 172.500$ $Z = -1.041$ $p > 0.05$

The variable level of studies had 4 categories: 0 = does not have studies; 1 = primary education level; 2 = secondary education level; 3 = higher studies.

(a) Facial emotion recognition task. A task of facial emotion recognition (FER) was included to ensure children did not fail the pretend emotions task due to difficulties recognizing the emotions of anger and sadness. The FER task included six drawings of a girl showing six basic emotions (happy, sad, scared, angry, surprised, and disgusted; the drawings for the task are included in Sidera et al., 2017). The six drawings were placed in two lines in front of the child, and the experimenter labeled the emotions one by one (following a Latin-square design to counterbalance the order of presentation). After identifying a label for one emotion, the experimenter asked the child "*Could you point to the girl looking...?*" After this question, the experimenter said "OK" and proceeded to label the following emotion. In the present study, only the results of the emotions of anger and sadness were considered. Children who pointed correctly to the faces of anger and sadness were included in the final sample of the study, whereas those who failed at least one of these emotions were excluded.

(b) Pretend emotions task. A task with silent videos (lasting about one minute each) was used to evaluate children's reasoning that the emotions used in pretend play contexts may be expressed with playful purposes. This task consisted of a warm-up phase and a test phase. In the warm-up, children were again shown the drawings of sadness and anger from the FER task, and were asked about the emotion expressed by the girl in each of the two drawings: "Can you tell me how this girl feels"? Children who responded incorrectly were given corrective feedback (the correct label was stated). Moreover, in order to make children familiar with the words that the present study used to refer to the distinction between pretense and reality (we used two Catalan expressions for making this difference: "de veritat" and "de mentida"), the experimenter performed some real actions and some pretend actions. First of all, the experimenter did two actions without feedback, and then four more actions with feedback. The first action without feedback was the real action of drinking water. Before doing the action, the experimenter explained it to the participant: "Now I'm going to do a real action, ok? I will really drink water." Then, the experimenter drank some water from a glass, and said: "Did you see? I really drank water." The second action without feedback was pretending to drink water. The experimenter also explained the action beforehand ("Now I am going to do a pretend action. I will pretend to drink water"). The pretend action was carried out in an obvious pretend way (the glass was empty, the lips did not touch the glass, and the movements were exaggerated as it is usual in pretend play), and after the action the experimenter said: "Did you see? I pretended to drink water. I pretended to drink, but in reality I did not drink." Afterwards, the experimenter carried out the four actions with feedback. Before carrying out each action, the experimenter said: "Ok, X, now I am going to do an action and you have to tell me whether it is a real or a pretend action, ok?" After that, the experimenter carried out the action and asked whether it was real or pretend. For example: "X, am I really cutting the paper or am I pretending to cut the paper?" After their response, children were given corrective feedback. For example: "Yes, very good, I pretended to cut the paper with the scissors, but I did not really cut it," or "Really cutting the paper? No, I pretended to cut the paper with the scissors, but I did not really cut it."

In the test phase, eight silent videos were presented of children acting out real or pretend emotions (four videos of real emotions and four of pretend emotions) following a Latin-square design to counterbalance the order of presentation. However, in the present study we were interested in how children reason about emotions expressed in pretend play contexts, so only the four videos depicting pretend emotions were analyzed. In these videos, two characters were practicing pretend play and at the end the image froze with one of the characters (the "protagonist") expressing pretend anger or pretend sadness (two videos of each emotion were used). In the pretend sadness videos, one character played the role of the baby and the other the role of the mother; the mother became angry after the baby misbehaved (did not want to eat or sit down in a chair), so the baby pretended to be sad. In the pretend anger videos, two children pretended that a doll was misbehaving (throwing pretend food or knocking down a tower of blocks) and one of them pretended to be angry toward the doll.

At the end of each video, while the image was frozen, two questions were asked about the protagonists:

Test question: *"Is the child really angry/sad or is she pretending to be angry/sad?"*
Justification: *"Why do you think she is angry/sad (or pretending to be angry/sad)?"*

(The word "angry" was used for the pretend anger videos and the word "sad" for the pretend sadness videos).

Hence, the test question evaluated whether children understood the expressed emotion as real or pretense, and thus that emotions may have a pretend purpose. One point was given for each correct answer in the test questions, so the total score for the pretend emotions task ranged from 0 to 4. Regarding justifications, they were divided into the following categories:

1. *Emotion*. When children justified their response to the test question with reference to the emotional expression or the emotion of the protagonist.
2. *Event/behavior.* When children justified their response to the test question by referring to the event in the video that triggered the protagonist's emotion (e.g., *"because the doll knocked down the tower"*), or when they referred to the protagonist's behavior (*"because the girl is telling the doll off"*).
3. *Play.* When children justified their response to the test question by arguing that the protagonist was playing (e.g., *"they were just pretending with the doll"*) or explaining that the children were just pretending so the emotion of the protagonist must be understood as pretense (e.g., *"because it was the girl who knocked down the tower, not the doll").*
4. *Non-response.* When children did not answer, said they did not know the answer, or gave a non-sensical answer.
5. *Other.* Answers that included more than one of the previous categories were included in this category.

Two authors of the study categorized all responses into one of the five previous categories, and their categorizations were compared. The number of observed agreements was 91.46% of the observations, while the Kappa equaled 0.884 (*SE* = 0.021). Differences between judges were resolved by discussion.

Finally, the categories event/behavior and play were merged in some analyses, as both include information related to the context of the story represented in the videos.

(c) Non-verbal cognitive ability test. The children's non-verbal ability was evaluated by means of the Pattern Construction subtest from the British Ability Scales, 2nd edition (Spanish version by Arribas and Corral, 2011). The Ability Scores of the test were used, as they consider the specific items administered to each child.

Aside from the tasks administered to the children, their teachers were also asked to respond to a language assessment questionnaire [the Language Proficiency Profile LPP-2 by Bebko and McKinnon (1993)], the data from which were not used in the present study. The teachers also responded to a demographic questionnaire in order to provide background information about the children (date of birth, number of siblings, school enrollment, existence of learning difficulties, parental education, mother tongue of the parents, and language used with the child).

Procedure

The children were tested in a quiet room in their schools. Administration of the tasks lasted between 35 and 55 min and took place in one session. The data were analyzed using IBM SPSS version 23. Non-parametric tests were used, as the data did not meet the criteria of normal distributions. The one-sample Wilcoxon signed-rank test was used to compare children's scores to chance level. The Mann-Whitney's *U* test was used to compare the scores between the two age groups. The Wilcoxon test was used to compare the scores of the pretend sadness with the pretend anger videos. Finally, the Chi-Square test was used to compare frequencies of responses between the different justification categories.

RESULTS

Scores for the Pretend Emotions Task

The younger children's mean in the pretend emotions task was 2.5 (out of 4; *SD* = 1.30) and the older children's mean was 3.84 (*SD* = 0.37), close to the maximum. Children's scores in each age group were compared to chance expectation (two points was considered as the chance level, because the task involved four dichotomous responses) using the one-sample Wilcoxon signed-rank test. Both groups obtained scores above chance (young group: $Z = 2.043$, $p = 0.041$; older group: $Z = 6.168$, $p < 0.001$). On the other hand, Mann-Whitney's U test showed significant differences between the two age groups ($U = 281.500$; $Z = -0.5447$; $p < 0.001$). The mean percentile in the non-verbal cognitive ability score was 60.42 ($SD = 25.15$) in the young group and 63.84 ($SD = 20.55$) in the older group, and according to Mann-Whitney's U test no significant differences existed between the two age groups in terms of the percentile of non-verbal ability ($p > 0.05$).

When type of emotion was taken into account, age differences were observed for both sadness and anger, the older children doing better than the younger children. Moreover, both the younger and older children obtained better scores in the pretend sadness videos than in the pretend anger videos (see **Table 2**). When we compared children's scores for each type of emotion at each age group to the expected chance level (1 point), we observed that older children obtained scores above chance in both emotions (anger: $Z = 6.000$, $p = 0.000$; sadness: $Z = 6.557$, $p < 0.001$), while young children scored above chance for sadness ($Z = 3.889$, $p < 0.001$) but not for anger ($p = 0.414$).

On the other hand, the development of the recognition of pretend sadness and pretend anger is shown in **Figure 1**. The same pattern of development is observed, but with a better performance for the emotion of sadness than for that of anger. Specifically, children reached the maximum score for sadness at the age of 5, and a near-to-ceiling score for anger at the age of 6.

TABLE 2 | Means (and SD) for the pretend emotions task by age and type of emotion.

	Anger	Sadness	Anger-sadness comparison (Wilcoxon)
Younger children	0.89 (0.82)	1.61 (0.73)	$Z = -3.802$ $p < 0.001$
Older children	1.84 (0.37)	2 (0)	$Z = -2.646$ $p = 0.008$
Age group comparison (Mann-Whitney)	$U = 292.000$ $Z = -5.361$ $p < 0.001$	$U = 580.000$ $Z = 1246.500$ $p = 0.001$	

score range 0–2.

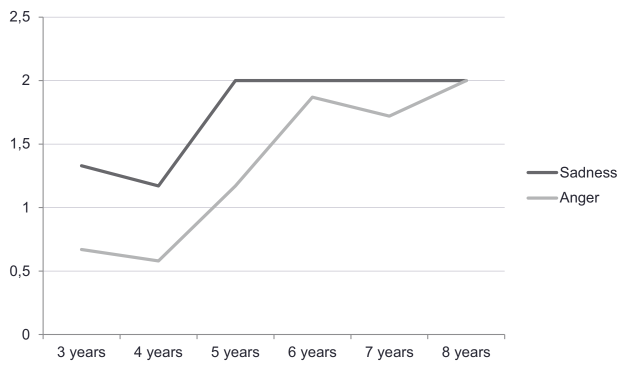


FIGURE 1 | Developmental understanding of pretend sadness and pretend anger as a function of age.

TABLE 3 | Mean number (and SD) for justifications used in the four scenarios of the pretend emotions task.

	Emotion	Event/behavior	Play	Non-response
Younger children	0.75 (1.08)	1.81 (1.37)	0.75 (1.08)	0.69 (1.17)
Older children	1.33 (1.25)	0.91 (1.13)	1.40 (1.07)	0.35 (0.923)
Age group comparison (Mann-Whitney)	$U = 5611.500$ $Z = -2.221$ $p = 0.026$	$U = 486.000$ $Z = -2.990$ $p = 0.003$	$U = 485.000$ $Z = -2.987$ $p = 0.003$	$U = 625.000$ $Z = -1.917$ $p = 0.055$

As there were four videos, the maximum number of justifications for each category was 4.

Justifications for Responses in the Pretend Emotions Task

Regarding the mean number of each type of justification used by the children (see **Table 3**), we observed that in young children the most used category was event/behavior, while older children mostly used the categories emotion and play. Furthermore, age differences were found in how the children justified pretend emotions: among the older children, there was a significantly higher use of the emotion and play categories and lower use of the event/behavior category. Also, the decrease with age in the number of justifications in the non-response category was close to significant.

Following this, we analyzed the type of justification as a function of the emotion involved in the videos (anger vs. sadness; see **Table 4**). The children were found to use different types of justification as a function of emotion type. In the anger videos, children mostly used play and event/behavior justifications, while in the sadness videos the most commonly used justifications were emotion and event/behavior. In fact, the category emotion was used significantly more in the sadness than in the anger videos, while the children more frequently used the play category for the anger videos than for the sadness videos.

We also analyzed the use of the different categories in the anger and sadness situations as a function of the age group (see **Table 5**). For the emotion category both the young and the older group followed the same pattern: children used this type of justification more in the sadness than in the anger situation. For the event/behavior category, the young group showed a higher use of this category in the anger situations, and no differences existed in the older group. The opposite occurred in the play

TABLE 4 | Mean number (and SD) for justifications used in the pretend emotions task as a function of emotion type.

	Emotion	Event/behavior	Play	Non-response
Anger videos	0.23 (0.53)	0.73 (0.83)	0.85 (0.83)	0.19 (0.51)
Sadness videos	0.84 (0.87)	0.56 (0.75)	0.28 (0.58)	0.32 (0.67)
Anger-sadness comparison (Wilcoxon)	$Z = -5.202$ $p < 0.001$	$Z = -1.737$ $p = 0.082$	$Z = -4.739$ $p < 0.001$	$Z = -1.978$ $p = 0.048$

As there were two videos for each type of emotion, the maximum number of justifications for each category was 2.

TABLE 5 | Mean number (and SD) for justifications used in the pretend emotions task as a function of emotion type and age group.

		Emotion	Event/behavior	Play	Non-response
Young group	Anger	0.17 (0.45)	1.08 (0.87)	0.44 (0.74)	0.31 (0.62)
	Sadness	0.61 (0.77)	0.72 (0.78)	0.28 (0.62)	0.39 (0.73)
Anger-sadness comparison (Wilcoxon)		$Z = -3.234$ $p = 0.001$	$Z = -2.166$ $p = 0.030$	$Z = -1.261$ $p = 0.207$	$Z = -0.758$ $p = 0.448$
Old group	Anger	0.28 (0.59)	0.44 (0.67)	1.19 (0.76)	0.09 (0.37)
	Sadness	1.02 (0.91)	0.42 (0.70)	0.28 (0.55)	0.26 (0.62)
Anger-sadness comparison (Wilcoxon)		$Z = -4.122$ $p < 0.001$	$Z = -0.179$ $p = 0.858$	$Z = -4.786$ $p < 0.001$	$Z = -2.333$ $p = 0.020$

category: the older group showed a higher use in the anger situation than in the sadness situation. Finally, the older group showed a higher use of the non-response category in the sadness situation, but no differences existed in the young group.

In order to evaluate the relationship between each justification category and correct responses to the test questions for the pretend emotions task, the proportion of correct responses was calculated for each justification category. The proportion of justifications labeled as play and considered correct was 0.98. This proportion was 0.89 for the emotion category, 0.70 for the non-response category, and 0.63 for the event/behavior category. A Chi-Square test revealed significant differences in the proportion of correct responses between categories ($\chi^2 = 44.601$, $p < 0.001$). When the proportions of correct responses were compared between the different categories in pairs, significant differences were observed between the following categories: emotion and event/behavior ($\chi^2 = 16.845$, $p < 0.001$), emotion and play ($\chi^2 = 5.390$, $p = 0.020$), emotion and non-response ($\chi^2 = 6.996$, $p = 0.008$), play and event/behavior ($\chi^2 = 35.578$, $p = 0.000$), and play and non-response ($\chi^2 = 22.236$, $p < 0.001$). No differences were found between the categories event/behavior and non-response ($p > 0.05$).

Proportional use of the emotion and contextual categories (the latter including the categories event/behavior and play) was compared in children who gave correct responses. The children responded with a contextual category in 0.59 of correct responses; with an emotion category in 0.29 of cases; with a non-response category in 0.11 of cases; and in the other category in 0.004 of cases. Therefore, the contextual category was the most widely used category for correct responses. However, when the variable type of emotion was taken into account, a higher use of the contextual category over the emotion category (for correct responses) was noted for the anger videos, but not for the sadness videos (see **Table 6**). Thus, for the sadness videos, the proportion of correct responses was very similar in the contextual and emotion categories. When the Chi-Square test was used to compare the proportion of correct responses for emotion vs. contextual in the anger and sadness videos, significant differences were observed between the two conditions in the use of the two categories ($\chi^2 = 38.234$, $p < 0.001$).

TABLE 6 | Proportional use of the different justifications in the pretend emotions task as a function of type of emotion and age group.

		Emotion	Contextual	Non-response	Others
All children	Anger videos	0.11	0.82	0.07	
	Sadness Videos	0.43	0.42	0.14	0.01
Young group	Anger videos	0.03	0.84	0.13	0
	Sadness Videos	0.31	0.53	0.16	0
Old group	Anger videos	0.14	0.81	0.05	0
	Sadness Videos	0.51	0.35	0.13	0.01

Finally, the proportional use of the emotion and contextual categories (in children who gave correct responses) was compared as a function of the age group for each emotion (see **Table 6**). In the young group, the contextual category was the most used both for the anger and sadness videos. The older group also used the contextual category more than the emotion category for the anger videos, but they used the emotion category more for the sadness videos. The Chi-Square confirmed that the use of the emotion and contextual categories was similar in the young and older groups for the anger videos ($p > 0.05$), while the age groups differed in the frequency of the use of these categories for the sadness videos ($\chi^2 = 6.028$, $p = 0.014$).

DISCUSSION

Previous literature has shown that young children are capable of realizing other children may pretend to be sad for playful reasons (see Sidera et al., 2013). The current study found that this knowledge develops gradually between the ages of 4 and 6, and is well-established from the age of 6 years, as all children in the older group recognized the expression of pretend sadness. This older group also performed well at the pretend anger task, although significant differences in their understanding of pretend anger and pretend sadness existed. Furthermore, younger children (aged 3 to 5) performed worse than older children, especially in the pretend anger task. So while the young group showed some understanding that sadness may be expressed for playful intentions, their scores in the pretend anger tasks were not above chance level. More research is needed to confirm the possibility that children's capacity to interpret pretend sadness is better than their capacity to interpret pretend anger, and to discard the possibility that methodological differences between the two tasks accounted for the differences we found.

The results of this study suggest that recognizing negative pretend emotions (for anger and sadness) is easier than discriminating genuine from non-genuine expressions. This statement is supported if we compare our results to those of Dawel et al. (2015); while they found that compared to adults, children aged 8–12 had difficulties discriminating genuine sadness, in our study even young children recognized expressions of pretend sadness quite well. In sum, we can state that being able to recognize facial expressions is not enough for recognizing pretend emotions (let us recall that all children in our study had successfully completed a task of recognizing the emotions of sadness and anger, but some failed the pretend emotions task); that said, children are capable of identifying pretend sadness before they are capable of distinguishing genuine from non-genuine facial expressions of sadness out of context (as in Dawel et al. study); hence, it is likely that children use other informational cues to identify pretend emotions. Indeed, the capacity to use contextual information to detect authenticity in emotional expressions appears very early. Walle and Campos (2014) showed that 19-month-olds are capable of detecting authenticity in emotional expressions based on contextual information (expressing pain in a situation where a hammer did not hit the hand).

In this study, we grouped the types of reasoning children use to justify whether an emotion is pretend or not into three categories: referring to the protagonist's emotional expression (emotion), considering the context of pretend play (play), or referring to the context of the story depicted in the video and/or the behavior of its protagonists (event/behavior). The results showed differences between the age groups, since the younger children (3- to 5-year-olds) tended to justify the pretend emotions expressed by the protagonists of the story by referring to their context or the behavior of their characters, while the older children (6- to 8-year-olds) mostly referred to the protagonist's emotional expression and the context of pretend play. Furthermore, the justifications emotion and play, which were used more frequently by the older children, were the ones most associated with correct responses. Therefore, we can state that there are developmental differences in how children explain whether an expressed emotion is pretend or real: as children grow older they do not consider the general context of the story as much, but rather focus more on the fact that the depicted story is set in a pretend play situation; similarly, they do not focus much on the general behavior of the protagonist, but specifically on their emotional expression. These results are in accordance with those found by Sidera (2009), whose study involved children being told stories where the protagonists simulated sadness or happiness in a pretend play context. When they were asked to justify the external and internal emotion of the protagonist, 6-year-olds were more capable of considering that the protagonists were involved in a playful situation than 4-year-olds.

In our study, we observed differences in the reasoning children used according to the type of pretend emotion expressed in the videos. For anger, most children referred to the play context or to the event described in the story or the behavior of the protagonist rather than the protagonist's emotional expression. This was true for both age groups, although older children more frequently used the play category, in accordance with the developmental differences commented above. For sadness, children's justifications were mostly based on the emotional expression of the protagonist. When age groups were considered we found that young children mostly used the event/behavior category, while older children mostly used the emotion category. Before we discuss a possible explanation of these differences, it is worth mentioning some of the justifications given by the children. We need to consider that a proportion of the children did not justify their response, and also that the least successful justification referred to the event and/or the behavior of the protagonist. This is possibly because the latter involves considering elements of context or behavior (beyond those related to playing) that are less relevant for interpreting facial expression correctly.

Success in the event/behavior category was near chance level. Therefore, the behavior or situation/event in which the emotional expression is integrated would not be useful in this situation for detecting pretend emotions, while knowledge of the general context of play in which the emotion is simulated would be. Therefore, as found in other studies (Balconi and Carrera, 2007; Nelson et al., 2013; Widen et al., 2015) with regard to later emerging emotions, prior history or, in this case, viewing the emotional expression to be identified in a story (where children play), facilitates recognition that the emotional expression is a pretend one. Children's references to the protagonist's emotional expression were also associated with correct responses, possibly because this is linked to children's capacity to capture the exaggerated elements of the facial expression. Although we cannot conclude this from the data in our study, the study by Walle and Campos (2014) does support the view that infants as young as 19 months of age are sensitive to exaggerated emotional displays and may use the level of exaggeration of an emotion in order to judge its authenticity or communicative value. Interestingly, in our study older children used mostly the emotion category for justifying pretend sadness, while they mostly used the play category for justifying pretend anger. We will try to interpret this next, by looking at the categories used for the correct responses.

Finally, when the results of the correct answers were only grouped into two categories (contextual vs. emotion), it was found that in the case of anger, children mostly used contextual clues (and not emotion) to judge whether the emotion was pretend or not (both in the young and the older group). In the case of sadness, children used both categories similarly when the whole sample was taken into account. But when age groups were considered, we found that young children's interpretations were more based on contextual cues while older children used emotion cues. Gnepp (1983) found that even preschoolers were capable of considering both emotional and contextual cues when presented with pictures where the facial expression of the protagonist was incongruent to the context. In this sense, the age changes in the justifications for pretend emotions would not be attributable to the inability of young children considering one or another type of cue. In this sense, a possible explanation for our results is that it is easier to detect (or express) pretend sadness than pretend anger from facial cues; this would explain why children relied more on the expressed emotion in the pretend sadness videos (since it was enough for children, especially for the ones in the older group, to interpret the communicative intention of the protagonist), whereas in the pretend anger videos children needed to seek more contextual cues (and especially cues related to the play situation in the older group) to interpret the pretend emotion and give an answer. Future research should clarify whether this explanation is correct, or whether differences are due to methodological issues.

This study had some limitations. First, silent videos were used to control for the influence of information from the intonation of speech, although obviously there is normally sound and language when we are exposed to the emotional expressions of others. Research into the recognition of emotional expressions by adults has shown that this is modulated by linguistic stimuli, and it is therefore necessary to advance the recognition of pretend emotions through more ecologically valid situations (Park and Itakura, 2019), which include the information provided from the prosody that accompanies speech as well as from some vocal bursts. Moreover, there is evidence that anger and sadness may be differentiated from the expression of other emotions in different modalities (Keltner et al., 2019), but it is yet to be investigated whether this is the case for pretend emotions. Similarly, the level of the intensity of the emotions from the facial emotion recognition task or from the pretend emotions task were not controlled. We must also bear in mind that in the present study children were asked to justify whether the emotions expressed by other children were pretense or not, meaning they were asked to give explicit responses, whereas if implicit behaviors were sought, then different, and perhaps interesting, results may also be obtained. Furthermore, the study of emotional expressions suggests that they are expressed in prototypical multimodal patterns of behavior with important variations (Keltner et al., 2019), a theory that also needs to be investigated for pretend emotions.

To sum up, then, in this study we have found that children aged 3 to 5 are capable of detecting pretend sadness in other children, at least in a contextualized situation, but still have difficulties with pretend anger. This may be due to the fact that facial cues are not as evident for pretend anger, and they have to seek more contextual cues. When doing so, older children are more aware when a character's behavior should be interpreted as pretend play, and therefore also interpret their emotional expressions in these terms.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Comité d'ètica i bioseguretat de la recerca de la Universitat de Girona. Reference: CEBRU0024-2019. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

All authors contributed to the study conception and design, commented on previous versions of the manuscript, read and approved the final manuscript. FS, ES, and AA prepared the material and collected and analyzed the data. ES and FS wrote the first draft of the manuscript.

FUNDING

This study was funded by the Spanish Ministerio de Economía y Competitividad and the Fondo Europeo de Desarrollo Regional (PSI2015-69419-R; MINECO-FEDER). Funds for open access were from Department of Psychology, University of Girona, Spain.

REFERENCES

Arribas, D., and Corral, S. (2011). *BAS2. Escalas de Aptitudes Intelectuales*. Madrid: TEA Ediciones.

Balconi, M., and Carrera, A. (2007). Emotional representation in facial expression and script: a comparison between normal and autistic children. *Res. Dev. Disabil.* 28, 409–422. doi: 10.1016/j.ridd.2006.05.001

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., and Pollak, S. D. (2019). Emotional expressions reconsidered: challenges to inferring emotion from human facial movements. *Psychol. Sci. Public Interest* 20, 1–68. doi: 10.1177/1529100619832930

Barrett, L. F., Mesquita, B., and Gendron, M. (2011). Context in emotion perception. *Curr. Dir. Psychol. Sci.* 20, 286–290. doi: 10.1177/0963721411422522

Bebko, J. M., and McKinnon, E. E. (1993). *The Language Proficiency Profile-2 (Unpublished assessment tool)*. Toronto: York University.

Dawel, A., Palermo, R., O'Kearney, R., and McKone, E. (2015). Children can discriminate the authenticity of happy but not sad or fearful facial expressions, and use an immature intensity-only strategy. *Front. Psychol.* 6:462. doi: 10.3389/fpsyg.2015.00462

de Gelder, B., Meeren, H. K., Righart, R., Van den Stock, J., Van de Riet, W. A., and Tamietto, M. (2006). Beyond the face: exploring rapid influences of context on face processing. *Prog. Brain Res.* 155, 37–48. doi: 10.1016/S0079-6123(06)55003-4

Denham, S. A. (1998). *Emotional Development in Young Children*. New York, NY: Guilford Press.

Felleman, E. S., Carlson, C. R., Barden, R. C., Rosenberg, L., and Masters, J. C. (1983). Children's and adults' recognition of spontaneous and posed emotional expressions in young children. *Dev. Psychol.* 19, 405–413. doi: 10.1037/0012-1649.19.3.405

Gnepp, J. (1983). Children's social sensitivity: inferring emotions from conflicting cues. *Dev. Psychol.* 19, 805–814. doi: 10.1037/0012-1649.19.6.805

Harris, P. L., Donnelly, K., Guz, G. R., and Pitt-Watson, R. (1986). Children's understanding of the distinction between real and apparent emotion. *Child Dev.* 57, 895–909. doi: 10.2307/1130366

Keltner, D., Tracy, J. L., Sauter, D., and Cowen, A. (2019). What basic emotion theory really says for the twenty-first century study of emotion. *J. Nonverbal Behav.* 43, 195–201. doi: 10.1007/s10919-019-00298-y

Kromm, H., Färber, M., and Holodynski, M. (2015). Felt or false smiles? volitional regulation of emotional expression in 4-, 6-, and 8-year-old children. *Child Dev.* 86, 579–597. doi: 10.1111/cdev.12315

Lawrence, K., Campbell, R., and Skuse, D. (2015). Age, gender, and puberty influence the development of facial emotion recognition. *Front. Psychol.* 6:761. doi: 10.3389/fpsyg.2015.00761

Maassarani, R., Gosselin, P., Montembeault, P., and Gagnon, M. (2014). French-speaking children's freely produced labels for facial expressions. *Front. Psychol.* 5:555. doi: 10.3389/fpsyg.2014.00555

Misailidi, P. (2006). Young children's display rule knowledge: understanding the distinction between apparent and real emotions and the motives underlying the use of display rules. *Soc. Behav. Pers.* 34, 1285–1296. doi: 10.2224/sbp.2006.34.10.1285

Mizokawa, A. (2011). Young children's understanding of pretend crying: the effect of context. *Br. J. Dev. Psychol.* 29, 489–503. doi: 10.1348/026151010X519964

Morgan, J. K., Izard, C. E., and King, K. A. (2010). Construct validity of the emotion matching task: preliminary evidence for convergent and criterion validity of a new emotion knowledge measure for young children. *Soc. Dev.* 19, 52–70. doi: 10.1111/j.1467-9507.2008.00529.x

Nelson, N. L., Hudspeth, K., and Russell, J. A. (2013). A story superiority effect for disgust, fear, embarrassment, and pride. *Br. J. Dev. Psychol.* 31, 334–348. doi: 10.1111/bjdp.12011

Nelson, N. L., and Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *J. Exp. Child Psychol.* 110, 52–61. doi: 10.1016/j.jecp.2011.03014

Park, Y. H., and Itakura, S. (2019). Causal information over facial expression: modulation of facial expression processing by congruency and causal factor of the linguistic cues in 5-Year-Old Japanese children. *J. Psycholinguist. Res.* 48, 987–1004. doi: 10.1007/s10936-019-09643-0

Pons, F., Harris, P. L., and de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: developmental periods and hierarchical organization. *Eur. J. Dev. Psychol.* 1, 127–152. doi: 10.1080/17405620344000022

Reichenbach, L., and Masters, J. C. (1983). Children's use of expressive and contextual cues in judgments of emotion. *Child Dev.* 54, 993–1004. doi: 10.2307/1129903

Repacholi, B. M. (1998). Infants' use of attentional cues to identify the referent of another person's emotional expression. *Dev. Psychol.* 34, 1017–1025. doi: 10.1037/0012-1649.34.5.1017

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

Ruba, A. L., Meltzoff, A. N., and Repacholi, B. M. (2019). How do you feel? preverbal infants match negative emotions to events. *Dev. Psychol.* 55, 1138–1149. doi: 10.1037/dev0000711

Saarni, C. (1999). *The Development of Emotional Competence*. New York, NY: Guilford Press.

Saarni, C., Campos, J. J., Camras, L. A., and Witherington, D. (2007). "Emotional development: action, communication, and understanding," in *Handbook of Child Psychology: Social, Emotional, and Personality Development*, W. Eisenberg Samon and R. M. Lerner (eds.) (Hoboken, NJ: John Wiley & Sons Inc), 226–299.

Scharfe, E. (2000). "Development of emotional expression, understanding, and regulation in infants and young children," in R. Bar-On and J. D. A. Parker (eds.) *The Handbook of Emotional Intelligence: Theory, Development, Assessment, and Application at Home, School, and in the Workplace* (San Francisco, Ca: Jossey-Bass Inc), 244–262.

Sidera, F. (2009). *La Comprensió Infantil de la Distinció entre L'emoció Externa i l'emoció interna en Situacions D'engany i de joc de ficció*. Ph.D. disseration, Universitat de Girona, Girona.

Sidera, F., Amadó, A., and Martínez, L. (2017). Influences on facial emotion recognition in deaf children. *J. Deaf Stud. Deaf Educ.* 22, 164–177. doi: 10.1093/deafed/enw072

Sidera, F., Amadó, A., and Serrat, E. (2013). Are you really happy? children's understanding of real vs. pretend emotions. *Curr. Psychol.* 32, 18–31. doi: 10.1007/s12144-012-9159-9

Sidera, F., Serrat, E., Rostan, C., and Serrano, J. (2012). Children's attribution of beliefs about simulated emotions. *Stud. Psychol.* 54, 67–80.

Sidera, F., Serrat, E., Rostan, C., and Sanz-Torrent, M. (2011). Do children realize that pretend emotions might be unreal? *J. Genet. Psychol.* 172, 40–55. doi: 10.1080/00221325.2010.504761

Song, R., Over, H., and Carpenter, M. (2016). Young children discriminate genuine from fake smiles and expect people displaying genuine smiles to be more prosocial. *Evol. Hum. Behav.* 37, 490–501. doi: 10.1016/j.evolhumbehav.2016.05.002

Sorce, J. F., Emde, R. N., Campos, J. J., and Klinnert, M. D. (1985). Maternal emotional signaling: its effect on the visual cliff behavior of 1-year-olds. *Dev. Psychol.* 21, 195–200. doi: 10.1037/0012-1649.21.1.195

Székely, E., Tiemeier, H., Arends, L. R., Jaddoe, V. W., Hofman, A., Verhulst, F. C., et al. (2011). Recognition of facial expressions of emotions by 3-years-olds. *Emotion* 11, 425–435. doi: 10.1037/a0022587

Theurel, A., Witt, A., Malsert, J., Lejeune, F., Fiorentini, C., Barisnikov, K., et al. (2016). The integration of visual context information in facial emotion recognition in 5-to 15-year-olds. *J. Exp. Child Psychol.* 150, 252–271. doi: 10.1016/j.jecp.2016.06.004

Walle, E. A., and Campos, J. J. (2014). The development of infant detection of inauthentic emotion. *Emotion* 14, 488–503. doi: 10.1037/a0035305

Widen, S. C. (2013). Children's interpretation of facial expressions: the long path from valence-based to specific discrete categories. *Emot. Rev.* 5, 72–77. doi: 10.1177/1754073912451492

Widen, S. C., Pochedly, J. T., and Russell, J. A. (2015). The development of emotion concepts: a story superiority effect in older children and adolescents. *J. Exp. Child Psychol.* 131, 186–192. doi: 10.1016/j.jecp.2014.10.009

Widen S. C., and Russell J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Dev. Psychol.* 39, 114–128. doi: 10.1037/0012-1649.39.1.114

Widen, S. C., and Russell, J. A. (2008). "Young children's understanding of others' emotions," in *(Eds.), Handbook of emotions*, M. Lewis, J. M. Haviland-jones, and L. F. Barrett (eds.) (New York, NY: Guilford Press), 348–363.

Widen, S. C., and Russell, J. A. (2010). Differentiation in preschooler's categories of emotion. *Emotion* 10, 651–661. doi: 10.1037/a0019005

Zeman, J., and Garber, J. (1996). Display rules for anger, sadness, and pain: it depends on who is watching. *Child Dev.* 67, 957–973. doi: 10.2307/1131873

Conflict of Interest: 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 Computational Neuroscience

ORIGINAL RESEARCH
published: 05 January 2021
doi: 10.3389/fncom.2020.587702

Performance Analysis With Different Types of Visual Stimuli in a BCI-Based Speller Under an RSVP Paradigm

Ricardo Ron-Angevin[1]*, *M. Teresa Medina-Juliá*[1], *Álvaro Fernández-Rodríguez*[1], *Francisco Velasco-Álvarez*[1], *Jean-Marc Andre*[2], *Veronique Lespinet-Najib*[2] and *Liliana Garcia*[2]

[1] UMA-BCI Group, Departamento de Tecnología Electrónica, Universidad de Málaga, Malaga, Spain, [2] Laboratoire IMS, CNRS UMR5218, Cognitique Team, Bordeaux INP-ENSC, Talence, France

Brain-Computer Interface (BCI) systems enable an alternative communication channel for severely-motor disabled patients to interact with their environment using no muscular movements. In recent years, the importance of research into non-gaze dependent brain-computer interface paradigms has been increasing, in contrast to the most frequently studied BCI-based speller paradigm (i.e., row-column presentation, RCP). Several visual modifications that have already been validated under the RCP paradigm for communication purposes have not been validated under the most extended non-gaze dependent rapid serial visual presentation (RSVP) paradigm. Thus, in the present study, three different sets of stimuli were assessed under RSVP, with the following communication features: white letters (WL), famous faces (FF), neutral pictures (NP). Eleven healthy subjects participated in this experiment, in which the subjects had to go through a calibration phase, an online phase and, finally, a subjective questionnaire completion phase. The results showed that the FF and NP stimuli promoted better performance in the calibration and online phases, being slightly better in the FF paradigm. Regarding the subjective questionnaires, again both FF and NP were preferred by the participants in contrast to the WL stimuli, but this time the NP stimuli scored slightly higher. These findings suggest that the use of FF and NP for RSVP-based spellers could be beneficial to increase information transfer rate in comparison to the most frequently used letter-based stimuli and could represent a promising communication system for individuals with altered ocular-motor function.

Keywords: brain computer-interface (BCI), rapid serial visual presentation (RSVP), electroencephalography (EEG), P300, N170, famous faces, neutral pictures

Edited by:
Jian K. Liu,
University of Leicester,
United Kingdom

Reviewed by:
Jing Jin,
East China University of Science and Technology, China
Zulay Lugo,
Institut Pere Mata, Spain

***Correspondence:**
Ricardo Ron-Angevin
rron@uma.es

***Received:** 27 July 2020*
***Accepted:** 23 November 2020*
***Published:** 05 January 2021*

Citation:
Ron-Angevin R, Medina-Juliá MT, Fernández-Rodríguez Á, Velasco-Álvarez F, Andre J-M, Lespinet-Najib V and Garcia L (2021) Performance Analysis With Different Types of Visual Stimuli in a BCI-Based Speller Under an RSVP Paradigm. Front. Comput. Neurosci. 14:587702. doi: 10.3389/fncom.2020.587702

INTRODUCTION

Brain computer interfaces (BCI) was first described by Vidal (1973) as a man-computer dialogue using observable and controllable neuroelectric events. That is, BCIs are a type of system that allow users to interact with their environment, using no muscular movements but only their brain activity (Nicolas-Alonso and Gomez-Gil, 2012). Therefore, these systems serve as a last communication channel between severely motor-disabled patients, such as amyotrophic lateral sclerosis (ALS) patients or those with brainstem injuries in a locked-in state (LIS), and their environment.

The most frequently control signal BCI systems use is the brain bioelectricity recorded through electroencephalography (EEG) (Nicolas-Alonso and Gomez-Gil, 2012; Rezeika et al., 2018). The EEG data is processed, and different brain components could be studied depending on the stimulus type and system that is desired to be controlled. The most typical components used in BCI are steady-state visual evoked potentials (SSVEP), event-related potentials (ERP) and sensorimotor rhythms (SMR) (Nicolas-Alonso and Gomez-Gil, 2012). The present study will focus on ERP components, which are evoked after the appearance of an infrequent stimulus. The most studied component of this type is the P300 component, which was first discovered by Sutton et al. (1965) and described as a positive amplitude waveform alteration that reaches peak amplitude at about 300 ms after a sensory stimulus. This potential is mostly recorded in the parietal area (Polich, 2007).

This P300 component is usually employed as a control signal for a type of BCI system which is called a virtual speller (Rezeika et al., 2018). The first P300 based BCI speller was proposed by Farwell and Donchin (1988). This speller consisted of a 6 × 6 matrix table of letters and numbers, whose rows and columns were highlighted (i.e., the characters color turned from gray to white) pseudorandomly in order to evoke the P300 component each time the target character was highlighted. As a consequence, this BCI speller presentation paradigm is called row-column paradigm (RCP). On the other hand, to consistently elicit and classify the P300 component, users are often asked to focus their attention on their desired target letter and count the number of times it flashes, and a classification algorithm differentiates the target letter between many non-targets. Other temporal components generated earlier or later to P300 (P100, N170, N250, N400) are equally analyzed to detect stimuli features (Zheng et al., 2012; Jiang et al., 2017; Tian et al., 2018).

Different variations on the highlighting type and nature of the characters have been studied, such as the shape, color and size of the characters (Salvaris and Sepulveda, 2009; Ryan et al., 2017; Fernández-Rodríguez et al., 2019b) in order to improve system performance (classification accuracy, information transfer rate or ERP amplitude). Regarding the nature of stimuli, it has been demonstrated that the presentation of famous faces (FF) instead of letters leads to an improvement in performance (Kaufmann et al., 2011; Li et al., 2015). Other set of images, such as neutral images, might also help to increase performance as compared to letters (Fernández-Rodríguez et al., 2019a). Moreover, the study of Kellicut-Jones and Sellers (2018) suggests that the FF paradigm might not be significantly better than neutral images in RCP. On the other hand, in the single character presentation (SCP) paradigm –which consists of illuminating the matrix stimuli one by one– the use of faces (non-famous) seemed to increase performance as compared to neutral images (inanimate objects) (Zhao et al., 2011). Nevertheless, these study results should be carefully considered as they are derived from a small sample size. Even though this study is not completely adequate, these findings might suggest that a difference in performance might exist depending on the stimulus presentation paradigm used, in particular when applying FF and neutral images.

The stimulus presentation paradigms RCP and SCP present their stimuli in different locations of the monitor screen, but RCP presents them by row and column groups, and SCP, individually. However, this type of presentation paradigm might not be the most suitable for some patients with motor disabilities who also have no or residual ocular mobility, as the performance of these paradigms is greatly decreased under covert attention conditions (Brunner et al., 2010; Treder and Blankertz, 2010). Different type of visual gaze-independent BCIs have been researched by the literature in order to prevent this limitation. According to the BCI-Spellers review by Rezeika et al. (2018), two groups of main gaze-independent spellers have been proposed by previous literature: (i) those that display the stimuli to be selected in different close positions to control the speller under covert attention, such as Chroma Speller (Acqualagna et al., 2013), Geospell (Aloise et al., 2012), Gaze-Independent Block Speller (GIBS) (Pires et al., 2011) and Hex-O-Spell (Treder and Blankertz, 2010); or (ii) those based on rapid serial visual presentation (RSVP), which sequentially presents stimuli in the center of the screen (Acqualagna and Blankertz, 2013). The authors of this review stated that the RSVP-based BCIs show promising results and have been the most widely used to date.

Different visual configurations of the stimuli under RSVP had also been studied in the literature to increase the system performance for different applications like face recognition or RSVP spellers (Lees et al., 2018). In a recent study, Chen et al. (2016) tested if the characteristics of the stimuli can affect the performance of the system using colored balls, gray dummy faces and colored dummy faces. For each paradigm, six different stimuli were presented (six colors and six dummy face expressions). They found that the combination of colors and dummy face expressions could improve the bit rate. Regarding RSVP spellers, a previous study found a trend in which using colors and different capitalizations might improve the accuracy and bit rate compared to black letters (Acqualagna and Blankertz, 2013). Furthermore, the study of Won et al. (2018) proposed a RSVP speller whose colored stimuli were placed in six different near central positions. They found that using different locations for the letters increased the accuracy of the system in contrast to the classical RSVP paradigm.

Nevertheless, studies regarding the nature of the stimuli under RSVP have barely been carried out for communication purposes (i.e., RSVP spellers). In a preliminary study, neutral images and letters were compared in RSVP (Fernández-Rodríguez et al., 2019c). The results of this work showed that neutral images did not offer significant benefits as compared to letters under the RSVP paradigm. In the same way, to our knowledge, any studies regarding RSVP have compared FF to letters and it would be interesting to determine the efficacy of FF under a RSVP paradigm. However, the results of Fernández-Rodríguez et al. (2019c) should also be carefully considered as a small sample size was applied and no metrics regarding the user experience were considered (such as fatigue, preference and control). To better understand the effect of this sort of stimuli when applying an RSVP paradigm, an extended and complete study regarding neutral images and FF against letters should be carried out.

We hypothesized that using alternative stimuli under RSVP –i.e., famous faces and neutral pictures– instead of letters would increase system performance and user experience of the RSVP-based spellers, as previously demonstrated in the RCP and SCP presentation paradigms. Therefore, the aim of this study was to compare and evaluate the performance of three different types of stimuli (letters, famous faces and neutral pictures) as feasible communication stimuli for a gaze-independent BCI speller. The evaluation was carried out in terms of objective parameters (specifically, accuracy, information transfer rate and brain waveform analysis) and a subjective questionnaire regarding the perception of the participants. The main contribution of this study would be to experimentally (in)validate the usability of alternative stimuli under the RSVP paradigm for communication purposes.

METHODS

Participants

Eleven French participants (aged 19.91 ± 0.83) took part in the present study. None of the participants had previous experience in the use of BCI systems. The study was approved by the Ethics Committee of the University of Malaga and met the ethical standards of the Helsinki Declaration. According to self-reports, none of the participants had any history of neurological or psychiatric illness. In addition, all of them provided written consent trough a protocol reviewed by the ENSC-IMS (Ecole Nationale Supérieur de Cognitique – Intégration du Matériau su Système) Cognitive and UMA-BCI teams.

Data Acquisition and Signal Processing

The EEG was recorded using the electrode positions: Fz, Cz, Pz, Oz, P3, P4, PO7, and PO8, according to the 10/20 international system. All channels were referenced to the right earlobe, using FPz as the ground.

The EEG was amplified through a 16 channel biosignal amplifier gUSBamp (Guger Technologies). The amplifier settings were from 0.5 to 100 Hz for the band-pass filter, the notch (50 Hz) was on, and the sensitivity was 500 μV. The signal was then digitized at a rate of 256 Hz. EEG data collection and processing were controlled by the *UMA-BCI Speller* software (Velasco-Álvarez et al., 2019), which serves as the front-end to BCI2000 (Schalk et al., 2004). Likewise, when the brain signal was recorded by the *UMA-BCI Speller*, a pass-band filter from 0.1 to 60 Hz was applied, and the notch filter was on at 60 Hz.

A stepwise linear discriminant analysis (SWLDA) of the data was performed to obtain the weights for the P300 classifier and calculate the accuracy. Alternative classification methods of the EEG signal have been proposed by the literature (Lotte et al., 2018; Xiao et al., 2020), however the SWLDA algorithm has been widely used and validated (Krusienski et al., 2008; Lees et al., 2018). Furthermore, this last is the algorithm that BCI2000 software, and thus the UMA-BCI Speller, has implemented.

According to the specifications described in the Wiki page of BCI2000[1], the EEG channels used and their respective weights in the classification matrix are dependent of specific parameters of the user. The different ERP components are commonly found in certain brain zones and certain latencies; but when analyzed particularly for each user, the specific channels and latencies may be different from one another (Luck, 2014). These weights are calculated in the calibration task. The time frame considered to train the classifier was from 0 to 800 ms after the onset of a stimulus (target or non-target). Note that the selection of the channels and calculation of the classification weights were automatically done by the classifier that the BCI2000 software has implemented.

Spelling Paradigms

Three different RSVP paradigms were evaluated in the present work. The only difference between paradigms was the type of stimulus used: (i) white letters (WL), (ii) famous faces (FF), and (iii) neutral pictures (NP) (**Figure 1**).

Each paradigm presented nine different stimuli (**Table 1**). In the WL paradigm, the letters used were A, B, C, E, L, M, O, R, and S. On the other hand, each character in the FF stimuli was chosen so that the character's name or surname had to start with the same letter as the one used in the WL paradigm (e.g., W. Allen for the letter A, or Beyoncé for the letter B). Finally, for the NP stimuli, the criterion was the same: the picture had to start, in French, with the same letter as the one used in the WL paradigm (e.g., the picture of a tree –*arbre*, in French– for the letter A, or a boat –*bateau*, in French– for the letter B). The relationship between each stimulus and image (face or picture) was explicitly declared by the research staff to participants in order to avoid any mistake. See **Table 1** for the letters and their corresponding image names (face and picture). The images used in the experiment are not shown in this paper due to copyright reasons.

The number of elements was selected in order to avoid a target selection time that was too long, as the aim of this study was to validate the different sets of stimuli under RSVP for communication purposes. In previous studies with this kind of paradigm, an even smaller number of elements has been used to validate hypotheses (Chen et al., 2016; Fernández-Rodríguez et al., 2019c).

The duration of each stimulus presentation was equal to 187.5 ms and the inter stimuli interval (ISI) was equal to 93.75 ms. Therefore, the stimulus onset asynchrony (SOA) had a duration equal to 281.25 ms. The time for completing a sequence (i.e., single presentation or flashing of every stimuli) was 2.44 s. The pause time between one selection and the start of the next (i.e., between completed sets of sequences) was equal to 5 s.

The flashing stimuli were presented in the center of the screen. The dimensions regarding the type of stimuli were as follows: letters, around 3 × 4 cm; faces, around 6 × 8.5 cm; and Pictures, around 12 × 8.5 cm.

[1] https://www.bci2000.org/mediawiki/index.php/Main_Page

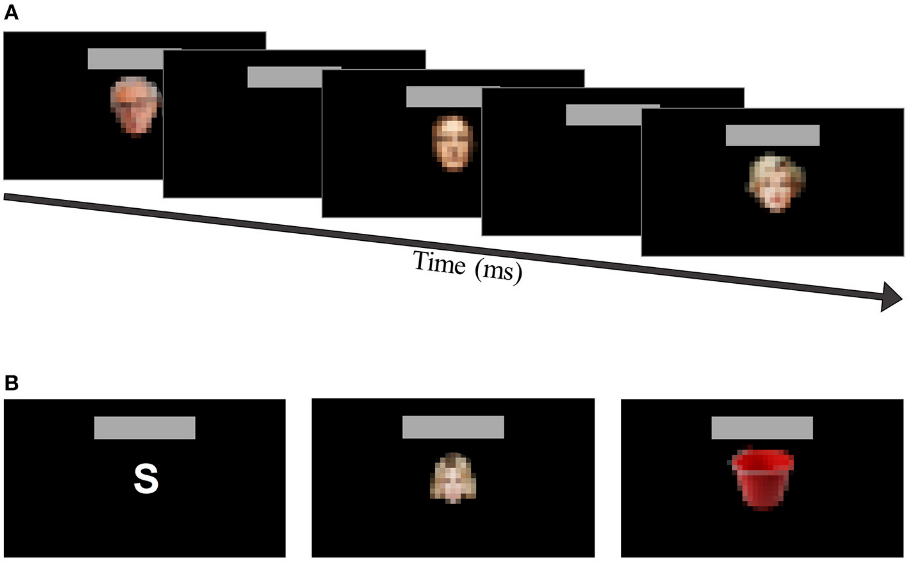


FIGURE 1 | (A) RSVP paradigm over time with the Famous Faces (FF) interface as an example; **(B)** Example of a stimulus representation with its equivalent and corresponding white letter (WL, "S"), famous face (FF, "Shakira") and neutral picture (NP, "*Seau*"). Note that due to copyright reasons, the images presented are pixelated in this figure.

TABLE 1 | Presentation of the nine stimuli names contained in each set of stimuli.

Stimuli set		
WL	**FF**	**NP**
A	W. Allen	*Arbre* (tree)
B	Beyoncé	*Bateau* (boat)
C	H. Clinton	*Cloche* (bell)
E	A. Einstein	*Eau* (water)
L	J. Lennon	*Lit* (bed)
M	M. Monroe	*Main* (hand)
O	B. Obama	*Ours* (bear)
R	D. Radcliffe	*Roue* (wheel)
S	Shakira	*Seau* (bucket)

WL, white letter; FF, famous face; NP, neutral image. Each row contains the letter, famous name, and picture names that are related. The English translation of each word contained in the NP column is presented inside brackets.

Procedure

A within-subject design was used, so that all users went through all experimental conditions. The experiment was carried out in one session. The order of the paradigms was counterbalanced across participants in order to prevent any undesired effects, such as learning or fatigue. Each condition consisted of two parts: (i) an initial calibration task to obtain the specific signal patterns associated with each user and (ii) an online task in which the user actually controlled the interface. Therefore, the main difference between both tasks was that in the first task the user did not receive any feedback.

For both phases, the task was to write different four-letter words. In the case of the calibration phase, the participant had to write two French words ("MARE," pond in English, and "CLOS," enclosed plot in English), so the total number of selections for this task was 8 letters. On the other hand, for the online phase, the user had to write three French words ("MALE," male in English, "ROSE," pink in English, and "BOLS," bowl in English), so the number of selections would be 12 letters. Participants were told during the pause between selections which image (famous face or neutral picture) or letter they have to focus on in the next run. They were not asked to memorize the sets of stimuli used in the experiment (letter, face and picture related) as the purpose of this study was to test the effect of this type of stimuli in a preliminary RSVP-based speller. A short break between words (variable at the request of the user) was employed. The number of sequences (i.e., the number of times that each stimulus –target and non-target– was presented) was pre-fixed to 6 in the calibration task and adapted in the online phase depending on the user performance in the calibration phase. The number of sequences selected for the online task was two trials more than the minimum number of trials required to obtain 100% accuracy in the calibration phase.

At the end of the session, the user had to complete a questionnaire regarding his/her experience during the control of the paradigm.

Evaluation

Four parameters were used to evaluate the effect of the RSVP paradigm and stimulus type on the performance: (i) the *accuracy* in the calibration and online phases, (ii) the *information transfer rate* (*ITR*) (Wolpaw et al., 1998) in the calibration and online phases, (iii) the analysis of the event-related waveform during the calibration phase, and (iv) a subjective questionnaire.

Accuracy was defined as the number of correctly predicted selections divided by the total number of predicted selections, multiplied by 100. While for the online task this last definition was applied, for the calibration phase, the accuracy percentage was computed by the signal classifier after the classification of the word using the data from each sequence. The SWLDA classification algorithm applied was the one proposed by BCI2000.

The *ITR* (bits/min) is an objective measure to determine the communication speed of the system. This parameter considers accuracy, the number of elements available in the interface and time to select one element:

$$ITR = \frac{log_2\, N + P\, log_2\, P + \,(1\, -\, P)\, log_2\, \frac{1-P}{N-1}}{T}$$

where P is the accuracy of the system, N is the number of elements available at the interface and T is the time needed to complete a trial (i.e., select an element).

The *ITR* was calculated similarly to the accuracy for both the calibration and the online tasks. It should be noted that the pause between selections was not considered when calculating the *ITR*.

The grand average of the ERP waveforms (from 0 to 800 ms) was analyzed in order to evaluate how the three different stimuli types affected the waveforms of the target, non-target and amplitude difference between target and non-target stimuli. In addition, to carry out a more exhaustive analysis concerning the ERP components frequently used in a BCI, a grand average topography was also carried out for target, non-target and amplitude difference between target and non-target stimuli. Next components were included in the topographical analyses: P100 (60–110 ms), N170 (110–180 ms), P300 (450–520 ms), and N400 (520–570 ms). These topographical maps were statistically compared between conditions. The interval time for each component were chosen according to previous literature and the specific EEG signal obtained in the present study (e.g., Tanaka, 2018; Mijani et al., 2019), except for the N400 component which was selected only according the EEG signal obtained. This last issue is discussed in the Discussion section.

To perform these analyses (i.e., comparison between conditions regarding the grand average of ERP waveforms and the grand average of topography), a baseline from −200 to 0 ms was used for the electrodes, and this was low-pass filtered at 30 Hz. Statistical analyses were carried out using EEGLAB (Delorme and Makeig, 2004), with which a false discovery rate (FDR) correction was applied.

Finally, a subjective questionnaire –specially configured for this experiment– was applied to investigate the experience of the users during the control of the spellers. This questionnaire required that the users scored the different conditions from 0 to 10 using a visual analog scale (VAS) according to the following dimensions: level of fatigue (*fatigue*), complexity of use (*complex*), level of speed felt during presentation of the stimuli (*speed*) and level of stress (*stress*). Where 0 is the lowest value and 10 is the highest for the *fatigue*, *complex* and *stress* dimensions. For the case of the *speed* dimension, 0 would mean that the interface had an adequate speed of stimuli presentation, and 10 would mean that the speed of the stimuli presentation was too fast.

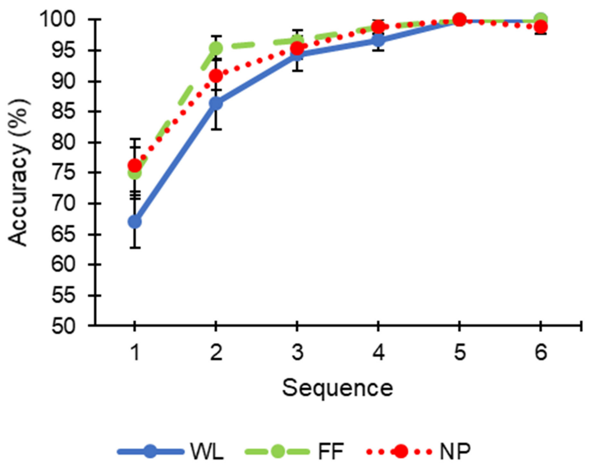


FIGURE 2 | Accuracy (mean ± standard error) of each condition (WL, white letters; FF, familiar faces; NP, neutral pictures) as a function of the number of sequences during the calibration task.

RESULTS

In this section, the different results are presented in different sections. First, the results of the calibration task (i.e., performance metrics and ERP waveforms) are presented, followed by the performance metrics of the online phase, and finally, the subjective questionnaire analysis.

Calibration Task

Performance Metrics

In order to find out if there were any significant differences between the different conditions, a Student's t-test was performed for repeated samples for each of the sequences. The *accuracy* (**Figure 2**) did not show significant differences for any sequence. However, the variable *ITR* (**Figure 3**) showed significant differences for the first sequence between conditions WL and NP [$t_{(10)} = 2.24$; $p = 0.049$] (**Supplementary Table 1**). Likewise, some marginally significant differences were revealed when the average *accuracy* and *ITR* of all sequences were calculated (WL, 90.74 ± 5.44% and 20.95 ± 3.88 bits/min; FF, 94.32 ± 3.5% and 23.75 ± 3.34 bits/min; NP, 93.39 ± 4.36% and 23.23 ± 4 bits/min). Specifically, WL was observed to offer a marginally significant worst performance than FF [*accuracy*, $t_{(10)} = 2.161$; $p = 0.056$; *ITR*, $t_{(10)} = 2.175$; $p = 0.055$] and NP [*ITR*, $t_{(10)} = 1.89$; $p = 0.088$].

ERP Waveform

Regarding the grand average ERP waveform, the statistical analyses showed significant differences between conditions at an early time interval (around 80–110 ms) for target stimuli in Cz and PO7 (**Figure 4**).

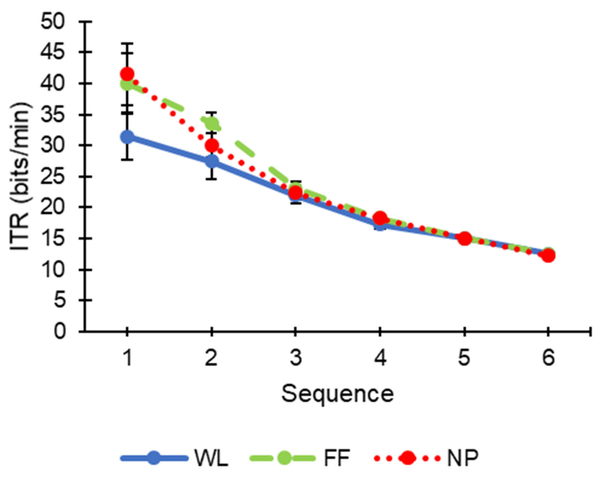


FIGURE 3 | Information transfer rate (ITR, mean ± standard error) of each condition (WL, white letters; FF, familiar faces; NP, neutral pictures) as a function of the number of sequences during the calibration task.

On the other hand, regarding the grand average topography of the P100, N170, P300, and N400 components, only the P100 (60–110 ms) component showed significant differences in channels Cz, Pz, Oz, P3, P4, PO7, and PO8 for the target stimuli (**Figure 5**). These differences could indicate a difference in early processing depending on the type of visual stimulus. Specifically, it appeared that the FF condition obtained lower grand average ERP amplitude values than those obtained by the WL and NP conditions (**Figure 4**).

Online Task

The *accuracy* and *ITR* results achieved, as well as the number of sequences used by each participant in the online task, are shown in **Table 2**. In regard to the *accuracy* obtained for the online task (second main column of **Table 2**), the Student's *t*-test between conditions showed no significant differences between NP and the rest of conditions [NP vs. WL, $t_{(10)} = 0.183$; $p = 0.859$; NP vs. FF, $t_{(10)} = 0.957$; $p = 0.361$]. However, a comparison between WL and FF conditions showed a trend close to significance [$t_{(10)} = 1.961$; $p = 0.078$]. On the other hand, for the *ITR* (third main column of **Table 2**), significant differences were found between the WL and FF conditions [$t_{(10)} = 2.973$; $p = 0.014$], but not between WL and NP [$t_{(10)} = 0.595$; $p = 0.565$] nor NP and FF [$t_{(10)} = 1.261$; $p = 0.236$].

Subjective Questionnaires

With reference to the results obtained in the questionnaire (**Figure 6**), the condition NP, compared to WL, was found to be associated with significantly less *fatigue* [$t_{(10)} = 2.262$; $p = 0.047$] and more appropriate speed presentation –*speed*– [$t_{(10)} = 3.13$; $p = 0.011$]. Note that the comparison of FF and NP was near nominal significance for *speed* [$t_{(10)} = 2.085$; $p = 0.064$]. All statistical comparisons made between conditions for the subjective questionnaires can be observed in **Supplementary Table 2**.

DISCUSSION

In this study we tested different kinds of stimuli –white letters (WL), famous faces (FF) and neutral picture (NP)– under a rapid serial visual presentation (RSVP) paradigm to analyse the system performance in terms of classification accuracy, information transfer rate (ITR), ERP waveform and user experience (*fatigue*, *complexity*, *speed,* and *stress* level). Main results showed that FF and NP might produce, respectively, better performance and better user experience compared to WL. These results suggest that the stimuli proposed (FF and NP) could enhance the system performance, and thus communication, of this type of gaze-independent BCI.

Calibration Task

Performance Metrics

Main results regarding *accuracy* showed that, in the first sequence, the NP condition had a significantly higher *ITR* in contrast to the WL condition. These results are especially interesting for those cases in which higher communication speed is preferred even though *accuracy* is partially decreased. In fact, the *accuracy* reached by the NP condition in the first sequence (76.18 ± 14.24%) was higher than 70%, which is the minimum *accuracy* recommended by Kübler et al. (2001), and normally used by the BCI community, to enable an efficient communication system. The FF condition achieved similar results in *accuracy* and *ITR* (75 ± 13.69% and 40.01 ± 16.1 bits/min) to the NP (76.18 ± 14.24% and 41.46 ± 16.62 bits/min) in the first sequence, but it was still slightly lower and, thus, did not reach statistical significance when compared with WL, neither for *accuracy* nor *ITR* (WL: 67.01 ± 13.99% and 31.45 ± 12.58 bits/min).

For the rest of sequences, it can be observed that the higher the number of sequences the more similar the results for the different conditions (**Figures 2**, **3**). Nevertheless, from the average *accuracies* and *ITR*s throughout the sequence, it was observed that the values of the FF and NP conditions showed marginally significant better performance than the WL condition (*accuracy*, $p = 0.056$; *ITR*, $p = 0.055$). Therefore, the tendency is toward the WL condition showing a worse performance than FF and NP.

ERP Waveform

Significant differences were obtained in the analysis of the grand average ERP waveform, particularly in early time intervals in channels Cz and PO7 for target stimuli between conditions (**Figure 4**). These significances were corroborated in the topographical analyses (**Figure 5**). The component P100 (60–110 ms) offered significant differences between conditions. Thus, it can be affirmed that there are differences in early neural processing depending on the type of visual stimulus. Specifically, it appeared that the FF condition obtained lower grand average ERP amplitude values than those obtained by the WL and NP conditions. Furthermore, observing the grand

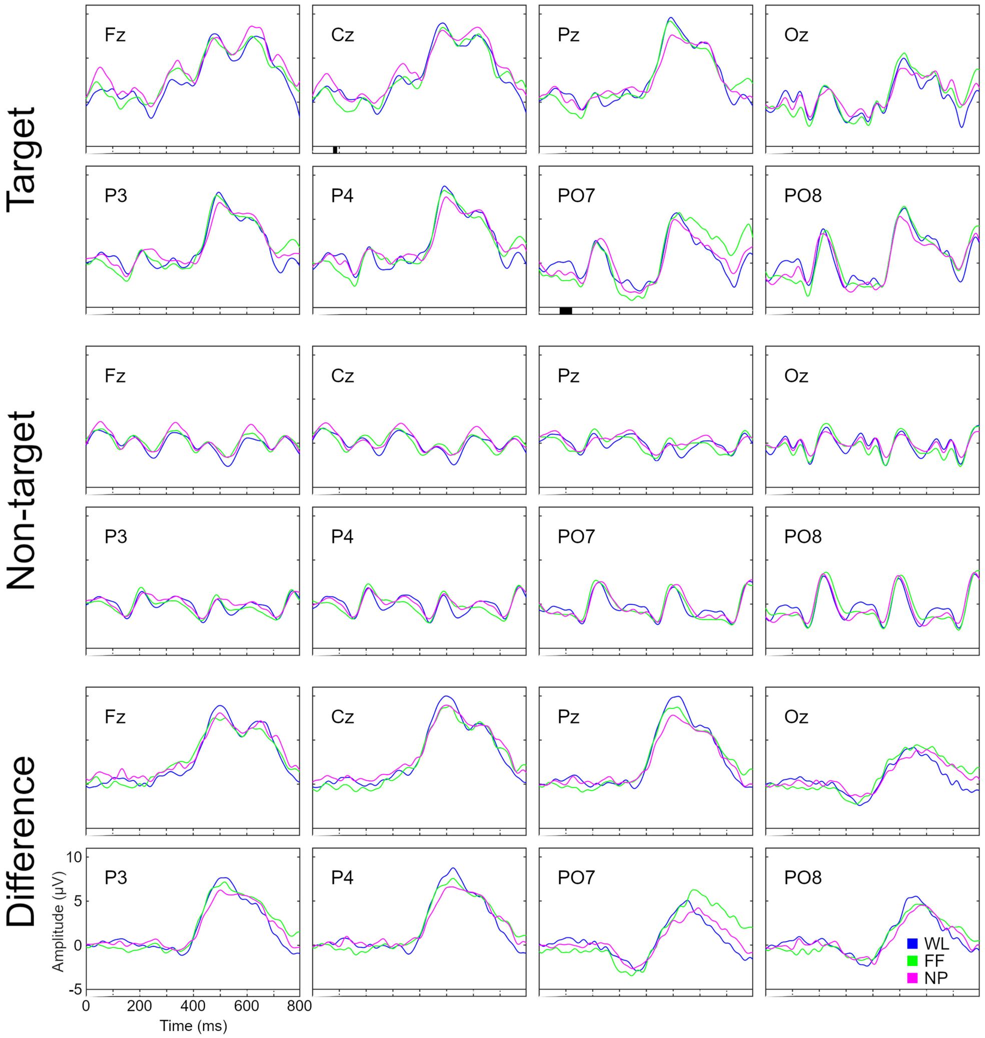


FIGURE 4 | Grand average ERP waveform for target, non-target and amplitude difference between target and non-target stimuli signals in all used channels (Fz, Cz, Pz, Oz, P3, P4, PO7, and PO8) for the three conditions: white letters (WL), familiar faces (FF), and neutral pictures (NP). These plots were obtained from the EEG data recorded during the calibration phase.

average ERP amplitude at the following milliseconds (**Figure 4**), a possible N170 component is observed in the three conditions in almost every channel. Nevertheless, this potential is especially pronounced for the FF condition –although not significant– in contrast to those obtained by the WL and NP conditions. These results would fit with previous BCI literature, as N170 is a potential related to facial recognition (Kaufmann et al., 2011; Kellicut-Jones and Sellers, 2018).

Regarding later potentials, the P300 potential was clearly the most distinctive and largest component in all the channels for the three conditions. On the other hand, the N400 –which is related to familiar face recognition– was not found as reported in previous studies (Dijkstra et al., 2020). We deduced that it might have different latency because of the stimuli presentation used, or that it might have been delayed or even partially canceled by the P300 component (which had large –but common– amplitude and latency). Most probably, the N400 potential was not found in the present study as the paradigm applied in this experiment did not use any type of semantic incongruity, which have been related by the literature with the increase in the N400 potential (Eimer, 2000).

The function of the classifier is to discriminate between target and non-target stimuli. The positive correlation between amplitude of ERP waveform and performance in a visual ERP-based BCI has been previously demonstrated (Mak et al., 2012). It could be considered that a larger difference between target and non-target stimuli for any of the studied ERP components could increase the classifier performance. Thus, the results obtained in the ERP waveforms (**Figures 4**, **5**) might correlate with what was obtained in the calibration phase regarding performance

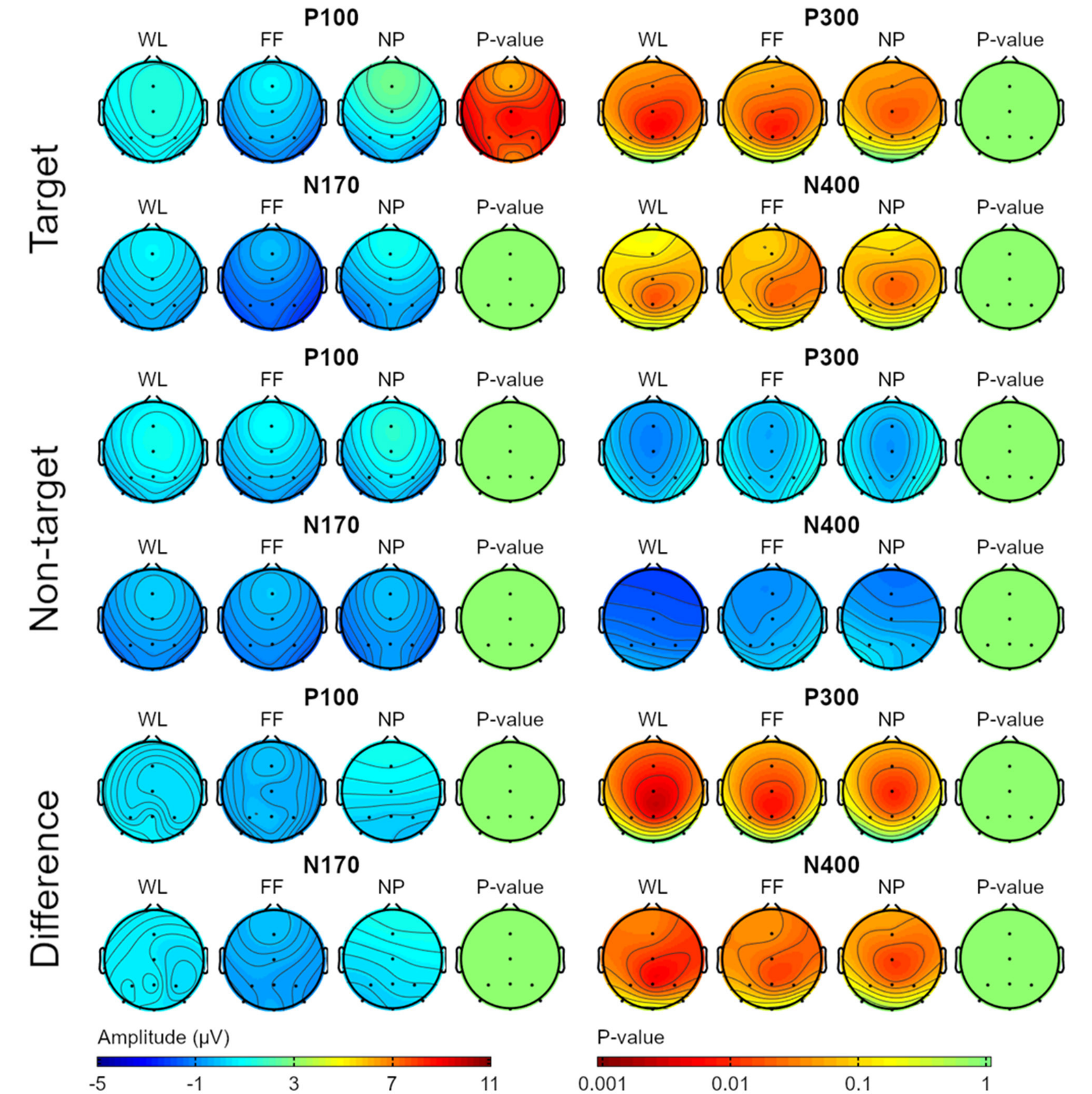


FIGURE 5 | Topographical scalp map of each condition (WL, white letters; FF, familiar faces; NP, neutral pictures) for the next components: P100 (60–110 ms), N170 (110–180 ms), P300 (450–520 ms), and N400 (520–570 ms). These plots were obtained from the EEG data recorded during the calibration phase.

(**Figures 2**, **3** and **Supplementary Table 1**). Specifically, the significant differences obtained in the first sequence of the ITR variable (**Figure 3**), between WL and NP, could be related to those found in the P100 component (**Figure 5**). Likewise, the higher performance of FF vs. WL in the first sequences of the calibration phase could be related to the grand average ERP amplitude of the N170 component presented for FF (**Figure 4**).

Online Task

For the online task, the FF condition achieved a significantly higher *ITR* as compared to the WL condition (17.2 ± 5.86 and 13.27 ± 5.13 bits/min, respectively), and a higher *accuracy,* which showed a trend close to significance ($p = 0.078$), was observed between these two conditions (WL, 85.24 ± 9.74%; FF, 90.53 ± 9.33%). On the other hand, the performance of the NP condition

TABLE 2 | Accuracy (%), ITR (bits/min) and number of sequences used (mean ± standard deviation) of the three conditions (WL, white letters; FF, familiar faces; NP, neutral pictures) during the online task.

User	Accuracy			ITR			Number of sequences		
	WL	FF	NP	WL	FF	NP	WL	FF	NP
U1	79.2	87.5	100	14.29	26.68	18.78	3	2	4
U2	83.3	100	66.7	9.57	15.03	5.94	5	5	5
U3	79.2	95.8	87.5	14.29	16.55	17.79	3	4	3
U4	95.8	100	95.8	22.06	25.04	33.10	3	3	2
U5	91.7	91.7	87.5	9.91	14.86	17.79	6	4	3
U6	87.5	83.3	95.8	8.89	15.95	13.24	6	3	5
U7	87.5	75	87.5	13.34	12.71	17.79	4	3	3
U8	91.7	100	83.4	19.82	25.04	7.99	3	3	6
U9	75	83.3	62.5	7.63	7.97	6.46	5	6	4
U10	100	100	87.5	18.78	15.03	10.67	4	5	5
U11	66.7	79.2	91.67	7.42	14.29	9.90	4	3	6
Mean	85.24 ± 9.74	90.53 ± 9.33	85.99 ± 11.67	13.27 ± 5.13	17.2 ± 5.86	14.5 ± 7.84	4.18 ± 1.17	3.73 ± 1.19	4.18 ± 1.33

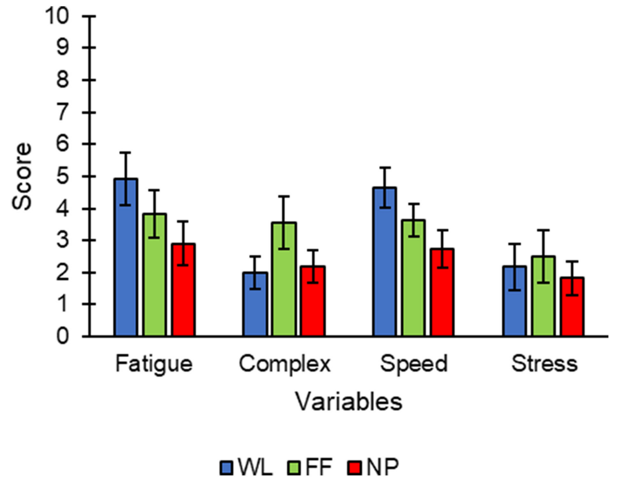


FIGURE 6 | Scores (mean ± standard error) of each condition (WL, white letters; FF, familiar faces; NP, neutral pictures) for the variables collected in the subjective questionnaire.

(85.99 ± 3.52%, 14.5 ± 2.36 bits/min) seemed to be placed in the middle and no significant differences were revealed as compared to the other two conditions (**Table 2**). Therefore, once again, the WL condition was found to be the least appropriate for the RSVP paradigm. These found observations go in the same direction as other authors suggest that the WL condition could be the less appropriate for the RCP paradigm than FF (Kaufmann et al., 2011; Kellicut-Jones and Sellers, 2018).

Comparing the obtained results with those of previous studies that also assessed RSVP spellers using only letters as stimuli, it can be observed that the reported *accuracy* and *ITR* values of this study are similar to those reported in the literature (Acqualagna and Blankertz, 2013; Chennu et al., 2013; Lin et al., 2018; Won et al., 2018; Fernández-Rodríguez et al., 2019c). To the best of our knowledge, the FF condition has not been used before for communication purposes under RSVP, and the NP condition has only been evaluated in a preliminary study (Fernández-Rodríguez et al., 2019c). Thus, the performance achieved by the FF or NP paradigms cannot be fairly compared to any other study, highlighting the novelty of this work.

The performance of the present study (especially the ITR) can be consider lower than the performance obtained by those studies that applied the RCP paradigm. This lower performance is essentially related to the time needed by the RSVP paradigm to present every stimuli in comparison to the one needed by RCP (Chennu et al., 2013). However, it should be noted that the RCP paradigm needs ocular mobility to be efficiently controlled what might limit its use for some patients (Brunner et al., 2010; Treder and Blankertz, 2010).

Subjective Questionnaire

Remarkably, the overall results of the subjective questionnaire were positive for the three interfaces, since all the average values were below 5 points (considering that the highest possible score was 10) for different subjective measurements (*fatigue*, *complex*, *speed,* or *stress*): WL, between 2 and 5 points (3.43 ± 2.56); FF, between 3 and 4 points (3.38 ± 2.35); and NP between 1 and 3 points (2.41 ± 1.87). Regarding the specific variables, the NP condition seemed to be the condition that gave the best results in terms of fatigue produced (*fatigue*) and interface speed adequacy (*speed*). In fact, NP offered a lower *fatigue* and *speed* vs. WL. Also, it is worth noting that the NP condition was near nominal significant to show better results than FF in the *speed* variable ($p = 0.064$). In addition to *fatigue* and *speed*, the NP condition showed the best results (i.e., the lowest values) for the *stress* parameter. On the other hand, the FF condition showed the highest scores for *complex* (although this was not significant). This last result is in contrast with observations at a global level for both the calibration and the online task, where FF generally presented better results.

Interestingly, the ERP waveforms (**Figures 4**, **5**) might correlate with what was declared by the participants regarding their subjective perspective of the paradigms (**Figure 6** and **Supplementary Table 2**). First, most probably we could not find more statistically differences in the ERP waveforms because the overall results of the subjective questionnaire were positive for the three spellers. Therefore, even though the NP condition obtained the best results for the *fatigue*, *speed* and *stress* parameters, these improvements might not highly affect the brain signal. Furthermore, the FF paradigm was declared as the most *complex* in a non-significant manner. This could be related to the non-significant differences obtained in the P300 potential, an ERP component previously associated with the complexity of the task (Käthner et al., 2014).

These results should be considered, especially in those cases where these applications want to be controlled during long sessions (either in the case of patients or healthy users), in which high levels of fatigue can diminish both user performance and satisfaction (Käthner et al., 2014).

Future Studies

Future studies might investigate more deeply why the effect of pictures or faces has not been as great as that observed for RCP in previous works (e.g., Kaufmann et al., 2011 and Fernández-Rodríguez et al., 2019a). Likewise, it would be interesting to further study whether the novel findings obtained under RCP in reference to the applied stimuli –for example, green famous faces, self-face paradigm or very small lateral stimuli (Li et al., 2015; Xu et al., 2018; Lu et al., 2020, respectively)– can be transferred to RSVP.

Furthermore, there are different BCI works in the literature that propose paradigms with reduced number of stimuli such as target selection in consecutive steps (Treder et al., 2011) or the T9 keyboard (Ron-Angevin et al., 2015). It would be interesting to test our proposed stimuli (face or picture) in this sort of reduced paradigms.

Finally, a further research to improve the system performance of the presented paradigms with images (face and pictures) would be also interesting. These improvements could be related to the type of classification algorithm used (Xiao et al., 2020), the creation of a generic model to decrease the calibration time (Jin et al., 2020), or even the application of hybrid systems which use different type of control signals (Xu et al., 2020).

CONCLUSION

The aim of this work was to assess the impact of three different types of stimuli under RSVP for communication purposes: WL, FF, and NP. In general terms, it seems that both the FF and NP conditions have a tendency to offer a better performance as compared to the WL condition, either for objective measurements (both for FF and NP in the calibration, and for FF in the online task) or for subjective measurements (in particular for NP).

Concerning any comparison between FF and NP, it is difficult to choose a recommended approach for potential users, because while the online task proved better for the FF condition, the NP condition achieved better scores in the subjective questionnaires. It is worth considering whether this performance improvement is more important than considering the subjective preference of the NP interface. It should be remembered that there were no significant differences between FF and NP throughout the study. Therefore, we estimate that the choice between the use of FF or NP will depend on the specific conditions and preferences of each user. However, it is clear that the WL condition should seldom be considered as the most suitable choice for a user.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Comité Ético de Experimentación de la Universidad de Málaga (CEUMA). CEUMA registry number: 51-2019-H. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RR-A, LG, MM-J, J-MA, and VL-N contributed to the conception and design of the study. ÁF-R, MM-J, and FV-Á performed the statistical analysis. MM-J and ÁF-R wrote the first draft of the manuscript. RR-A and LG were in charge of the supervision of the project. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was partially supported by the project SICCAU: RTI2018-100912-B-I00 (MCIU/AEI/FEDER, UE) and the University of Malaga (Universidad de Málaga). Furthermore, this project has received support from Bordeaux INP in the framework of the guest professor call obtained in 2018-2019.

ACKNOWLEDGMENTS

The authors would like to specially thank the college students Maurine Jouault, Maelle Abadie, Martin Filosa, Quentin Pestre-Sorge and Maelle Cloarec for their cooperation during the experiments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fncom.2020.587702/full#supplementary-material

REFERENCES

Acqualagna, L., Treder, M. S., and Blankertz, B. (2013). "Chroma speller: isotropic visual stimuli for truly gaze-independent spelling," in *International IEEE/EMBS Conference on Neural Engineering, NER* (San Diego, CA), 1041–1044.

Acqualagna, L., and Blankertz, B. (2013). Gaze-independent BCI-spelling using rapid serial visual presentation (RSVP). *Clin. Neurophysiol.* 124, 901–908. doi: 10.1016/j.clinph.2012.12.050

Aloise, F., Aricò, P., Schettini, F., Riccio, A., Salinari, S., Mattia, D., et al. (2012). A covert attention P300-based brain-computer interface: geospell. *Ergonomics* 55, 538–551. doi: 10.1080/00140139.2012.661084

Brunner, P., Joshi, S., Briskin, S., Wolpaw, J. R., Bischof, H., and Schalk, G. (2010). Does the 'P300' speller depend on eye gaze? *J. Neural Eng.* 7:056013. doi: 10.1088/1741-2560/7/5/056013

Chen, L., Jin, J., Daly, I., Zhang, Y., Wang, X., and Cichocki, A. (2016). Exploring combinations of different color and facial expression stimuli for gaze-independent BCIs. *Front. Comp. Neurosci.* 10:5. doi: 10.3389/fncom.2016.00005

Chennu, S., Alsufyani, A., Filetti, M., Owen, A. M., and Bowman, H. (2013). The cost of space independence in P300-BCI spellers. *J. NeuroEng. Rehabil.* 10:82. doi: 10.1186/1743-0003-10-82

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

Dijkstra, K. V., Farquhar, J. D. R., and Desain, P. W. M. (2020). The N400 for brain computer interfacing: complexities and opportunities. *J. Neural Eng.* 17:022001. doi: 10.1088/1741-2552/ab702e

Eimer, M. (2000). Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clin. Neurophysiol.* 111, 694–705. doi: 10.1016/S1388-2457(99)00285-0

Farwell, L. A., and Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr. Clin. Neurophysiol.* 70, 510–523. doi: 10.1016/0013-4694(88)90149-6

Fernández-Rodríguez, Á., Medina-Juliá, M. T., Velasco-Álvarez, F., and Ron-Angevin, R. (2019c). "Preliminary results using a P300 brain-computer interface speller: a possible interaction effect between presentation paradigm and set of stimuli," in *Advances in Computational Intelligence. IWANN 2019. Lecture Notes in Computer Science*, eds. I. Rojas, G. Joya, and A. Catala (Cham: Springer), 371–381.

Fernández-Rodríguez, Á., Velasco-Álvarez, F., Medina-Juliá, M. T., and Ron-Angevin, R. (2019a). Evaluation of emotional and neutral pictures as flashing stimuli using a P300 brain-computer interface speller. *J. Neural Eng.* 16:056024. doi: 10.1088/1741-2552/ab386d

Fernández-Rodríguez, Á., Velasco-Álvarez, F., Medina-Juliá, M. T., and Ron-Angevin, R. (2019b). Evaluation of flashing stimuli shape and colour heterogeneity using a p300 brain-computer interface speller. *Neurosci. Lett.* 709:134385. doi: 10.1016/j.neulet.2019.134385

Jiang, L., Wang, Y., Cai, B., Wang, Y., and Wang, Y. (2017). Spatial-temporal feature analysis on single-trial event related potential for rapid face identification. *Front. Comp. Neurosci.* 11:106. doi: 10.3389/fncom.2017.00106

Jin, J., Li, S., Daly, I., Miao, Y., Liu, C., Wang, X., et al. (2020). The study of generic model set for reducing calibration time in P300-based brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 3–12. doi: 10.1109/TNSRE.2019.2956488

Käthner, I., Wriessnegger, S. C., Müller-Putz, G. R., Kübler, A., and Halder, S. (2014). Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain – computer interface. *Biol. Psychol.* 102, 118–129. doi: 10.1016/j.biopsycho.2014.07.014

Kaufmann, T., Schulz, S. M., Grünzinger, C., and Kübler, A. (2011). Flashing characters with famous faces improves ERP-based brain-computer interface performance. *J. Neural Eng.* 8:056016. doi: 10.1088/1741-2560/8/5/056016

Kellicut-Jones, M. R., and Sellers, E. W. (2018). P300 brain-computer interface: comparing faces to size matched non-face stimuli. *Brain Comp. Interfaces* 5, 30–39. doi: 10.1080/2326263X.2018.1433776

Krusienski, D. J., Sellers, E. W., McFarland, D. J., Vaughan, T. M., and Wolpaw, J. R. (2008). Toward enhanced P300 speller performance. *J. Neurosci. Methods* 167, 15–21. doi: 10.1016/j.jneumeth.2007.07.017

Kübler, A., Neumann, N., Kaiser, J., Kotchoubey, B., Hinterberger, T., and Birbaumer, N. P. (2001). Brain-computer communication: self-regulation of slow cortical potentials for verbal communication. *Arch. Phys. Med. Rehabil.* 82, 1533–1539. doi: 10.1053/apmr.2001.26621

Lees, S., Dayan, N., Cecotti, H., McCullagh, P., Maguire, L., Lotte, F., et al. (2018). A review of rapid serial visual presentation-based brain-computer interfaces. *J. Neural Eng.* 15:aa9817. doi: 10.1088/1741-2552/aa9817

Li, Q., Liu, S., Li, J., and Bai, O. (2015). Use of a green familiar faces paradigm improves P300-speller brain-computer interface performance. *PLoS ONE* 10:0130325. doi: 10.1371/journal.pone.0130325

Lin, Z., Zhang, C., Zeng, Y., Tong, L., and Yan, B. (2018). A novel P300 BCI speller based on the triple RSVP paradigm. *Sci. Rep.* 8:3350. doi: 10.1038/s41598-018-21717-y

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., et al. (2018). A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *J. Neural Eng.* 15:031005. doi: 10.1088/1741-2552/aab2f2

Lu, Z., Li, Q., Gao, N., and Yang, J. (2020). The self-face paradigm improves the performance of the P300-speller system. *Front. Comp. Neurosci.* 13:93. doi: 10.3389/fncom.2019.00093

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique. 2nd Edn.* (Cambridge: MIT Press).

Mak, J. N., McFarland, D. J., Vaughan, T. M., McCane, L. M., Tsui, P. Z., Zeitlin, D. J., et al. (2012). EEG correlates of P300-based Brain-Computer Interface (BCI) performance in people with amyotrophic lateral sclerosis. *J. Neural Eng.* 9:026014. doi: 10.1088/1741-2560/9/2/026014

Mijani, A. M., Shamsollahi, M. B., and Hassani, M. S. (2019). A novel dual and triple shifted RSVP paradigm for P300 speller. *J. Neurosci. Methods* 328: 108420. doi: 10.1016/j.jneumeth.2019.108420

Nicolas-Alonso, L. F., and Gomez-Gil, J. (2012). Brain computer interfaces, a review. *Sensors* 12, 1211–1279. doi: 10.3390/s120201211

Pires, G., Nunes, U., and Castelo-Branco, M. (2011). "GIBS block speller: toward a gaze-independent P300-based BCI", in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* (Boston, MA), 6360–6364.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Rezeika, A., Benda, M., Stawicki, P., Gembler, F., Saboor, A., and Volosyak, I. (2018). Brain–computer interface spellers: a review. *Brain Sci.* 8:57. doi: 10.3390/brainsci8040057

Ron-Angevin, R., Varona-Moya, S., and Silva-Sauer, L. D. (2015). Initial test of a T9-like P300-based speller by an ALS patient. *J. Neural Eng.* 12:046023. doi: 10.1088/1741-2560/12/4/046023

Ryan, D. B., Townsend, G., Gates, N. A., Colwell, K., and Sellers, E. W. (2017). Evaluating brain-computer interface performance using color in the P300 checkerboard speller. *Clin. Neurophysiol.* 128, 2050–2057. doi: 10.1016/j.clinph.2017.07.397

Salvaris, M., and Sepulveda, F. (2009). Visual modifications on the P300 speller BCI paradigm. *J. Neural Eng.* 6:046011. doi: 10.1088/1741-2560/6/4/046011

Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N., and Wolpaw, J. R. (2004). BCI2000: a general-purpose Brain-Computer Interface (BCI) system. *IEEE Trans. Biomed. Eng.* 51, 1034–1043. doi: 10.1109/TBME.2004.827072

Sutton, S., Braren, M., Zubin, J., and John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science* 150, 1187–1188. doi: 10.1126/science.150.3700.1187

Tanaka, H. (2018). Face-sensitive P1 and N170 components are related to the perception of two-dimensional and three-dimensional objects. *NeuroReport* 29, 583–587. doi: 10.1097/WNR.0000000000001003

Tian, Y., Zhang, H., Pang, Y., and Lin, J. (2018). Classification for single-trial N170 during responding to facial picture with emotion. *Front. Comp. Neurosci.* 12:68. doi: 10.3389/fncom.2018.00068

Treder, M. S., and Blankertz, B. (2010). (C)Overt attention and visual speller design in an ERP-based brain-computer interface. *Behav. Brain Funct.* 6:28. doi: 10.1186/1744-9081-6-28

Treder, M. S., Schmidt, N. M., and Blankertz, B. (2011). Gaze-independent brain-computer interfaces based on covert attention and feature attention. *J. Neural Eng.* 8:066003. doi: 10.1088/1741-2560/8/6/066003

Velasco-Álvarez, F., Sancha-Ros, S., García-Garaluz, E., Fernández-Rodríguez, Á., Medina-Juliá, M. T., and Ron-Angevin, R. (2019). UMA-BCI speller: an easily configurable P300 speller tool for end users. *Comp. Methods Prog. Biomed.* 172, 127–138. doi: 10.1016/j.cmpb.2019.02.015

Vidal, J. J. (1973). Toward direct brain-computer communication. *Ann. Rev. Biophys. Bioeng.* 2, 157–180. doi: 10.1146/annurev.bb.02.060173.001105

Wolpaw, J. R., Ramoser, H., McFarland, D. J., and Pfurtscheller, G. (1998). EEG-based communication: improved accuracy by response verification. *IEEE Trans. Rehabil. Eng.* 6, 326–333. doi: 10.1109/86.712231

Won, D. O., Hwang, H. J., Kim, D. M., Muller, K. R., and Lee, S. W. (2018). Motion-based rapid serial visual presentation for gaze-independent brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 334–343. doi: 10.1109/TNSRE.2017.2736600

Xiao, X., Xu, M., Jin, J., Wang, Y., Jung, T. P., and Ming, D. (2020). Discriminative canonical pattern matching for single-trial classification of ERP components. *IEEE Trans. Biomed. Eng.* 67, 2266–2275. doi: 10.1109/TBME.2019.2958641

Xu, M., Han, J., Wang, Y., Jung, T. P., and Ming, D. (2020). Implementing over 100 command codes for a high-speed hybrid brain-computer interface using concurrent P300 and SSVEP features. *IEEE Trans. Biomed. Eng.* 67, 3073–3082. doi: 10.1109/TBME.2020.2975614

Xu, M., Xiao, X., Wang, Y., Qi, H., Jung, T. P., and Ming, D. (2018). A Brain-computer interface based on miniature-event-related potentials induced by very small lateral visual stimuli. *IEEE Trans. Biomed. Eng.* 65, 1166–1175. doi: 10.1109/TBME.2018.2799661

Zhao, Q., Onishi, A., Zhang, Y., Cao, J., Zhang, L., and Cichocki, A. (2011). "A novel oddball paradigm for affective BCIs using emotional faces as stimuli.", in *Neural Information Processing, ICONIP* (Shanghai), 279–286.

Zheng, X., Mondloch, C. J., and Segalowitz, S. J. (2012). The timing of individual face recognition in the brain. *Neuropsychologia* 50, 1451–1461. doi: 10.1016/j.neuropsychologia.2012.02.030

Conflict of Interest: 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: 15 January 2021
doi: 10.3389/fpsyg.2020.580287

Detection of Genuine and Posed Facial Expressions of Emotion: Databases and Methods

*Shan Jia[1,2], Shuo Wang[3], Chuanbo Hu[2], Paula J. Webster[3] and Xin Li[2]**

[1] State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, Wuhan, China, [2] Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, United States, [3] Department of Chemical and Biomedical Engineering, West Virginia University, Morgantown, WV, United States

Facial expressions of emotion play an important role in human social interactions. However, posed expressions of emotion are not always the same as genuine feelings. Recent research has found that facial expressions are increasingly used as a tool for understanding social interactions instead of personal emotions. Therefore, the credibility assessment of facial expressions, namely, the discrimination of genuine (spontaneous) expressions from posed (deliberate/volitional/deceptive) ones, is a crucial yet challenging task in facial expression understanding. With recent advances in computer vision and machine learning techniques, rapid progress has been made in recent years for automatic detection of genuine and posed facial expressions. This paper presents a general review of the relevant research, including several spontaneous vs. posed (SVP) facial expression databases and various computer vision based detection methods. In addition, a variety of factors that will influence the performance of SVP detection methods are discussed along with open issues and technical challenges in this nascent field.

Keywords: facial expressions analysis, spontaneous expression, posed expression, expressions classification, countermeasure

OPEN ACCESS

Edited by:
Rufin VanRullen,
Centre National de la Recherche Scientifique (CNRS), France

Reviewed by:
Haojiang Ying,
Soochow University, China
Jeffrey Wagman,
Illinois State University, United States

****Correspondence:***
Xin Li
xin.li@mail.wvu.edu

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

Received: *05 July 2020*
Accepted: *09 December 2020*
Published: *15 January 2021*

Citation:
Jia S, Wang S, Hu C, Webster PJ and Li X (2021) Detection of Genuine and Posed Facial Expressions of Emotion: Databases and Methods. Front. Psychol. 11:580287. doi: 10.3389/fpsyg.2020.580287

1. INTRODUCTION

Facial expressions, one of the main channels for understanding and interpreting emotions among social interactions, have been studied extensively in the past decades (Zuckerman et al., 1976; Motley and Camden, 1988). Most existing research works have focused on automatic facial expression recognition based on Ekman's theories (Ekman and Keltner, 1997), which suggests six basic emotions universal in all cultures, including happiness, surprise, anger, sadness, fear, and disgust. However, are facial expressions always the mirror of our innermost emotions as we have believed for centuries? Recent research (Crivelli et al., 2015) has found that facial expressions do not always reflect our true feelings. Instead of reliable readouts of people's emotional states, facial expressions tend to be increasingly posed and even deliberately to show our intentions and social goals. Therefore, understanding the credibility of facial expressions in revealing emotions has become an important yet challenging task in human behavioral research especially among the studies of social interaction, communication, anthropology, personality, and child development (Bartlett et al., 1999).

In the early 2000s, Ekman's suggestion (Ekman, 2003) that a small number of facial muscles are not readily subject to volitional control, has laid the foundation for distinguishing between spontaneous and posed facial expressions. Ekman called these "reliable facial muscles" and claimed that activities of these muscles communicate the presence of specific emotions (Ekman, 2009). This set of muscles therefore became particularly trustworthy emotion-specific cues to identify genuine experienced emotions because they tend to be difficult to produce voluntarily. Early research of discriminating genuine facial expressions from posed ones heavily relied on a variety of observer-based systems (Mehu et al., 2012) targeting these muscles. Rapid advances in computer vision and pattern recognition especially deep learning techniques have recently opened up new opportunities for automatic and efficient identification of these cues for SVP facial expression detection. A variety of SVP facial expression detection methods (Valstar et al., 2006; Dibeklioglu et al., 2010; Wu et al., 2014; Huynh and Kim, 2017; Park et al., 2020), as well as publicly available databases (Wang et al., 2010; Pfister et al., 2011; Mavadati et al., 2016; Cheng et al., 2018), have been proposed for facial expression credibility analysis.

As of 2020, there has been no systematic survey yet to summarize the advances of SVP facial expression detection in the past two decades. To fill in this gap, we present a general review of this pioneering work as well as the most recent studies in this field including both existing SVP databases and automatic detection algorithms. Through literature surveys and analysis, we have organized existing SVP detection methods into four categories (action units, spatial patterns, visual features, and hybrid) and identified a number of factors that will influence the performance of SVP detection methods. Furthermore, we attempt to provide new insights into the remaining challenges and open issues to address in the future.

2. SPONTANEOUS VS. POSED FACIAL EXPRESSION DATABASES

Early studies investigating facial expressions are mostly based on posed expressions due to the ease with which this data is collected where the subjects are asked to display or imitate each basic emotional expression. Spontaneous facial expressions, however, as natural expressions, need to be induced by various stimuli, such as odors (Simons et al., 2003), photos (Gajšek et al., 2009), and video clips (Pfister et al., 2011; Petridis et al., 2013). There have been several databases with single or multiple facial expressions collected to promote the research in automatic facial expression credibility detection. In this section, we first focus on databases with both spontaneous and posed facial expressions summarizing their details and characteristics. Then we review some databases with a single emotion category (either posed or spontaneous) but can provide rich data for detection of SVP facial expressions from different resources.

Table 1 provides an overview of existing SVP facial expression databases with both spontaneous and posed expressions. The MMI facial expression database (Pantic et al., 2005) was first collected with only posed expressions for facial expression recognition. Later, data with three spontaneous expressions (disgust, happiness, and surprise) were added with audio-visual recordings based on video clips as stimuli (Valstar and Pantic, 2010). USTC-NVIE (Wang et al., 2010) is a visible and infrared thermal SVP database. Six spontaneous emotions consisting of image sequences from onset to apex[1], were also induced by screening carefully selected videos, while the posed emotions consist of apex images. CK+ database (Lucey et al., 2010), UvA-NEMO (Dibeklioğlu et al., 2012), and MAHNOB database (Petridis et al., 2013) all focused on the subject's smile, which is the easiest emotional facial expression to pose voluntarily. Specifically, the video sequences in the CK+ database were fully coded based on the Facial Action Coding System (FACS) (Ekman, 1997) with facial action units (AUs) as emotion labels, while videos in MAHNOB recorded both smiles and laughter with microphones, visible, and thermal cameras.

The SPOS Corpus database (Pfister et al., 2011) included six basic SVP emotions, with labels for onset, apex, offset and ends with two annotators according to subjects' self-reported emotions. The BioVid dataset (Walter et al., 2013) specifically targeted pain with heat stimulation, and both biosignals (such as skin conductance level [SCL], electrocardiogram [ECG], electromyogram (EMG), and electroencephalography [EEG]) and video signals were recorded. DISFA and DISFA+ databases (Mavadati et al., 2013, 2016) contain spontaneous and posed facial expressions, respectively, with 12 coded AUs labeled using FACS and 66 landmark points. In addition to basic facial expressions, DISFA+ also includes 30 facial actions by asking participants to imitate and pose specific expressions. Originally proposed for the ChaLearn LAP Real vs. Fake Expressed Emotion Challenge in 2017, the SASE-FE database (Wan et al., 2017; Kulkarni et al., 2018) collected six expressions by asking participants to pose artificial facial expressions or showing participants video clips to induce genuine expressions of emotion. **Figure 1** illustrates several examples of video clips selected by psychologists to induce specific emotions in this database. Most recently, a large scale 4D database, 4DFAB (Cheng et al., 2018), was introduced with 6 basic SVP expressions, recorded in four different sessions spanning over a 5-year period. This is the first work to investigate the use of 4D spontaneous behaviors in biometric applications.

We further introduce some databases widely-used in the emotion detection field with either posed or spontaneous facial expressions, which can provide rich data with different resources for SVP facial expression detection. **Table 2** shows the details of these popular facial expression databases. The Karolinska Directed Emotional Face (KDEF) dataset (Lundqvist et al., 1998) contains 4,900 images (562 × 762 pixels) from 70 subjects, each with seven posed emotional expressions taken from five different angles. Oulu-CASIA (Zhao et al., 2011) is a

[1] Onset, apex, along with offset, and neutral, are four possible temporal segments of facial actions during the expression development (generally in the order of neutral→ onset → apex → offset → neutral). In the onset phase, muscles are contracting and changes in appearance are growing stronger. In the apex phase, the facial action is at a peak with no more changes in appearance. The offset phase describes that the muscles of the facial action are relaxing and the face returns to its original and neutral appearance, where there are no signs of activation of the investigated facial action.

TABLE 1 | Description of SVP facial expression databases with both spontaneous and posed facial expressions (AU-Action Units).

Dataset	Expression	#Sub	#M/F	Age	#P/S	Format	Feature	References
MMI	Multiple	25	13/12	20–32	2489/392	Video	Audio-visual; single and combinations of AUs	Valstar and Pantic, 2010
USTC-NVIE	Multiple	215	157/58	17–31	-/-	Frame	Visible + infrared thermal images	Wang et al., 2010
CK+	Smile	210	65/145	18–50	593/122	Frame	Multiple posed expressions, only un-posed smile, FACS coded	Lucey et al., 2010
SPOS Corpus	Multiple	7	4/3	/	51/147	Frame	Visible + infrared	Pfister et al., 2011
UvA-NEMO	Smile	400	215/185	8–76	643/597	Video	The largest smile database	Dibeklioğlu et al., 2012
MAHNOB	Smile	22	12/10	∼28	563/101	Video	Audio-visual, thermal recording	Petridis et al., 2013
BioVid	Pain	90	45/45	18–65	630/8700	Video	Biopotential signals, depth information	Walter et al., 2013
DISFA	Multiple	27	15/12	18–50	0/54	Video	AU labels and landmarks	Mavadati et al., 2013
DISFA+	Multiple	9	4/5	18–50	644/0	Frame	AU labels, 42 facial actions	Mavadati et al., 2016
SASE-FE	Multiple	50	-/-	19–36	300/300	Video	3 subsets	Wan et al., 2017
4DFAB	Multiple	180	120/60	5–75	-/-	4D video	Dynamic high-resolution 3D faces, 79 face landmarks	Cheng et al., 2018

A
B
C
D
E
F
WEIRD WORMS SHOOTS WHITE GOO ALL OVER HUMAN HAND
PUBLIC FREAKOUT ANGRY PEOPLE GONE CRAZY COMPILATION (Part 2) ACTV
Happy dogs doing funny things
Disgusting! Parasites, zits, insects in people's ears & more: prepare to lose your lunch - TomoNews
50 Heartbreaking Movie Moments | SUPERCUT
The Unexpected John Cena Prank | Hidden Camera
The 7 basic emotions - Do you recognise all facial expressions?

FIGURE 1 | Screenshots of video clips to induce specific emotions in SASE-FE database (Copyright permission is obtained from Kulkarni et al., 2018). These video stimuli contain either specific scenes (such as **A,D,E**), objects (such as **B,C**), or the target emotions themselves (such as **D–F**) for emotion elicitation.

NIR-VIS posed expression database with 2,880 image sequences collected from 80 subjects. Six basic expressions are recorded in the frontal direction under three different lighting conditions. Another widely-used posed expression dataset is the Japanese Female Facial Expressions (JAFFE) (Lyons et al., 2020), which consists of 213 grayscale images with seven emotions from 10 Japanese females. In terms of spontaneous expressions, the MPI dataset (Kaulard et al., 2012) collects 55 expressions with high diversity in three repetitions, two intensities, and three recording angles from 19 German subjects. The Binghamton-Pittsburgh 3D Dynamic Spontaneous (BP4D-Spontaneous) (Zhang et al., 2014) dataset collects both 2D and 3D videos of 41 participants from different races.

There are also facial expression databases with rich data collected in the wild, such as the Real-world Affective Database (RAF-DB) (Li S. et al., 2017), Real-world Affective Faces Multi Label (RAF-ML) (Li and Deng, 2019), and Aff-wild database (Kollias et al., 2019), or collected from movies, such as the Acted Facial Expressions in the Wild (AFEW) and its static subset Static Facial Expressions in the Wild (SFEW). These kinds of data are of great variability to reflect the real-world situations (please refer to recent surveys [Huang et al., 2019; Saxena et al., 2020] for more details about these facial expression databases).

3. DETECTION OF GENUINE AND POSED FACIAL EXPRESSIONS

Posed facial expressions, due to their deliberate and artificial nature, always differ from genuine ones remarkably in terms of intensity, configuration, and duration, which have been explored as distinct features for SVP facial expression recognition. Based on different distinct clues, we classify existing methods into four categories: *muscle movement (action units) based, spatial patterns based, texture features based, and hybrid methods*.

3.1. Muscle Movement (Action Units) Based

Early research on distinguishing genuine facial expressions from posed ones rely a lot on the analysis of facial muscle movements. This class of methods is based on the assumption that some specific facial muscles are particularly trustworthy cues due to the intrinsic difficulty of producing them voluntarily (Ekman, 2003). In these studies, the Facial Action Coding System (FACS) (Ekman and Rosenberg, 2005) is the most widely-used tool for decomposing facial expressions into individual components of muscle movements, called Action Units (AUs), as shown in **Figure 2A**. Several studies have explored the differences of muscle movements (AUs) in spontaneous and posed facial expressions, including the AU's amplitude, maximum speed, and duration (please refer to **Figure 2B** for an example).

It is known that spontaneous smiles have a smaller amplitude, but a larger and more consistent relation between amplitude and duration than deliberate, posed smiles (Baloh et al., 1975). Based on this observation, a method in Cohn and Schmidt (2003) used timing and amplitude measures of smile onsets for detection and achieved the recognition rate of 93% with a linear discriminant analysis classifier (LDA). The method in Valstar et al. (2006) was the first attempt to automatically determine whether an observed facial action was displayed deliberately or spontaneously. They proposed to detect SVP brow actions based on automatic detection of three AUs (AU1, AU2, and AU4) and their temporal segments (onset, apex, offset) produced

TABLE 2 | Description of facial expression databases with either spontaneous (S) or posed (P) facial expressions.

Dataset	Expression	#Sub	#M/F	Age	#P/S	Format	Feature	References
KDEF	Multiple	70	35/35	20–30	P-4900	Image	5 different angles	Lundqvist et al., 1998
Oulu-CASIA	Multiple	80	59/21	23–58	P-2880	Image	Visible + infrared	Zhao et al., 2011
JAFFE	Multiple	10	0/10	/	P-213	Image	Japanese female, grayscale images	Lyons et al., 2020
MPI	Multiple	19	9/10	20–30	S-1045	Image	German participants, high diversity	Kaulard et al., 2012
BP4D-Spontaneous	Multiple	41	18/23	18–29	S-328	Video	Multiple races, both 2D + 3D videos	Zhang et al., 2014

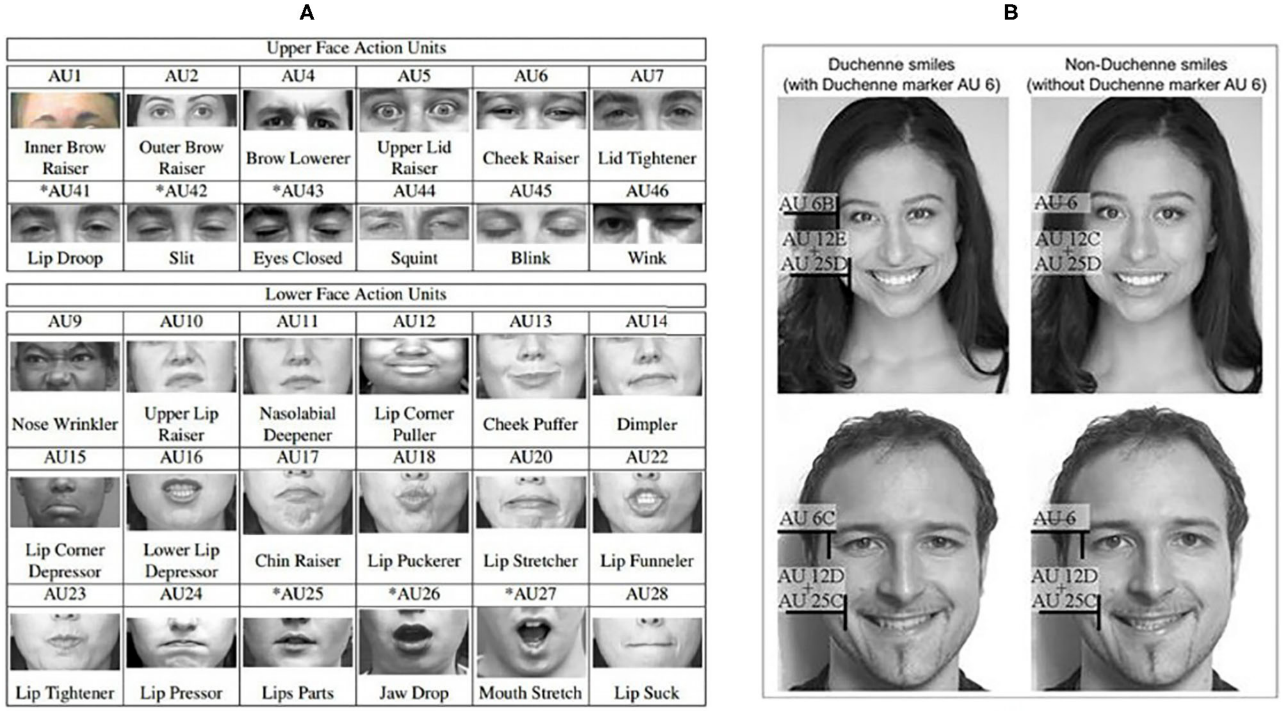


FIGURE 2 | Examples of Facial Action Coding System (FACS) Action Units (AUs) **(A)** Upper and lower face AUs (Copyright permission is obtained from la Torre De et al., 2015), **(B)** Different AUs in Duchenne (genuine) smiles (AU 6, 12, 25) and non-Duchenne smiles (AU12, 25) (Copyright permission is obtained from Bogodistov and Dost, 2017).

by movements of the eyebrows. Experiments on the combined databases have achieved 98.80% accuracy. Later works (Bartlett et al., 2006, 2008) extracted five statistic features (median, maximum, range, first-to-third quartile difference) of 20 AUs in each video segment for classification of posed and spontaneous pain. They reported a 72% classification accuracy on their own dataset. To detect SVP smiles, the method in Schmidt et al. (2009) quantified lip corner and eyebrow movement during periods of visible smiles and eyebrow raises, and found maximum speed and amplitude were greater and duration shorter in deliberate compared to spontaneous eyebrow raises. Aiming at multiple facial expressions, the method (Saxen et al., 2017) generated a 440-dimensional statistic feature space from the intensity series of seven facial AUs, and increased the performance to 73% by training an ensemble of Rank SVMs on the SASE-FE database. Alternatively, recent work in Racoviţeanu et al. (2019) used the AlexNet CNN architecture on 12 AU intensities to obtain the features in a transfer learning task. Training on the DISFA database, and testing on SPOS, the method achieved an average accuracy of 72.10%. A brief overview of these methods has been shown in **Table 3**.

3.2. Spatial Patterns Based

This category of methods aim at exploring spatial patterns based on temporal dynamics of different modalities, such as facial landmarks and shapes of facial components. A multimodal system based on fusion of temporal attributes including tracked points of the face, head, and shoulder were proposed in Valstar et al. (2007) to discern posed from spontaneous smiles. Best results were obtained with late fusion of all modalities of 94% on 202 videos from the MMI database. Specifically regarding smiles, a study in Van Der Geld et al. (2008) analyzed differences in tooth display, lip-line height, and smile width between SVP smiles. They revealed several findings in SVP smiling differences.

TABLE 3 | A brief overview of muscle movement based spontaneous vs. posed (SVP) detection methods (AU-action unit; LDA-linear discriminant analysis [classifier]; SVM-support vector machine).

References	Method (features)	Expression	AU	Classification	Database	Accuracy (%)
Cohn and Schmidt, 2003	Using timing and amplitude measures of smile onsets	Smile	6, 12, 15, 17	LDA	Self-collected	93.00
Valstar et al., 2006	Temporal dynamics of brow actions based on AUs and their temporal segments (onset, apex, offset)	Multiple (6)	1, 2, 4	Relevance Vector Machine	MMI+DS118+ CK+(262)	90.80
Bartlett et al., 2008	Statistic features of 20 AUs in each video segment	Pain	1, 2, 4–7, 9, 10, 12, 14, 15, 17, 18, 20, 23–26	Non-linear SVM	Self-collected	72.00
Schmidt et al., 2009	Maximum speed and amplitude of movement onset of lip corner and eyebrow; AFIA to measure movement	Smile	6, 12, 14, 15, 17, 23, 24, 50	(-)	Self-collected	(-)
Saxen et al., 2017	statistic features (440-dimensional) from the intensity time series of 7 facial AUs	Multiple (6)	1, 2, 4, 6, 9, 12, 25	Rank SVMs	SASE-FE	73.00
Racoviţeanu et al., 2019	AlexNet CNN architecture on 12 AU intensities to obtain the features in a transfer learning manner	Multiple (6)	1, 2, 4-6, 9, 12, 15, 17, 20, 25, 26	SVM	DISFA, SPOS	72.10

For example, maxillary lip-line heights in genuine smiles were significantly higher than those in posed smiles. When compared to genuine smiling, the tooth display in the (pre)molar area of posed smiling decreased by up to 30%, along with a significant reduction of smile width. Spatial patterns based on distance and angular features for eyelid movements were used in Dibeklioglu et al. (2010) and achieved 85 and 91% accuracy in discriminating SVP smiles on the BBC and CK databases, respectively. Based on fusing dynamics signals of eyelids, cheeks, and lip corners, more recent methods (Dibeklioğlu et al., 2012, 2015) achieved promising detection results on several SVP smile databases.

In multiple SVP facial expression detection studies, different schemes for spatial pattern modeling were used, including Restricted Boltzmann Machines (RBMs) based in Wang et al. (2015, 2016), Latent Regression Bayesian Network based in Gan et al. (2017), and interval temporal restricted Boltzmann machine (IT-RBM) in Wang et al. (2019). Results on several SVP databases confirmed the discriminative power and reliability of spatial patterns in distinguishing genuine and posed facial expressions. Similarly, Huynh and Kim (2017) used mirror neuron modeling and Long-short Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997) with parametric bias to extract features in the spatial-temporal domain from extracted facial landmarks, and achieved 66% accuracy on the BABE-FE database. **Table 4** includes an overview of these spatial pattern based detection methods.

3.3. Texture Features Based

Texture features based, such as Littlewort et al. (2009) designed a two-stage system to distinguish faked pain from real pain. It consisted of a detection stage for 20 facial actions using Gabor features and a SVM classification stage. The two-stage system achieved 88% accuracy on the UvA-NEMO dataset. Another method (Pfister et al., 2011) proposed a new feature (Completed local binary patterns from Three Orthogonal Planes [CLBP-TOP]), and fused the NIR and VIS modalities with the Multiple Kernel Learning (MKL) classifier, which achieved outstanding detection performance of 80.0% on the SPOS database. Finally, the approach in Gan et al. (2015) proposed to use pixel-wise differences between onset and apex face images as input features of a two-layer deep Boltzmann machine to distinguish SVP expressions. They achieved 84.62 and 91.73% on the SPOS and USTC-NVIE databases, respectively.

More recently, Mandal et al. (2016) explored several features, including deep CNN features, local phase quantization (LPQ), dense optical flow and histogram of gradient (HOG), to classify SVP smiles. With Eulerian Video Magnification (EVM) for micro-expression smile amplification, the HOG features outperformed other features with an accuracy of 78.14% on the UvA-NEMO Smile Database. Instead of using pixel-level differences, the method (Xu et al., 2017) designed a new layer named "comparison layer" for the deep CNN to generate high-level representations of the differences of onset and apex images, and verified its effectiveness on SPOS (83.34%) and USTC-NVIE (97.98%) databases. The latest work (Tavakolian et al., 2019) presents a Residual Generative Adversarial Network (R-GAN) based method to discriminate SVP pain expression by magnifying the subtle changes in faces. Experimental results have shown the state-of-the-art performance on three databases, with 91.34% on UNBC-McMaster (Lucey et al., 2011) with spontaneous pain expressions only, 85.05% on BiodVid, and 96.52% on STOIC (Roy et al., 2007) with posed expressions only. A brief overview of these methods is shown in **Table 5**.

3.4. Hybrid Methods

Hybrid methods combined different classes of features for discriminating SVP facial expressions. Experiments on still images were conducted in Zhang et al. (2011) to show that appearance features (e.g., Scale-Invariant Feature Transform [SIFT] (Lowe, 2004)) play a significantly more important role than geometric features (e.g., facial animation parameters [FAP] (Aleksic and Katsaggelos, 2006)) on SVP emotion discrimination, and fusion of them leads to marginal improvement over SIFT appearance features. The average classification accuracy of six

TABLE 4 | A brief overview of spatial patterns based spontaneous vs. posed (SVP) detection methods (SVM-support vector machine; RBM-Restricted Boltzmann Machines).

References	Method (features)	Expression	Classification	Database	Accuracy (%)
Valstar et al., 2007	Fusing temporal dynamics of head (6 features), face (12 points), and shoulder (5 points) modalities	Smile	GentleSVM-Sigmoid	MMI (202)	94.00
Van Der Geld et al., 2008	Analyzing tooth display, lip position and smile width in a dental perspective	Smile	(-)	Self-collected	(-)
Dibeklioglu et al., 2010	Distance-based and angular features for eyelid movements	Smile	Naive Bayes	BBC, CK	85.00, 91.00
Dibeklioğlu et al., 2012	Fusing the dynamics of eyelid, cheek, and lip corner movements	Smile	linear SVM	BBC, SPOS, UvA-NEMO	90.00, 75.00, 87.02
Dibeklioğlu et al., 2015	Dynamics of eyelid, cheek, and lip corner movements	Smile	SVM	BBC, SPOS, UvA-NEMO, MMI	90.00, 78.75, 92.10, 89.69
Wang et al., 2015, 2016	Spatial pattern modeling based on multiple RBMs and incorporating gender and expression categories as privileged information	Multiple (6)	RBMs	SPOS, USTC-NVIE, MMI	76.07, 92.61, 89.79
Gan et al., 2017	Spatial patterns based on Latent Regression Bayesian Network from he displacements of facial feature points	Multiple (6)	Bayesian Networks	SPOS, USTC-NVIE	76.07, 98.74
Huynh and Kim, 2017	Spatial-temporal features using mirror neuron modeling and LSTM with parametric bias from facial landmarks	Multiple (6)	Gradient boosting	SASE-FE	66.70
Wang et al., 2019	Universal spatial patterns and complicated temporal patterns using IT-RBM dynamic model	Multiple (6)	Bayesian network	SPOS, DISFA+	83.76, 96.24

TABLE 5 | A brief overview of texture features based spontaneous vs. posed (SVP) detection methods.

References	Method (features)	Expression	Classification	Database	Accuracy (%)
Littlewort et al., 2009	Gabor features based	Pain	Gaussian SVM	UvA-NEMO	88.00
Pfister et al., 2011	Spatiotemporal local texture descriptor (CLBP-TOP), fusing the NIR and VIS modalities	Multiple (6)	MKL	SPOS	80.00
Liu and Wang, 2012	Temperature features from Infrared thermal images	Multiple (6)	Bayesian Networks	USTC-NIVE	76.70
Gan et al., 2015	A two-layer deep Boltzmann machine model based	Multiple (6)	Haarcascades	SPOS, USTC-NVIE	84.62, 91.73
Mandal et al., 2016	Several features: using CNN face features, LPQ, dense optical flow and HOG, and HOG with the best result	Smile	Linear SVM	UvA-NEMO	78.14
Xu et al., 2017	Learned features based on CNN from the difference of structural changes between the onset and apex images	Multiple (6)	Linear SVM	SPOS, USTC-NVIE	83.34, 97.98
Tavakolian et al., 2019	Encoding the dynamic and appearance of a video into an image map based on spatiotemporal pooling, then using R-GAN model for discrimination	Pain	Softmax	BioVid Heat Pain, STOIC, UNBC-McMaster	85.05, 96.52, 91.34

emotions is 79.4% (the emotion of surprise achieved the best result of 83.4% while anger had the worst at 77.2% accuracy) on the USTC-NVIE database. Sequential geometric features based on facial landmarks and texture features using HOG were combined in Li L. et al. (2017). A temporal attention gated model is designed for HOG features, combining with LSTM autoencoder (eLSTM) to capture discriminative features from facial landmark sequences. The proposed model performed well on most emotions on SASE-FE database, with an average accuracy of 68%. Mandal and Ouarti (2017) fused subtle (micro) changes by tracking a series of facial fiducial markers with local and global motion based on dense optical flow, and achieved 74.68% accuracy using combined features from the eyes and lips, slightly better than using only the lips (73.44%) and using only the eyes (71.14%) on the UvA-NEMO smile database. A different hybrid method in Kulkarni et al. (2018) combined learned static CNN representations from still images with facial landmark trajectories, and achieved promising performance

TABLE 6 | A brief overview of hybrid methods for SVP detection.

References	Method (features)	Expression	Classification	Database	Accuracy (%)
Zhang et al., 2011	SIFT appearance based features and FAP geometric features	Multiple (6)	RBF SVM	USTC-NVIE	79.40
Li L. et al., 2017	Combining sequential geometric features based on facial landmarks and texture features using HOG	Multiple (6)	Sigmoid	SASE-FE	68
Mandal and Ouarti, 2017	Fusing subtle (micro) changes by tracking a series of facial fiducial markers with local and global motion based on dense optical flow	Smile	SVM	UvA-NEMO	74.68
Kulkarni et al., 2018	Combining learned static CNN representations from still images with facial landmark trajectories	Multiple (6)	Linear SVM	SASE-FE	70.20
Saito et al., 2020	Combining hardware (16 sensors embedded with the smart eyewear) with software-based method to get geometric and temporal features	Smile	Linear SVM	Self-collected	94.60

not only in emotion recognition, but also in detecting genuine and posed facial expressions on the BABE-FE database with data augmentation (70.2% accuracy). Most recently, Saito et al. (2020) combined hardware (16 sensors embedded with the smart eye-wear) with a software-based method to extract geometric and temporal features to classify smiles into either "spontaneous" or "posed," with an accuracy of 94.6% on their own database. See **Table 6** for a brief summary of these hybrid SVP facial expression detection methods.

4. DISCUSSIONS

The studies reviewed in the previous section indicate two key factors in the research on automatic SVP facial expression detection: *collection of SVP facial expression data* and *design of automatic detection methods*. We first discuss our findings in existing studies from the perspective of data collection and detection methodology, respectively. Then, we attempt to address several new challenging issues, including the necessity of collecting diverse datasets as well as performing a unified evaluation in terms of detection accuracy and generalizability.

4.1. Data Collection

The databases for SVP facial expressions play a significant role in benchmarking the effectiveness and practicality of different detection schemes. From **Tables 2–5**, it can be observed that the detection performance of the same detection method can vary widely in different databases. Such performance differences can be attributed to several uncertainty factors of data collection. As the collection process is mostly based on recording subjects' facial expressions when they are shown various stimuli (such as movie clips), the data size, subject selection, recording environment, and stimuli materials, all have a direct effect on the visual quality of collected video data. A detailed discussion of these influencing factors is included below:

- Several methods in **Tables 3**, **4** have illustrated worse detection performance on the smaller SPOS dataset (with seven subjects) than that on the larger USTC-NVIE dataset (with 215 subjects). This is because using a limited number of samples will not only limit the detection ability of data-driven based methods but also weaken the detection performance in practical applications.
- In terms of subjects, both age and gender will affect the SVP facial expression detection. Dibeklioğlu et al. (2012) has explored the effect of subject age by splitting the UvA-NEMO smile database into young (age < 18) and adults (age $\geq$ 18 years), and found that eyelid-and-cheek features provided more reliable classification for adults, while lip-corner features performed better on young people. They further explored the gender effect in their completely automatic SVP detection method using dynamic features in different face regions and temporal phases (Dibeklioğlu et al., 2015). Experimental results showed that the correct classification rates on males were better than females for different facial region features. Such performance differences can be attributed to the fact that male subjects have more discriminative geometric features (distances between different landmark pairs) than female subjects. They also improved their detection performance by using age or gender as labels. Similarly, Wang et al. (2019) considered the influence of gender, and incorporated it as the input for performance improvement of their expression analysis model.
- The recording environment can vary greatly between studies in terms of the recording devices and lighting conditions. Most existing databases record images/videos of subjects in indoor controlled environments, which may limit the diversity of the data. In addition to visible images/videos, some studies have shown the impact of different modalities on improving the detection performance. Pfister et al. (2011) illustrated that the performance of fusion of NIR with visible images (80.0% accuracy) is better than using single NIR (78.2% accuracy) or visible images (72.0% accuracy) on the SPOS dataset. Although special devices are needed for data acquisition, the advantages of different modalities in revealing subtle features

deserve further investigation. It is also plausible to combine the information contained in multiple modalities for further performance improvement.

- In the collection of spontaneous facial expressions, different stimuli are often selected by those generating the face databases or by psychologists to induce specific emotions from participants. The stimuli determine the categories of facial expressions included in databases directly, which will further influence the evaluation of the database. Due to the differences in activation of muscles, such as with different intensities and in different facial regions, each emotion has varying difficulty levels in SVP expression detection. For example, happiness and anger can activate obvious muscles around the eye and mouth regions, which has been widely studied for feature extraction. Based on appearance and geometric features, Zhang et al. (2011) found that surprise was the easiest emotion for their model to classify correctly (83.4% accuracy on USTC-NVIE), followed by happiness with 80.5% accuracy, while disgust was the most difficult (76.1% accuracy). Similarly, Kulkarni et al. (2018) achieved better results in detecting SVP happiness (71.05% accuracy) and anger (69.40% accuracy), but worse results for disgust (63.05% accuracy) and contempt (60.85% accuracy) on the SASE-FE dataset. On the contrary, Li L. et al. (2017) obtained the highest accuracy (80%) for both disgust and happy, while 50% for contempt on the SASE-FE dataset. Overall, SVP happiness is relatively easy to recognize.

4.2. Detection Methodology

Performance differences can also be observed on the same dataset among approaches in different categories. Generally speaking, the methodology for SVP facial expression detection involves several modules, including data pre-processing, features extraction, and classification. These modules are discussed separately below:

- As each emotion has its own discriminative facial regions, data pre-processing to extract specific facial regions is needed not only in emotion recognition but also in posed vs. genuine classification. The study in Zhang et al. (2011) has found that in SVP emotion detection, the mouth region is more important for sadness; the nose is more important for surprise; both the nose and mouth regions are important for disgust, fear, and happiness, while the eyebrows, eyes, nose, and mouth are all important for anger. Another study (Liu and Wang, 2012) also explored different facial regions, including the forehead, eyes, nose, cheek, and mouth. Experimental results have shown that the forehead and cheek performed better than the other regions for most facial expressions (disgust, fear, sadness, and surprise), while the mouth region performed the worst for most facial expressions. Moreover, fusing all these regions achieved the best performance. In SVP smile detection, it was observed in Dibeklioğlu et al. (2012) that the discriminative power of the eyelid region is better than the cheek and lip corners. A different study in Mandal and Ouarti (2017) has found that lip-region features (73.44% accuracy on UvA-NEMO) outperformed the eye-region features (71.14% accuracy), while the combined features performed the best with 74.68% accuracy for SVP smile detection. Overall, fusion of multiple facial regions can improve the detection performance over individual features. Besides, varying video temporal segments (i.e., onset, apex, and offset) for feature extraction also leads to different levels of performance. Several studies (Cohn and Schmidt, 2003; Dibeklioğlu et al., 2012) have demonstrated that the onset phase performs best among individual phases in SVP facial expression detection.
- It is clear that the features extracted for distinguishing between posed and spontaneous facial expressions play a key role in detection performance. Most methods have explored temporal dynamics of different features for effective detection. We can observe from **Tables 2–5** that the detection performance varies greatly among different algorithms using the same database. The learned texture features from comparing the differences between images taken throughout the process of forming a facial expression proposed by Gan et al. (2015) and Xu et al. (2017) in **Table 5** performed better than muscle movement and spatial pattern based methods on the SPOS database, while on the USTC-NIVE database and smile SVP database UvA-NEMO, spatial patterns based methods achieve slightly higher accuracy than texture features, and significantly higher than other kinds of methods. Overall, texture features based and spatial patterns based methods show more promising detection abilities; but there still lacks a consensus about which type of features will be optimal for the task of SVP detection.
- The classifier used also has a great effect on most classification tasks, which has also been explored by researchers in distinction between spontaneous and posed facial expressions. Dibeklioglu et al. (2010) assessed the reliability of their features with continuous HMM, k-Nearest-Neighbor (kNN), and the Naive Bayes classifier, and found that the highest classification rate was achieved by the Naive Bayes classifier on two datasets. Pfister et al. (2011) compared support vector machine (SVM), Multiple Kernel Learning (MKL), and Random Forest decision tree (RF) classifier, and found RF outperformed SVM and MKL based on CLBP-TOP features on the SPOS database. Dibeklioğlu et al. (2015) compared Linear Discriminant, Logistic Regression, kNN, Naive Bayes, and SVM classifiers on UvA-NEMO smile dataset, and showed the outstanding performance of the SVM classifier under all testing scenarios. Racoviţeanu et al. (2019) also used SVM, combined with a Hard Negative Mining (HNM) paradigm, to produce the best performance among RF, SVM, and Multi-Layer Perceptron (MLP) classifiers. Overall, as the most widely-used classifier, SVM can provide outstanding performance on several databases. Whether recently developed deep learning-based classifiers can achieve further performance improvement remains to be explored.

4.3. Challenges and Opportunities

Based on the summary of existing studies in SVP facial expression detection, we further discuss the challenges in both data collection and detection methods for developing an automatic SVP facial expression recognizer.

The database creation procedure in existing SVP facial expression databases is diverse and there is a lack of a unified protocol or guidelines for high quality database collection. Several general steps are involved in the process of data collection, including subject selection, stimulus selection, recording process, and data annotation. In addition to the influencing factors that have been studied in existing detection methods, (e.g., the data size, age and gender of subjects, recording environment and devices, and stimuli materials [please see the details in section 4.1]), there are more factors that may influence the database quality and deserve further investigation. For example, most databases ignore the external factors, such as personality or mood of the participants in subject selection. Some databases gave an introduction to the experimental procedure for the subjects in advance (e.g., the USTC-NVIE dataset), while some gave no instructions to subjects on how they should react and what the aim of the study was (such as the MAHNOB dataset). In terms of the stimulus selection, there is no detailed description on how the video clip stimuli were selected by collectors or psychologists. Besides the recording environment, the recording distance, shooting angle, and more importantly, the order setup for recording different emotions (e.g., to reduce the interaction of different emotions, neutral clips were shown to subjects between segments in USTC-NVIE), will all have an effect on the quality of collected data. Further, unlike posed emotions which subjects are asked to display, spontaneous emotions induced by specific video clips, are more difficult to label. In the DISFA and MMI datasets, the data were annotated based on FACS coding of facial muscle actions. The USTC-NVIE and SPOS Corpus databases used self-reported data of subjects as the real emotion labels. We believe that designing a protocol to unify these procedures to conduct deeper investigations to determine their influence on SVP emotion detection will contribute to higher-quality and more credible SVP expressions collection.

Collection of SVP facial expression datasets that are large-scale and diverse in subjects selection, emotion categories, and recording environment (such as fully unconstrained environments) are also in high demand to reflect the real-world situations. Existing databases with both spontaneous and posed facial expressions of the same subjects are limited in data size due to the difficulty of data collection. Moreover, the arbitrary movement of subjects, low resolution or occlusion (e.g., the person may not be looking directly at the camera) may occur in a realistic interaction environment, which has not been taken into consideration by existing databases. Taking advantage of rich datasets proposed for emotion detection is one alternative to help realize the full potential of data-driven detection methods. In addition, using the strength of the detection methods to aid the database creation is also worth exploring. For example, based on the findings that the data properties, such as subject age and gender can contribute to improvement of detection performance, the subject distribution in terms of age, gender, race, and even personality, should be considered in data collection, which will not only improve the data diversity, but also inspire researchers to design more effective and practical detection methods.

Another challenge is the lack of a unified evaluation standard (such as experimental data and annotation) for SVP facial expression detection. Therefore, it is difficult to compare the diverse methods reviewed in this paper on a common experimental setting. Although several studies have reported promising detection accuracy on specific datasets, they have observed the apparent gap of performance between posed facial expressions detection and genuine ones. For example, based on texture features, Liu and Wang (2012) found that it is much easier to distinguish all posed expressions (90.8% accuracy) than genuine ones (62.6%) using the USTC-NIVE database. Similarly, Mandal et al. (2016) also achieved higher classification accuracy of posed smiles than spontaneous ones (with over 10% gaps) on the UvA-NEMO dataset. However, two hybrid methods (Mandal and Ouarti, 2017; Kulkarni et al., 2018) both obtained higher accuracy in detecting genuine facial expressions than posed ones, with a 6% gap in methods (Mandal and Ouarti, 2017) on the UvA-NEMO Smile database, while an average of 7.9% gap in methods (Kulkarni et al., 2018) on the SASE-FE database. Such inconsistent differences, influenced by both feature extraction methods and databases, deserve to be reconciled in future research.

Furthermore, how to improve the generalization ability of SVP detection on multiple universal facial expressions, or improve the performance on a specific emotion based on its unique facial features, also deserves further investigation. Existing research can achieve promising detection performance on specific datasets under intra-dataset testing scenarios. However, few studies conduct cross-dataset testing evaluation (Cao et al., 2010) to show the detection robustness on facial expressions from unknown resources. Hybrid methods with fused features from multiple descriptors, multiple face regions, multiple image modalities, or multiple visual cues (such as including head movement and body gesture) require further investigation for the improvement of facial expression detection performance.

5. CONCLUSIONS

With the emerging and increasingly supported theory that facial expressions do not always reflect our genuine feelings, automatic detection of spontaneous and posed facial expressions have become increasingly important in human behavior analysis. This article has summarized recent advances of SVP facial expression detection over the past two decades. A total of sixteen databases and nearly thirty detection methods have been reviewed and analyzed here. Particularly, we have provided detailed discussions on existing SVP facial expression detection studies from the perspectives of both data collection and detection methodology. Several challenging issues have also been identified to gain a deeper understanding of this emerging field. This review is expected to serve as a good starting point for researchers who consider developing automatic and effective models for genuine and posed facial expression recognition.

One area that has not been covered by this paper is the 3D dynamic facial expression databases (Sandbach et al., 2012; Zhang et al., 2013). As 3D scanning technology (e.g., Kinect and LIDAR) rapidly advances, SVP detection from 3D, instead of 2D data, might become feasible in the near future. Can 3D information facilitate the challenging task of SVP facial expression detection? It remains to be explored. Research on SVP detection also has connections with other potential applications, such as Parkinson's disease (Smith et al., 1996), deception detection (Granhag and Strömwall, 2004), and alexithymia (McDonald and Prkachin, 1990). More sophisticated computational tools, such as deep learning based methods might help boost the research progress in SVP detection. It is likely that the field of facial expression recognition and affective computing will continue to grow in the new decade.

AUTHOR CONTRIBUTIONS

SJ and CH did the literature survey together. SJ and XL jointly worked on the write-up. PW and SW handled the revision. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by an NSF CAREER Award (1945230), ORAU Ralph E. Powe Junior Faculty Enhancement Award, West Virginia University (WVU), WVU PSCoR Program, the Dana Foundation (to SW), an NSF Grant (OAC-1839909), and the WV Higher Education Policy Commission Grant (HEPC.dsr.18.5) (to XL). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

Aleksic, P. S., and Katsaggelos, A. K. (2006). Automatic facial expression recognition using facial animation parameters and multistream hmms. *IEEE Trans. Inform. Forens. Sec.* 1, 3–11. doi: 10.1109/TIFS.2005.863510

Baloh, R. W., Sills, A. W., Kumley, W. E., and Honrubia, V. (1975). Quantitative measurement of saccade amplitude, duration, and velocity. *Neurology* 25, 1065–1065. doi: 10.1212/WNL.25.11.1065

Bartlett, M., Littlewort, G., Vural, E., Lee, K., Cetin, M., Ercil, A., et al. (2008). "Data mining spontaneous facial behavior with automatic expression coding," in *Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction* (Berlin; Heidelberg: Springer), 1–20. doi: 10.1007/978-3-540-70872-8_1

Bartlett, M. S., Hager, J. C., Ekman, P., and Sejnowski, T. J. (1999). Measuring facial expressions by computer image analysis. *Psychophysiology* 36, 253–263. doi: 10.1017/S0048577299971664

Bartlett, M. S., Littlewort, G., Frank, M. G., Lainscsek, C., Fasel, I. R., Movellan, J. R., et al. (2006). Automatic recognition of facial actions in spontaneous expressions. *J. Multimed.* 1, 22–35. doi: 10.4304/jmm.1.6.22-35

Bogodistov, Y., and Dost, F. (2017). Proximity begins with a smile, but which one? Associating non-duchenne smiles with higher psychological distance. *Front. Psychol.* 8:1374. doi: 10.3389/fpsyg.2017.01374

Cao, L., Liu, Z., and Huang, T. S. (2010). "Cross-dataset action detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (San Francisco, CA), 1998–2005. doi: 10.1109/CVPR.2010.5539875

Cheng, S., Kotsia, I., Pantic, M., and Zafeiriou, S. (2018). "4DFAB: a large scale 4D database for facial expression analysis and biometric applications," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City), 5117–5126. doi: 10.1109/CVPR.2018.00537

Cohn, J. F., and Schmidt, K. L. (2003). "The timing of facial motion in posed and spontaneous smiles," in *International Journal of Wavelets Multiresolution & Information Processing* 2.02, 121–132.

Crivelli, C., Carrera, P., and Fernández-Dols, J.-M. (2015). Are smiles a sign of happiness? Spontaneous expressions of judo winners. *Evol. Hum. Behav.* 36, 52–58. doi: 10.1016/j.evolhumbehav.2014.08.009

Dibeklioğlu, H., Salah, A. A., and Gevers, T. (2012). "Are you really smiling at me? Spontaneous versus posed enjoyment smiles," in *European Conference on Computer Vision* (Berlin; Heidelberg: Springer), 525–538. doi: 10.1007/978-3-642-33712-3_38

Dibeklioğlu, H., Salah, A. A., and Gevers, T. (2015). Recognition of genuine smiles. *IEEE Trans. Multimed.* 17, 279–294. doi: 10.1109/TMM.2015.2394777

Dibeklioglu, H., Valenti, R., Salah, A. A., and Gevers, T. (2010). "Eyes do not lie: spontaneous versus posed smiles," in *Proceedings of the 18th ACM International Conference on Multimedia* (Firenze), 703–706. doi: 10.1145/1873951.1874056

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

Ekman, P. (2009). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (Revised Edition)*. New York, NY: WW Norton & Company.

Ekman, P., and Keltner, D. (1997). "Universal facial expressions of emotion," in *Nonverbal Communication: Where Nature Meets Culture*, eds U. Segerstrale and P. Molnar (Francisco, CA: California Mental Health Research Digest), 27–46.

Ekman, P., and Rosenberg, E. (2005). *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Encoding System (FACS)*. California, CA: Oxford University Press.

Ekman, R. (1997). *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. New York, NY: Oxford University Press.

Gajšek, R., Štruc, V., Mihelič, F., Podlesek, A., Komidar, L., Sočan, G., et al. (2009). Multi-modal emotional database: AvID. *Informatica* 33, 101–106.

Gan, Q., Nie, S., Wang, S., and Ji, Q. (2017). "Differentiating between posed and spontaneous expressions with latent regression bayesian network," in *Thirty-First AAAI Conference on Artificial Intelligence*. San Francisco, CA.

Gan, Q., Wu, C., Wang, S., and Ji, Q. (2015). "Posed and spontaneous facial expression differentiation using deep Boltzmann machines," in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)* (Xian: IEEE), 643–648. doi: 10.1109/ACII.2015.7344637

Granhag, P. A., and Strömwall, L. A. (2004). *The Detection of Deception in Forensic Contexts*. Cambridge University Press.

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Huang, Y., Chen, F., Lv, S., and Wang, X. (2019). Facial expression recognition: a survey. *Symmetry* 11:1189. doi: 10.3390/sym11101189

Huynh, X.-P., and Kim, Y.-G. (2017). "Discrimination between genuine versus fake emotion using long-short term memory with parametric bias and facial landmarks," in *Proceedings of the IEEE International Conference on Computer Vision Workshops* (Venice), 3065–3072.

Kaulard, K., Cunningham, D. W., Bülthoff, H. H., and Wallraven, C. (2012). The MPI facial expression database—a validated database of emotional and conversational facial expressions. *PLoS ONE* 7:e32321. doi: 10.1371/journal.pone.0032321

Kollias, D., Tzirakis, P., Nicolaou, M. A., Papaioannou, A., Zhao, G., Schuller, B., et al. (2019). Deep affect prediction in-the-wild: Aff-wild database and challenge, deep architectures, and beyond. *Int. J. Comput. Vision* 127, 907–929. doi: 10.1007/s11263-019-01158-4

Kulkarni, K., Corneanu, C., Ofodile, I., Escalera, S., Baro, X., Hyniewska, S., et al. (2018). Automatic recognition of facial displays of unfelt emotions. *IEEE Trans. Affect. Comput.* 2018, 1–14. doi: 10.1109/TAFFC.2018.2874996

la Torre De, F., Chu, W.-S., Xiong, X., Vicente, F., Ding, X., and Cohn, J. (2015). "Intraface," in *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops*, Vol. 1 (Ljubljana). doi: 10.1109/FG.2015.7163082

Li, L., Baltrusaitis, T., Sun, B., and Morency, L.-P. (2017). "Combining sequential geometry and texture features for distinguishing genuine and deceptive emotions," in *Proceedings of the IEEE International Conference on Computer Vision Workshops* (Venice), 3147–3153. doi: 10.1109/ICCVW.2017.372

Li, S., and Deng, W. (2019). Blended emotion in-the-wild: Multi-label facial expression recognition using crowdsourced annotations and deep locality feature learning. *Int. J. Comput. Vision* 127, 884–906. doi: 10.1007/s11263-018-1131-1

Li, S., Deng, W., and Du, J. (2017). "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu), 2852–2861. doi: 10.1109/CVPR.2017.277

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

Liu, Z., and Wang, S. (2012). "Posed and spontaneous expression distinguishment from infrared thermal images," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)* (Tsukuba Science City: IEEE), 1108–1111.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* 60, 91–110. doi: 10.1023/B:VISI.0000029664.99615.94

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. (2010). "The extended cohn-kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops* (IEEE), 94–101. doi: 10.1109/CVPRW.2010.5543262

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. *Face Gesture* 2011, 57–64. doi: 10.1109/FG.2011.5771462

Lundqvist, D., Flykt, A., and Öhman, A. (1998). *The Karolinska Directed Emotional Faces (KDEF)*. CD ROM from Department of Clinical Neuroscience, Psychology Section, Karolinska Institutet, Stockholm, Sweden.

Lyons, M. J., Kamachi, M., and Gyoba, J. (2020). Coding facial expressions with gabor wavelets (IVC special issue). *arXiv* 2009.05938.

Mandal, B., Lee, D., and Ouarti, N. (2016). "Distinguishing posed and spontaneous smiles by facial dynamics," in *Asian Conference on Computer Vision* (Taipei: Springer), 552–566. doi: 10.1007/978-3-319-54407-6_37

Mandal, B., and Ouarti, N. (2017). "Spontaneous versus posed smiles—can we tell the difference?" in *Proceedings of International Conference on Computer Vision and Image Processing* (Honolulu: Springer), 261–271. doi: 10.1007/978-981-10-2107-7_24

Mavadati, M., Sanger, P., and Mahoor, M. H. (2016). "Extended DISFA dataset: investigating posed and spontaneous facial expressions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (Las Vegas), 1–8. doi: 10.1109/CVPRW.2016.182

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. Comput.* 4, 151–160. doi: 10.1109/T-AFFC.2013.4

McDonald, P. W., and Prkachin, K. M. (1990). The expression and perception of facial emotion in alexithymia: a pilot study. *Psychosom. Med.* 52, 199–210. doi: 10.1097/00006842-199003000-00007

Mehu, M., Mortillaro, M., Bänziger, T., and Scherer, K. R. (2012). Reliable facial muscle activation enhances recognizability and credibility of emotional expression. *Emotion* 12:701. doi: 10.1037/a0026717

Motley, M. T., and Camden, C. T. (1988). Facial expression of emotion: a comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting. *West. J. Commun.* 52, 1–22. doi: 10.1080/10570318809389622

Pantic, M., Valstar, M., Rademaker, R., and Maat, L. (2005). "Web-based database for facial expression analysis," in *2005 IEEE International Conference on Multimedia and Expo* (IEEE), 5. doi: 10.1109/ICME.2005.1521424

Park, S., Lee, K., Lim, J.-A., Ko, H., Kim, T., Lee, J.-I., et al. (2020). Differences in facial expressions between spontaneous and posed smiles: Automated method by action units and three-dimensional facial landmarks. *Sensors* 20:1199. doi: 10.3390/s20041199

Petridis, S., Martinez, B., and Pantic, M. (2013). The Mahnob laughter database. *Image Vision Comput.* 31, 186–202. doi: 10.1016/j.imavis.2012.08.014

Pfister, T., Li, X., Zhao, G., and Pietikäinen, M. (2011). "Differentiating spontaneous from posed facial expressions within a generic facial expression recognition framework," in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)* (Barcelona: IEEE), 868–875. doi: 10.1109/ICCVW.2011.6130343

Racoviţeanu, A., Florea, C., Badea, M., and Vertan, C. (2019). "Spontaneous emotion detection by combined learned and fixed descriptors," in *2019 International Symposium on Signals, Circuits and Systems (ISSCS)* (Iasi: IEEE), 1–4.

Roy, S., Roy, C., Éthier-Majcher, C., Fortin, I., Belin, P., and Gosselin, F. (2007). *Stoic: A Database of Dynamic and Static Faces Expressing Highly Recognizable Emotions*. Montréal, QC: Université De Montréal.

Saito, C., Masai, K., and Sugimoto, M. (2020). "Classification of spontaneous and posed smiles by photo-reflective sensors embedded with smart eyewear," in *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction* (Sydney), 45–52. doi: 10.1145/3374920.3374936

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

Saxen, F., Werner, P., and Al-Hamadi, A. (2017). "Real vs. fake emotion challenge: Learning to rank authenticity from facial activity descriptors," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 3073–3078. doi: 10.1109/ICCVW.2017.363

Saxena, A., Khanna, A., and Gupta, D. (2020). Emotion recognition and detection methods: a comprehensive survey. *J. Artif. Intell. Syst.* 2, 53–79. doi: 10.33969/AIS.2020.21005

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

Simons, G., Ellgring, H., and Smith Pasqualini, M. (2003). Disturbance of spontaneous and posed facial expressions in Parkinson's disease. *Cogn. Emot.* 17, 759–778. doi: 10.1080/02699930302280

Smith, M. C., Smith, M. K., and Ellgring, H. (1996). Spontaneous and posed facial expression in Parkinson's disease. *J. Int. Neuropsychol. Soc.* 2, 383–391. doi: 10.1017/S1355617700001454

Tavakolian, M., Cruces, C. G. B., and Hadid, A. (2019). "Learning to detect genuine versus posed pain from facial expressions using residual generative adversarial networks," in *2019 14th IEEE International Conference on Automatic Face Gesture Recognition (FG 2019)* (Lille: IEEE), 1–8. doi: 10.1109/FG.2019.8756540

Valstar, M., and Pantic, M. (2010). "Induced disgust, happiness and surprise: an addition to the mmi facial expression database," in *Proceedings of 3rd International Workshop on EMOTION (Satellite of LREC): Corpora for Research on Emotion and Affect* (Paris), 65.

Valstar, M. F., Gunes, H., and Pantic, M. (2007). "How to distinguish posed from spontaneous smiles using geometric features," in *Proceedings of the 9th International Conference on Multimodal Interfaces* (Nagoya), 38–45. doi: 10.1145/1322192.1322202

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* (Banff), 162–170. doi: 10.1145/1180995.1181031

Van Der Geld, P., Oosterveld, P., Bergé, S. J., and Kuijpers-Jagtman, A. M. (2008). Tooth display and lip position during spontaneous and posed smiling in adults. *Acta Odontol. Scand.* 66, 207–213. doi: 10.1080/00016350802060617

Walter, S., Gruss, S., Ehleiter, H., Tan, J., Traue, H. C., Werner, P., et al. (2013). "The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system," in *2013 IEEE International Conference on Cybernetics (CYBCO)* (Lausanne: IEEE), 128–131. doi: 10.1109/CYBConf.2013.6617456

Wan, J., Escalera, S., Anbarjafari, G., Jair Escalante, H., Baró, X., Guyon, I., et al. (2017). "Results and analysis of chalearn lap multi-modal isolated and continuous gesture recognition, and real versus fake expressed emotions challenges," in *Proceedings of the IEEE International Conference on Computer Vision Workshops* (Venice), 3189–3197. doi: 10.1109/ICCVW.2017.377

Wang, S., Liu, Z., Lv, S., Lv, Y., Wu, G., Peng, P., et al. (2010). A natural visible and infrared facial expression database for expression recognition and emotion inference. *IEEE Trans. Multimed.* 12, 682–691. doi: 10.1109/TMM.2010.2060716

Wang, S., Wu, C., He, M., Wang, J., and Ji, Q. (2015). Posed and spontaneous expression recognition through modeling their spatial patterns. *Mach. Vision Appl.* 26, 219–231. doi: 10.1007/s00138-015-0657-2

Wang, S., Wu, C., and Ji, Q. (2016). Capturing global spatial patterns for distinguishing posed and spontaneous expressions. *Comput. Vision Image Understand.* 147, 69–76. doi: 10.1016/j.cviu.2015.08.007

Wang, S., Zheng, Z., Yin, S., Yang, J., and Ji, Q. (2019). A novel dynamic model capturing spatial and temporal patterns for facial expression analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 2082–2095. doi: 10.1109/TPAMI.2019.2911937

Wu, P., Liu, H., and Zhang, X. (2014). "Spontaneous versus posed smile recognition using discriminative local spatial-temporal descriptors," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Florence: IEEE), 1240–1244. doi: 10.1109/ICASSP.2014.6853795

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)* (Hong Kong: IEEE), 769–774. doi: 10.1109/ICME.2017.8019373

Zhang, L., Tjondronegoro, D., and Chandran, V. (2011). "Geometry vs. appearance for discriminating between posed and spontaneous emotions," in *International Conference on Neural Information Processing* (Shanghai: Springer), 431–440. doi: 10.1007/978-3-642-24965-5_49

Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., et al. (2013). "A high-resolution spontaneous 3D dynamic facial expression database," in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (Shanghai), 1–6. doi: 10.1109/FG.2013.6553788

Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., et al. (2014). BP4D-spontaneous: a high-resolution spontaneous 3D dynamic facial expression database. *Image Vision Comput.* 32, 692–706. doi: 10.1016/j.imavis.2014.06.002

Zhao, G., Huang, X., Taini, M., Li, S. Z., and PietikäInen, M. (2011). Facial expression recognition from near-infrared videos. *Image Vision Comput.* 29, 607–619. doi: 10.1016/j.imavis.2011.07.002

Zuckerman, M., Hall, J. A., DeFrank, R. S., and Rosenthal, R. (1976). Encoding and decoding of spontaneous and posed facial expressions. *J. Pers. Soc. Psychol.* 34:966. doi: 10.1037/0022-3514.34.5.966

Conflict of Interest: 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 Neuroscience

ORIGINAL RESEARCH
published: 23 April 2021
doi: 10.3389/fnins.2021.609760

Deep Neural Networks for Depression Recognition Based on 2D and 3D Facial Expressions Under Emotional Stimulus Tasks

Weitong Guo[1,2,3,4], *Hongwu Yang*[2,4], *Zhenyu Liu*[1,3]*, *Yaping Xu*[1,2] and *Bin Hu*[1,3]*

[1] *School of Information Science Engineering, Lanzhou University, Lanzhou, China,* [2] *School of Educational Technology, Northwest Normal University, Lanzhou, China,* [3] *Gansu Provincial Key Laboratory of Wearable Computing, Lanzhou, China,* [4] *National and Provincial Joint Engineering Laboratory of Learning Analysis Technology in Online Education, Lanzhou, China*

The proportion of individuals with depression has rapidly increased along with the growth of the global population. Depression has been the currently most prevalent mental health disorder. An effective depression recognition system is especially crucial for the early detection of potential depression risk. A depression-related dataset is also critical while evaluating the system for depression or potential depression risk detection. Due to the sensitive nature of clinical data, availability and scale of such datasets are scarce. To our knowledge, there are few extensively practical depression datasets for the Chinese population. In this study, we first create a large-scale dataset by asking subjects to perform five mood-elicitation tasks. After each task, subjects' audio and video are collected, including 3D information (depth information) of facial expressions via a Kinect. The constructed dataset is from a real environment, i.e., several psychiatric hospitals, and has a specific scale. Then we propose a novel approach for potential depression risk recognition based on two kinds of different deep belief network (DBN) models. One model extracts 2D appearance features from facial images collected by an optical camera, while the other model extracts 3D dynamic features from 3D facial points collected by a Kinect. The final decision result comes from the combination of the two models. Finally, we evaluate all proposed deep models on our built dataset. The experimental results demonstrate that (1) our proposed method is able to identify patients with potential depression risk; (2) the recognition performance of combined 2D and 3D features model outperforms using either 2D or 3D features model only; (3) the performance of depression recognition is higher in the positive and negative emotional stimulus, and females' recognition rate is generally higher than that for males. Meanwhile, we compare the performance with other methods on the same dataset. The experimental results show that our integrated 2D and 3D features DBN is more reasonable and universal than other methods, and the experimental paradigm designed for depression is reasonable and practical.

Keywords: deep belief networks, facial expression, 3D deep information, affective rating system, depression recognition

OPEN ACCESS

Edited by:
Huiyu Zhou,
University of Leicester,
United Kingdom

Reviewed by:
Zheheng Jiang,
Lancaster University, United Kingdom
Jialin Lyu,
University of Leicester,
United Kingdom

****Correspondence:***
Bin Hu
bh@lzu.edu.cn
Zhenyu Liu
liuzhenyu@lzu.edu.cn

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

Received: *24 September 2020*
Accepted: *08 March 2021*
Published: *23 April 2021*

Citation:
Guo W, Yang H, Liu Z, Xu Y and Hu B
(2021) Deep Neural Networks for
Depression Recognition Based on 2D
and 3D Facial Expressions Under
Emotional Stimulus Tasks.
Front. Neurosci. 15:609760.
doi: 10.3389/fnins.2021.609760

1. INTRODUCTION

According to the World Health Organization (WHO), more than 350 million people of all ages suffer from depression disorder globally (Reddy, 2012). Depression (depressive disorder or clinical depression) is one of the most severe but prevalent mental disorders globally. Depression can induce severe impairments that interfere with or limit one's ability to conduct major life activities for at least 2 weeks. During at least 2 weeks, there is either a depressed mood or a loss of interest or pleasure, as well as at least four other symptoms that reflect a change in functioning, such as problems with sleep, eating, energy, concentration, self-image, or recurrent thoughts of death or suicide. Depression can occur at any age, and cases in children and adolescents have been reported[1]. Because of its harmfulness and recent prevalence, depression has drawn increasing attention from many related communities.

Although it is severe, depression is curable through medication, psychotherapy, and other clinical methods (National Collaborating Centre for Mental Health, 2010). The earlier that treatment can begin, the more effective it is. Thus, the early detection of depression is critical to controlling it at an initial stage and reducing the social and economic burden related to this disease. Traditional diagnosis approaches for depression are mostly based on patients self-reporting in clinic interviews, behaviors reported by relatives or friends, and questionnaires, such as the Patient Health Questionnaire (PHQ-9) (Kroenke and Spitzer, 2002) and the Beck Depression Inventory (BDI-II) (McPherson and Martin, 2010). However, all of them utilize subjective ratings, and their results tend to be inconsistent at different times or in different environments. During the diagnosis, several clinical experts must be involved to obtain a relatively objective assessment. As the number of depressed patients increases, early-stage diagnosis and re-assessments for tracking treatment effects are often limited and time consuming. Therefore, machine learning-based automatic potential depression risk detection or depression recognition is expected to facilitate objective and fast diagnosis to ensure excellent clinical care quality and fundamentally reduce potential harm in real life.

Under the influence of depression, behavior disorder-based signals for depression recognition are increasingly extensive, such as voices (Ooi et al., 2013; Yang et al., 2013; Nicholas et al., 2015; Jiang et al., 2017), facial expressions (Schwartz et al., 1976; Babette et al., 2005), gestures (Alghowinem et al., 2018), gaits (Michalak et al., 2009; Demakakos et al., 2015), and eye movements (Winograd-Gurvich et al., 2006; Alghowinem et al., 2013; Carvalho et al., 2015). This work focuses on using facial expressions to recognize patients with potential depression risk. The research on depression based on facial expression essentially utilize video or images (Gupta et al., 2014; Alghowinem, 2015; Pampouchidou et al., 2015, 2016a; Bhatia et al., 2017). To be more precise, the interests are localized on images, facial landmark points (Stratou et al., 2014; Morency et al., 2015; Nasir et al., 2016; Pampouchidou et al., 2016b), and/or facial action units (AUs) (Cohn et al., 2009; McIntyre et al., 2009; Williamson et al., 2014). However, the methods that adopt image analysis (the essence of the video-based method are still images analysis where videos are converted into images) are affected by environmental factors and instrument parameters, such as illumination, angle, skin color, and resolution power. If these factors are not addressed appropriately, the recognition performance will be affected. Several researchers (Gong et al., 2009; Zhao et al., 2010) proposed using in-depth information captured from 3D sensors, which is relatively illumination, angle, and skin color invariant. However, 3D points of information can lose the texture features of facial expression. Therefore, the fusion of 2D with 3D data can make up for each other to address these issues.

Depression recognition typically comprises two steps: feature extraction and recognition (depression or not/ depression severity). The quality of feature extraction directly affects the result of recognition. Conventional feature extraction methods for depression facial expression utilize geometric features, appearance features, and dynamics. These methods extract the displacement of facial edges, corners, coordinates (McIntyre et al., 2010; Bhatia et al., 2017), mean squared distance of all mouth landmarks to the mouth centroid (Gupta et al., 2014), and displacement from the mid-horizontal axis to depict the changes and intensity of basic expressions (Bhatia, 2016). The local binary pattern (LBP), LBP-TOP (Joshi et al., 2012), local Gabor binary pattern (LGBP-TOP) (Sidorov and Minker, 2014), local curvelet binary pattern (LCBP-TOP) (Pampouchidou et al., 2015), and LPQ from three orthogonal planes (LPQ-TOP) (Wen et al., 2015) extracted describe the texture changes in the facial region. Histogram of optical flow (Gupta et al., 2014), motion history histogram (MHH) (Meng et al., 2013), and space-time interest points (STIPs) (He et al., 2015) are extracted to describe the facial motion. These results indicated that depressed people display a lower performance when responding to positive and negative emotional content. Nevertheless, all those mentioned approaches are hand-crafted feature descriptors designed based on tremendous professional knowledge, and image processing is also complicated for hand-crafted features. However, our cognition of depression remains insufficient. Such features probably yield segmented representations of facial expressions and are insufficiently discriminative. Simultaneously, the dynamics are extracted from a video, which involves the effect of the environmental factors mentioned above. On the other hand, the time window is used to extract motion features (Pampouchidou et al., 2016a; He et al., 2018). The reported window lengths are 60 frames, 20 frames, 5 frames, or even 300 frames. However, the optimal window length cannot be determined because there are significant variations over time in the facial expression according to the particular person and experimental device.

In recent years, deep learning techniques have prevailed in audio- and video-based applications, especially in visual information processing (Girshick et al., 2014). The purpose of this study is to identify the patients at risk of depression. The selected subjects are outpatients, and the evaluated depression degree is moderate. Many samples with depression risk and

[1] Available at: http://medlineplus.gov/depress.html.

the normal control group had no noticeable expression changes in some stimulation tasks. Therefore, we chose the generative model deep belief network (DBN). The DBN-based deep learning method can hierarchically learn good representation from original data; thus, the learned facial features should be more discriminative than hand-crafted features for depression recognition. Long short-term memory (LSTM) is an effective and scalable model for learning problems related to sequential data and can capture long-term temporal dependencies. Facial expression is a dynamic process of continuous change, and it is a time-series signal on a timeline. Then facial expression motion is captured by LSTM used on the entire timeline.

The availability of clinical data is critical for the evaluation of methods for depression recognition. Because of the sensitivity of clinical data and privacy reasons, datasets for depression research are neither extensive nor free. It is why most research groups resort to generating their datasets. The current datasets are as follows: Pittsburgh, BlackDog, DAIC-WOZ, AVEC, ORI, ORYGEN, CHI-MEI, and EMORY, but only three of which are available. AVEC is the only fully public dataset available for free download, DAIC-WOZ is partly available, while Pittsburgh is also available, but not accessible now. The rest depression-related datasets are proprietary, and the corresponding research results are few. The securable datasets above provide the third-parties visual and audio features. Only AVEC discloses complete video recordings. However, these datasets are collected from non-Chinese subjects, which differ from Chinese subjects in terms of emotional expressiveness due to different cultural backgrounds. Thus, we used a structured experimental paradigm to construct a depression database specifically for Chinese subjects in conjunction with relevant psychiatric hospitals. To the best of our knowledge, the database we have established is the only database with complete data, a reasonable structure, and the largest number of subjects in China. Our dataset includes complete video recordings from typical webcam and microphone, and 3D 1,347 facial points scan from deep camera Kinect (Leyvand et al., 2011). Not only does a Kinect detect the human face, but it also provides real-time access to over 1,000 facial points in the 3D space irrespective of the color of the skin or the surrounding environment, illumination, or distance from the camera.

This paper builds on our previous work (Guo et al., 2019) by adding 2D static image information and 3D facial point motion information to identify depression, and it is a further improvement and summary of the original work. We build two different deep networks, respectively, one of which extracts static appearance feature using 2D images based on DBN, and the other learns the facial motion via 3D facial landmark points and facial AUs using DBN-LSTM. The two kinds of deep networks are then integrated by joint fine-tuning, which can further improve the overall performance. Therefore, our main contributions in this paper can be summarized as follows:

1. We designed a reasonable and effective experimental paradigm, collected diversified data and three kinds of samples (normal population, outpatients, and inpatients) combined with specialized hospitals, and constructed a large-scale dataset for depression analysis.
2. The two deep networks proposed can extract appearance features from 2D images and motion features from 3D facial landmark points. The integrated networks can achieve the fusion of static and dynamic features, which can improve recognition performance.
3. We have proved qualitatively and quantitatively that depressive prone groups show significant differences from healthy groups under positive and negative stimuli.

The following section briefly describes the related works on depression recognition based on facial expression. In section 3, we introduce the proposed depression recognition network structure. Dataset creation, experimental setting, results, and analysis are reported in section 4. Finally, some discussions and future works are provided in section 5.

2. RELATED WORK

2.1. Depression Recognition Based on Machine Learning

Machine learning tools for depression detection have access to the same streams of information that a clinician utilizes for diagnosis. For example, the variation of facial expression, gesture, voice, and language should occur in communication modality. Reduced emotional expression variability is commonly found in depression and connected with deficits in expression positive and negative emotion (Rottenberg et al., 2005). In the following, we briefly summarize some excellent research results on identifying depression from visual cues.

Wang et al. (2008) extracted geometric features from 28 regions formed by 58 2D facial landmarks to characterize facial expression changes. Probabilistic classifiers were employed to propagate the probabilities frame by frame and create a probabilistic facial expression profile. The results indicate that depressed patients exhibit different trends of facial expressions than healthy controls. Meng et al. (2013) employed motion history histogram (MHH) to capture motion information of facial expression. Local binary patterns (LBP) and edge orientation histogram (EOH) features were then extracted, and partial least square (PLS) was finally applied for prediction. These features were extracted from images. Nasir et al. (2016) employed perceptually motivated distance and area features obtained from facial landmarks to detect depression. The window-based representation of features was used to capture large-scale temporal contexts results. Anis et al. (2018) developed an interpretable method of measuring depression severity. Barycentric coordinates of facial landmarks and rotation matrix of 3D head motion were used to extract kinematic features, and a multi-class SVM was used to classify the depression severity.

The methods mentioned above are based on traditional machine learning methods to extract hand-crafted facial expression feature descriptors for depression analysis. Some studies have also utilized deep learning to extract high-level semantic features of facial expressions from raw video recordings for automatic depression detection. Jan et al. (2018) utilized convolutional neural networks (CNN) to extract many different visual primitive features from the facial expression frames, while feature dynamic history histogram (FDHH) was employed to capture the temporal movement on the features. Zhou et al. (2020) presented a DCNN regression model with a GAP layer for depression severity recognition from facial images. Different face regions were modeled, and then these models were combined to improve the overall recognition performance. The results indicated that the salient regions for patients with different depression levels were usually around the eyes and forehead. Melo et al. (2019) used two 3D CNNs to model the spatiotemporal dependencies in global and local facial regions captured in a video, and then combine the global and local 3D CNNs to improve the performance. The CNN-based method mentioned above requires a large amount of data to train the model. Once the amount of data is small, it is easy to fall into overfitting. By comparing the data volume of existing studies with our method, we found that the state-of-the-art research used about 4,350 min of video-based publicly available datasets, while the amount of video data we used was only about 2,080 min. Existing studies have shown that the generation model has a better classification effect than the discriminant model in low samples (Ng and Jordan, 2002). So, we finally choose to use the DBN model.

2.2. DBN

The DBN (Hinton et al., 2006) is a generative model that uses multiple layers of feature-detecting neurons. It can learn hierarchical representation from raw input data and can be effectively built by stacking a restricted Boltzmann machine (RBM) (Fischer and Igel, 2012) layer-by-layer and greedily training it. In our study, the Gaussian–Bernoulli RBM is adopted to use real-valued visible units to train the first layer of the DBN; binary hidden units are used for training the higher layers. For a Gaussian–Bernoulli RBM, the energy function of a joint configuration is given as Equation (1).

$$\mathrm{E}(\mathrm{V},\mathrm{H}) = \frac{1}{2\sigma^2}\sum_{i=1}^{m}\frac{(v_i-a_i)^2}{2} - \frac{1}{\sigma^2}\left(\sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij}v_ih_j + \sum_{j=1}^{m} b_jh_j\right) \tag{1}$$

where $a \in R^D$ and $b \in R$ are the biases for visible and hidden units, respectively. $w_{ij} \in R$ is the weight between the visible unit i and the hidden unit j, while m and n are the numbers of visible and hidden units, respectively. σ is a hyper-parameter. As there are no connections between units in the same layer, the conditional probability distributions are given by Equations (2) and (3).

$$\mathrm{P}\left(h_j = 1 \mid v\right) = \mathrm{sigmoid}\left(\frac{1}{\sigma^2}\left(\sum_{i=1}^{m} w_{ij}h_j + b_j\right)\right) \tag{2}$$

$$\mathrm{P}\left(v_i \mid h\right) = \mathcal{N}\left(a_i + \sum_{j=1}^{n} w_{ij}h_j, \sigma^2\right) \tag{3}$$

where $\mathcal{N}(\mu, v)$ is a Gaussian function with mean μ and variance v. (w, a, b) are the parameters of the RBM and are learned using contrastive divergence. The generated features are the posteriors of the hidden units in the case of given visible units. Finally, the top output values are classified using sigmoid activation and the stochastic gradient descent method is used to train the deep networks.

2.3. LSTM

The LSTM block has a memory cell that stores information with long-term dependencies. We use an LSTM with a conventional structure, as shown in **Figure 1**.

In **Figure 1**, $x^{(t)}$ is the input data in time step t (the current frame), $h^{(t-1)}$ is the hidden unit in time step $t-1$ (the previous

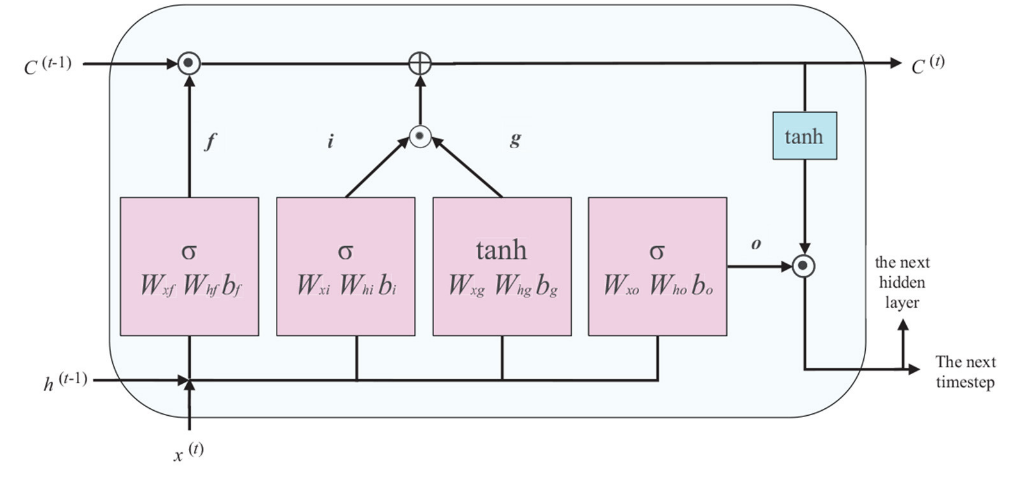

$C^{(t-1)}$
$C^{(t)}$
f
i
g
tanh
σ W_{xf} W_{hf} b_f
σ W_{xi} W_{hi} b_i
tanh W_{xg} W_{hg} b_g
σ W_{xo} W_{ho} b_o
o
the next hidden layer
The next timestep
$h^{(t-1)}$
$x^{(t)}$

FIGURE 1 | Deployment structure of long short-term memory (LSTM) unit.

frame), and $C^{(t-1)}$ represents the cell status of the previous time step, which is modified to obtain the cell status of the current timestep $C^{(t)}$. The flow of information in the LSTM is controlled by the computing unit described in **Figure 1**, namely the forget, input, and output gates. The specific process is described by the following equations:

forget gate f_t: control the retention and removal of features.

$$f_t = \text{sigmoid}\left(W_{xf}x^{(t)} + W_{hf}h^{(t-1)} + b_f\right) \tag{4}$$

input gate i_t: update cell status with input node g_t.

$$\begin{aligned} i_t &= \text{sigmoid}\left(W_{xi}x^{(t)} + W_{hi}h^{(t-1)} + b_i\right) \\ g_t &= \tanh\left(W_{xg}x^{(t)} + W_{hg}h^{(t-1)} + b_g\right) \\ C_t &= \left(C^{(t+1)} \odot f_i\right) \oplus \left(i_t \odot g_t\right) \end{aligned} \tag{5}$$

output gate O_t: update the value of a hidden unit.

$$\begin{aligned} O_t &= \text{sigmoid}\left(W_{xo}x^{(t)} + W_{ho}h^{(t-1)} + b_o\right) \\ h_t &= O_t \odot \tanh\left(C^{(t)}\right) \end{aligned} \tag{6}$$

where the weight matrix subscripts have the obvious meaning. For example, W_{hf} is the hidden-forget gate matrix, and W_{xi} is the input-input gate matrix. The bias terms b, the subscripts of f,i,g, and o, denote the corresponding door's bias.

2.4. Problem Setup

We find that the various kinds of effective methods proposed are based on 2D images (video is split into images) and 2D landmark point data (extracted from 2D images) by a survey of the current research on depression based on visual cues. The main limitations of 2D image-based analysis are problems associated with large variations in pose, illumination, angle, skin color, and resolution power. Nevertheless, depth information captured from 3D sensors is relatively posed and illumination invariant. Inspired by the idea of Aly et al. (2016), the fusion of 2D with 3D data can address these issues and cover the shortage of each other that 3D landmark points miss texture feature. Each expression can be decomposed into a set of semantic AUs, which exhibit in different facial areas and at different times with different intensities. Therefore, the dataset we build contains both 2D video and 3D landmark points and AUs information. In the paper, we propose a novel approach for depressive prone patients recognition based on two kinds of different DBN models combination, one of which extracts 2D appearance features from facial images collected by optical cameras, the other learns the facial motion from 3D facial points and facial AUs collected by a Kinect. The final decision result comes from the combination of the two networks. Finally, we evaluate all proposed deep models in our built dataset and analyze three aspects: gender, stimulus task, and affective valence.

3. PROPOSED APPROACH

3.1. The Framework of Deep Neural Networks-based Depression Recognition

We utilize the DBN and LSTM to potential depression risk recognition. We build two different deep networks: 2D static appearance deep network (2D-SADN), which is used to extract the static appearance features from images based on DBN. In other words, the network only focused on the analysis of appearance from static facial pictures in which a single image was used as input to the network and the network structure did not encode temporal information. 3D dynamic geometry deep network (3D-DGDN) based on combined of DBNs and LSTM, which capture the dynamic geometry features of 3D facial landmark points and AUs from Kinect. Expressions are inherently dynamic events consisting of onset, apex, and offset phases (Liu et al., 2014). Therefore, in the second network, we took the facial contour map composed of facial landmark points as input and used the position offset of the three-dimensional coordinate value on the time axis to obtain motion information. Finally, the two networks are integrated to improve the recognition performance. The overview of the proposed approach is shown in **Figure 2**.

3.1.1. The Structure of DBN Model

The designed basic DBN network is composed of four RBMs, as shown in **Figure 3**. First, Gibbs sampling and contrastive divergence are adopted to train RBM to maximize $\mathbb{E}_{V\sim p_{\text{data}}} \log p(v)$. The RBM parameter defines the parameters of the first layer of the DBN. Then, the second RBM is trained to approximately maximize $\mathbb{E}_{V\sim p_{\text{data}}} \mathbb{E}_{h^{(1)}\sim p^{(1)}(h^{(1)}|v)} \log p^{(2)}\left(h^{(1)}\right)$, where $p^{(1)}$ is the probability distribution represented by the first RBM, and $p^{(2)}$ is the probability distribution represented by the second RBM. That is, the second RBM is trained to simulate the distribution defined by the hidden unit sampling of the first RBM, which is driven by the input data. This process is repeated four times to add four hidden layers to the DBN, and each new RBM models the samples of the previous RBM. Each RBM defines another layer of DBN. Top-down fine-tuning is used to generate weights to guide the determination of the DBN model. At the top two levels, the weights are linked together so that the output of the lower level will provide a correlation to the top level, which will then link it to its associative memory. DBN can adjust the discriminant performance by using the labeled data and BP algorithm.

3.1.2. Learning the Static Appearance Deep Network

In the 2D-SADN, we train a DBN as shown above with four layers by oneself with three channels, and then average the predicted values of each channel. The result is the final predicted value, as shown in **Figure 4**.

3.1.3. Learning the Dynamic Geometry Deep Network

In the 3D dynamic geometry deep network (3D-DGDN), we build four different DBN models based on our designed basic DBN network, as shown in **Figure 5**. **Figure 5A** is a four-hidden layer DBN with facial points, named 4DBN; **Figures 5B,C** shows

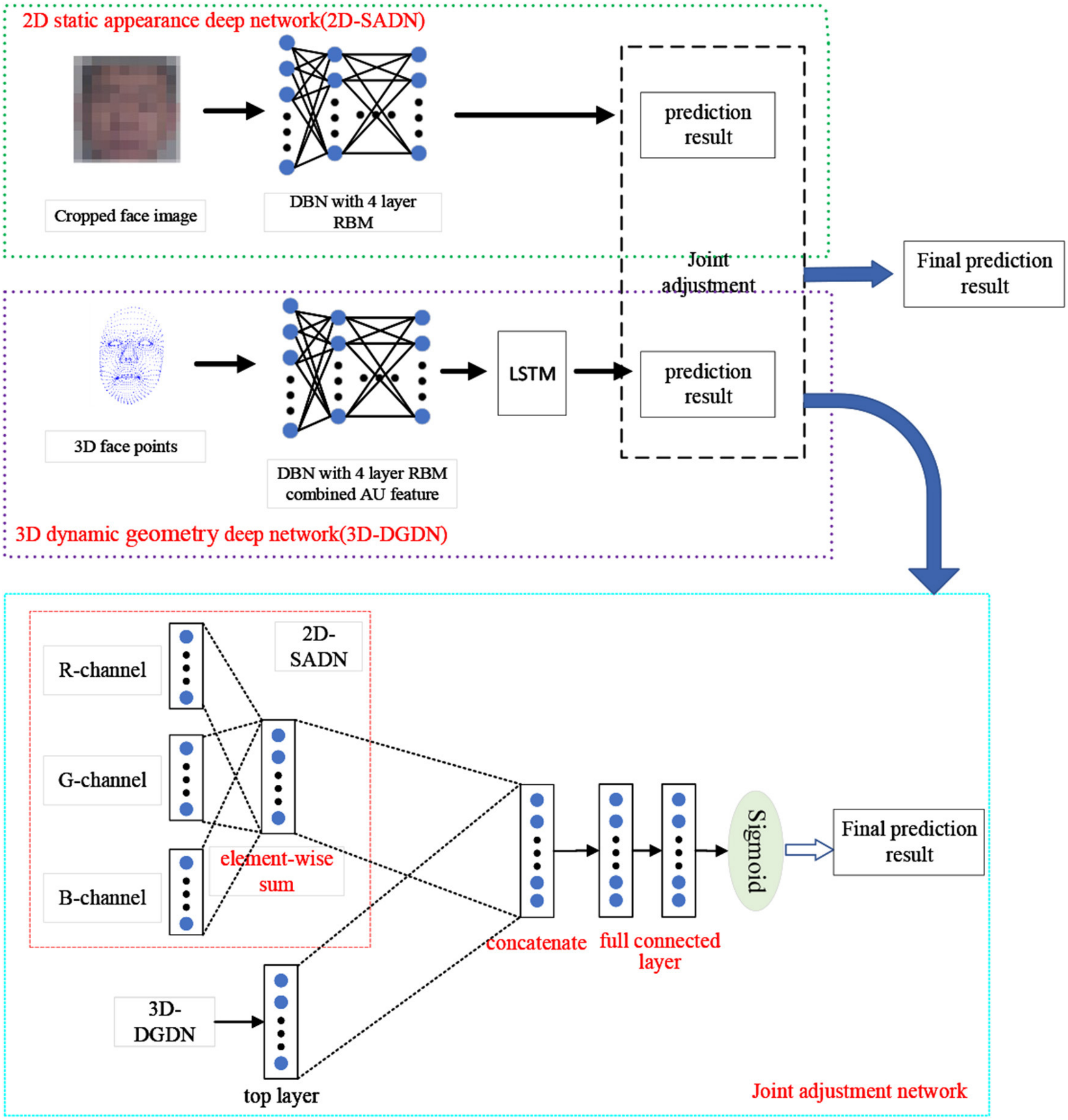


FIGURE 2 | The framework of proposed approach.

four hidden layer DBNs with facial points and AU, named AU-4DBN and 4DBN-AU. **Figure 5D** shows a four hidden layer DBNs with facial points and AU followed by a LSTM, named AU-4DBN-LSTM. In the meantime, we build a five-hidden-layer DBN using facial points as the input and find that the accuracy rate of the four-hidden layer-DBN is almost the highest in all of the stimulus tasks, therefore a four-hidden-layer network is used as the basic DBN structure and AU-4DBN-LSTM stands for 3D-DGDN. The details are as follows.

- *4DBN* is a four-hidden layer DBN only using facial points as the input, as shown **Figure 5A**.
- *4DBN-AU* is a four-hidden layer DBN based on 4DBN that uses AU and facial points as the input, as shown **Figure 5B**.
- *AU-4DBN* is a four-hidden-layer DBN with AU added at the penultimate layer, which is used as the input of stacking an extra RBM on the top, as shown in **Figure 5C**.
- *AU-4DBN-LSTM* is based on the AU-4DBN model to add the LSTM. That is, the output of the RBM on the top of AU-4DBN is used as the input to the first layer of the LSTM (Greff et al., 2016), which has two layers, as shown in **Figure 5D**.

3.1.4. Joint Fine-turning Method

The strategy of adding the corresponding position elements was used to connect the activation values of the top hidden layers of three channels to obtain the feature vector in the 2D-SADN. We then concatenated the feature vector and the activation values of

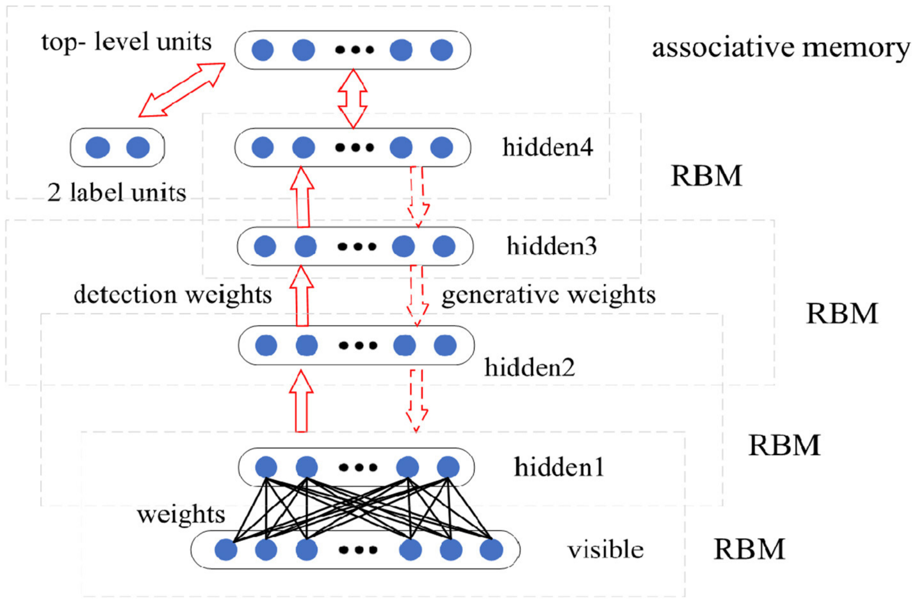


FIGURE 3 | The structure of designed deep belief network (DBN) model.

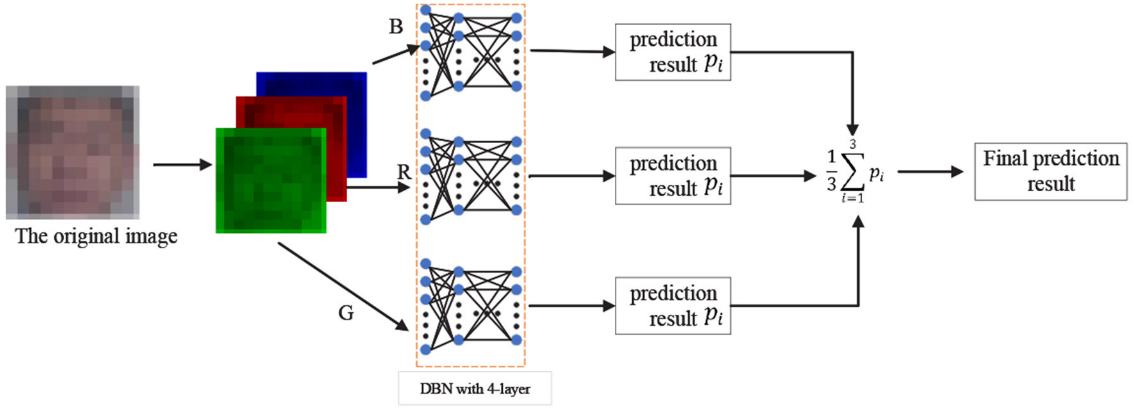


FIGURE 4 | 2D static appearance deep network.

the top hidden layers of 3D-DGDN. Finally, the concatenated values are used for inputs to a fully connected network with a sigmoid activation, as shown in **Figure 2**. Consequently, we integrate the network using a linear weighted sum of the loss function based on Jung et al. (2015), defined as follows:

$$L_{fusion} = \lambda_1 L_{2D-SADN} + \lambda_2 L_{3D-DGDN} + \lambda_3 L_{2D-3D} \quad (7)$$

where L is cross entropy loss function, and $L_{2D-SADN}$, $L_{3D-DGDN}$, and L_{2D-3D} are computed by 2D-SADN, 3D-DGDN, and the both, respectively. λ_1, λ_2, and λ_3 are turning parameters. In order to fully utilize the capabilities of the two models, we set the value of λ_1, λ_2, and λ_3 to 1, 1, and 0.1, respectively. Cross entropy loss function is defined as follows:

$$L_{2D-SADN} = -\sum_{i=1}^{n} y^{(i)} \log \hat{y}^{(i)}_{2D-SADN} + \left(1 - y^{(i)}\right) \log \left(1 - \hat{y}^{(i)}_{2D-SADN}\right) \quad (8)$$

$$L_{3D-DGDN} = -\sum_{i=1}^{n} y^{(i)} \log \widehat{y}^{(i)}_{3D-DGDN} + \left(1 - y^{(i)}\right) \log \left(1 - \widehat{y}^{(i)}_{3D-DGDN}\right) \quad (9)$$

$$L_{2D-3D} = -\sum_{i=1}^{n} y^{(i)} \log \hat{y}^{(i)}_{2D-3D} + \left(1 - y^{(i)}\right) \log \left(1 - \hat{y}^{(i)}_{2D-3D}\right) \quad (10)$$

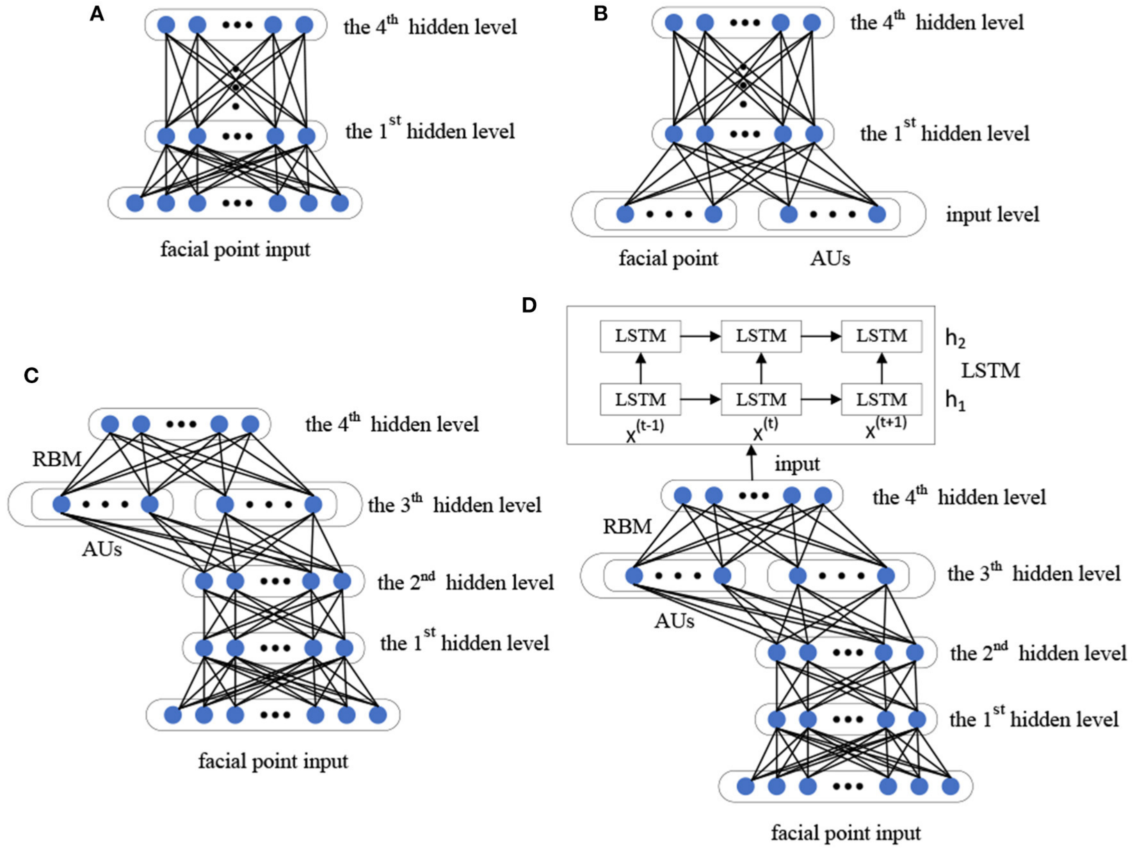


FIGURE 5 | Framework of four different deep belief network (DBN) models: **(A)** 4DBN, **(B)** 4DBN-AU, **(C)** AU-4DBN, **(D)** AU-4DBN-LSTM.

where n is the number of samples and $y^{(i)}$ is the ground-truth label of the ith sample. $\hat{y}^{(i)}_{2D-SADN}$, $\hat{y}^{(i)}_{3D-DGDN}$, and $\hat{y}^{(i)}_{2D-3D}$ are the ith output value of sigmoid activation of 2D-SADN, 3D-DGDN, and the integrated network, respectively. $\hat{y}^{(i)}_{2D-3D}$ is computed by logit values of network 2D-SADN and 3D-DGDN as follows:

$$\hat{y}^{(i)}_{2D-3D} = \sigma\left(l^{(i)}_{2D-SADN} + l^{(i)}_{3D-DGDN}\right) \tag{11}$$

where $l^{(i)}_{2D-SADN}$ and $l^{(i)}_{3D-DGDN}$ are the ith logit values of network 2D-SADN and 3D-DGDN, respectively. $\sigma(\bullet)$ is a sigmoid activation function.

$$l^{(i)}_k = \log\left(\frac{x_i}{1-x_i}\right)_k \quad \forall x_i \in (0,1) \tag{12}$$

where k means network 2D-SADN and 3D-DGDN, and x_i is the ith output value of sigmoid of network k. The final prediction is the index with the maximum value from the output of sigmoid of the integrated network as follows:

$$\hat{P} = \underset{i}{\operatorname{argmax}} \hat{y}^{(i)}_{2D-3D} \tag{13}$$

The paper uses 10-fold cross-validation to evaluate experiments for excluding the differences caused by individuality and over-fitting. Note that 80% of samples from the total participants are used for training, 10% for validation, and the rest of 10% for testing. Each fold includes the data from 42 participants for training, 5 participants for validation, and 5 participants for testing. We use accuracy to evaluate the proposed model performance. Accuracy is computed by the confusion matrix consisting of the number of true positives (TP), true negatives (TN), false positives (FP), and false negative (FN), defined as follows:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

where TP is the number of depression samples predicted to be depressed, TN is the number of healthy samples predicted to be healthy, FP is the number of healthy samples predicted to be depression, and FN is the number of depressed samples predicted to be healthy.

4. EXPERIMENTS

4.1. Depression Data Collection

To effectively obtain depression data, we cooperated with Tianshui Third People's Hospital in Gansu province to collect data. Data collection was accomplished in an isolated, quiet, and soundproof room without electromagnetic interference. Two people were present in the room at the same time: one was the clinician controlling the data collection process, and the other was the participant. The clinician operated one of the two computers and played all the stimulus tasks (film clips, voice responses, text reading, and picture description) sequentially. The stimulus materials were displayed to the participant on the second computer. Participants must evaluate their emotional state before and after completing each stimulus task. Each stimulus task was divided into positive, neutral, and negative stimuli. In order to prevent the stimulus of the previous emotional valence from affecting the next emotional valence stimulus, there is a 1-min break at the end of each material. Moreover, the order of valence stimulation presented to each subject was also different. Audio and video information on the participant was recorded by a webcam, a Kinect camera, and a microphone.

4.1.1. Participants

Every participant is rated by psychiatrists through interviews and questionnaires. The set of questionnaires required to be filled included the International Neuropsychiatric Interview (MINI) and the Patient Health Questionnaire-9 (PHQ-9). PHQ-9 was the main grouping criteria (health control:$<$ 5, patient: $\geq$ 5). PHQ-9 scores are treated as the label. In this experiment, the out-patient sample set included data from 52 males and 52 females; meanwhile, the control group also included data from 52 males and 52 females. Participants were excluded from the health group if they received a Beck depression inventory (BDI) score $>$ 5. The demographic characteristics of all participants are shown in **Table 1**. All participants provided informed consent.

4.1.2. Paradigm Design of Depression Experiments

Depressed individuals have negative self-schema in cognitive processing related to attention control disorder of emotional interference. The phenomena related to emotion include subjective experience, facial expression behavior, individual differences in nervous system response, and emotional response. In order to obtain useful data, we need to choose appropriate emotional induction methods. Using the classic oddball experimental paradigm of psychology (Li, 2014), we designed 5 stimuli tasks of 3 emotional valences to induce behavioral differences between healthy and depressed individuals. The tasks included:

1. *Watching film clips:* Three short film segments around 90 s each, one positive, another neutral, and the other negative, were disciplinarily presented. Participants were asked to watch the film clip and then describe their mood. The clips had previously been rated for their affective content (Gross and Levenson, 1995). The positive film clip is excerpted from cartoon "Larva Funny Bugs," the neutral film clip is excerpted from the documentary "Universe Millennium," and the negative film clip is excerpted from the movie "October Siege." The synchronized start of the stimuli with the recording enabled us to draw a correlation between facial activity and the stimuli.
2. *Replying to nine free-response questions:* Each participant was requested to respond to nine specific questions (three positives, three neutrals, and three negatives). These questions are designed based on DSM-IV and other depression scales such as the Hamilton depression rating scale (HDRS)[2]. Questions included, for instance, "what kind of lifestyle do you like?" "discuss a sad childhood memory," and "please evaluate yourself." The answers were synchronized with the facial activity recorded.
3. *Reading three phonetically balanced passages containing affective content:* The participants were presented with a paragraph of text on a computer screen and asked to finish the reading as naturally as possible. There are three reading materials. One of passages contained positive words (e.g., glorious, victory), and the other contained negative words (e.g., heart-broken, pain), which were selected from affective the ontology corpus created by Hongfei Lin[3]. The last one included neutral words (e.g., village, center) selected from the extremum table of affective Chinese words (Gong et al., 2011). The reading and recording commenced synchronously.
4. *Describing pictures:* The picture description section is to present 6 pictures in sequence on the computer screen. The first three pictures are facial expression pictures of three women divided into positive, neutral, and negative, and the last three pictures are three scene pictures divided into positive, neutral, and negative. All pictures were selected from the Chinese Facial Affective Picture System (CFAPS) (Gong et al., 2011). Participants were requested to observe the picture and then describe it. Reporting logs enabled correlations between image presentations and facial activity to be established.

4.1.3. Process of Affective Rating

The Self-Assessment Manikin (SAM) (Lang, 1980) which is an affective rating-scale system using a graphical figure that depicts

TABLE 1 | Demographic characteristic of the out-patients and control group: mean and (standard deviation).

Gender	Category	Number	Age	Education	PHQ-9	BDI
Male	Control group	52	39 (10.8)	11.8 (2.5)	0.8 (2.0)	6.4 (6.4)
	Out-patients	52	34.8 (11.1)	11.2 (3.4)	17.5 (5.6)	26.4 (12.8)
Female	Control group	52	34.7 (10.7)	12.3 (3.2)	0.3 (0.7)	4.7 (5.3)
	Out-patients	52	37.4 (10.4)	10.8 (4.0)	18.3 (5.6)	33.5 (13.2)

[2] Available at: http://ir.dlut.edu.cn/Group.aspx?ID=4.

[3] Available at: http://www.datatang.com/data/43216.

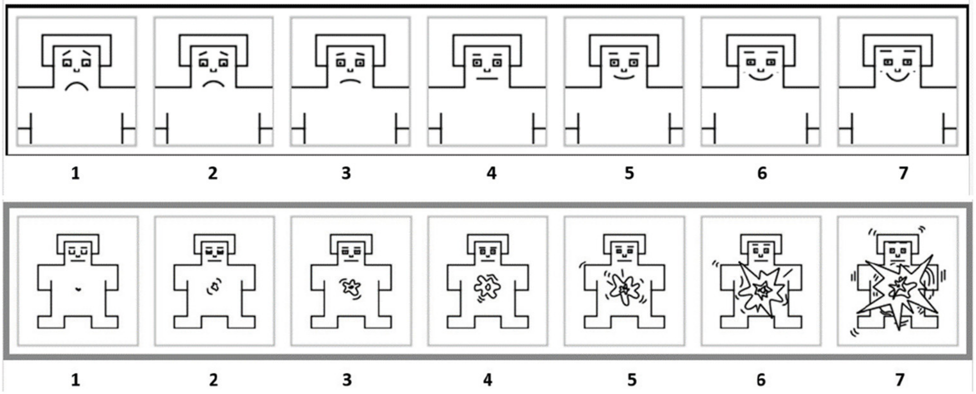


FIGURE 6 | Schematic diagram of affective rating. The top is valence rating. The bottom is arousal rating.

the dimensions of valence (from a smiling figure to a frowning figure) and arousal (from an excited to a relaxed figure), is used to measure a participant's emotion, as shown in **Figure 6**. The affective rating process consists of three parts: emotional pretest, emotion-eliciting tasks, and emotional posttest.

4.1.4. Constructing Depression Dataset

Every participant has to complete five stimulus tasks of three emotional valences in turn, resulting in 15 datasets: three datasets for watching film clips (one positive + one neutral + one negative), three datasets for nine interviews (three positive + three neutral + three negative), three datasets for text readings (one positive + one neutral + one negative), three datasets for expression image descriptions (one positive + one neutral + one negative), and three datasets for scene image descriptions (one positive + one neutral + one negative). The database consists of four folders for each participant, which are voice, video, emotional state, and information log. Fifteen monophonic speech recordings are made in the voice folder. A sampling rate of 44.1 kHz and a sampling depth of 24-bit are used for collecting speech signals. Speech recordings are saved in the uncompressed WAV format. Ambient noise should be lower than 60 dB. There are two types of data in the video folder. One is 15 video recordings of 640 × 480 pixel, 30 fps collected by a webcam, and saved as mp4; the other is 15 recordings obtained by the Kinect, and each recording contains two kinds of data: three-dimensional coordinates of 1,347 facial points and 17 AUs. Every facial point is a 3D point with X, Y, and Z coordinates. **Figure 7** shows that facial contours consist of 1,347 feature points. At present, many AU detection methods use feature point tracking, shape modeling, template matching, and neural network to recognize the AU features of the face. In this paper, the Kinect device automatically recognizes the AU of the face through the built-in API interface and the facial AU detection algorithm. The intensity of each AU in each frame was calculated. The intensity amplitude was between -1 and 1, and the facial expression was directly measured by digital features. Seventeen AUs recorded by Kinect are corresponding to AUs encoded by FACS. AU from Kinect can appear separately or in combination to show different expressions[4]. Fifteen three-dimensional facial points recordings and 15 AUs recordings are saved as CSV format. Participants are assessed for the two dimensions of valence and arousal before and after receiving different emotion-eliciting tasks to obtain 15 evaluation results, which are saved in the emotional state folder. The information related to the subjects is saved in the information log, which are name, gender, age, profession, education background, label, PHQ-9, BDI, and so on.

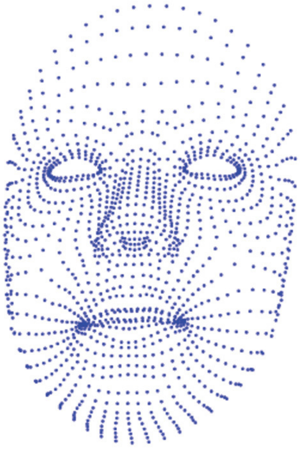


FIGURE 7 | Facial contour that consists of 1,347 vertices generated by Kinect.

4.2. Data Pre-processing

We intercepted the data which the subjects did not speak during the three tasks of watching the film clips, facial expression pictures description, and scene pictures description. The machine learning toolkit Dlib (King, 2009) is used to acquire facial region cropped and aligned according to each video's eye location. All images are resized to 100 × 100. Considering the problem of

[4] Available at: https://blog.csdn.net/u014365862/article/details/48212757.

many frame redundancy in the video clip, we adopt the sampling scheme of taking one frame every 100 frames. Experiments determine the sampling interval according to the number of frames or length of each video and the frequency of facial changes. Finally, we obtained 156,000 facial image data in any given stimulus task. Then, we adopt the same sampling strategy on the Kinect dataset to ensure that the two data are aligned on the time axis. These data were used for training, validation, and testing networks. The original image is then divided into three images by R, G, and B channels and used to train a DBN. Every image is flatted, and then the pixel value of which is normalized for input in 2D-SADN. In 3D-DGDN, three-dimensional coordinates of 1,347 facial points can be considered as one-dimension vector at frame t and defined as $\left[x_1^{(t)}y_1^{(t)}z_1^{(t)}\cdots x_n^{(t)}y_n^{(t)}z_n^{(t)}\right]^T$, where n is the total number of landmark points at frame t. $I-score$ standardization is used to normalize xyz-coordinates as follows:

$$\bar{x}_i^{(t)} = \frac{x_i^{(t)} - \mu}{\sigma} \tag{15}$$

where $x_i^{(t)}$ is x-coordinate of the ith facial landmark point at frame t, μ, and σ is mean value and standard deviation of x-coordinate at frame t, respectively. This process is also applied to $y_i^{(t)}$, $z_i^{(t)}$, and AUs. Finally, these normalized points are concatenated $\left[\bar{x}_1^{(t)}\bar{y}_1^{(t)}\bar{z}_1^{(t)}\cdots\bar{x}_n^{(t)}\bar{y}_n^{(t)}\bar{z}_n^{(t)}\right]^T$ or $\left[\bar{x}_1^{(t)}\bar{y}_1^{(t)}\bar{z}_1^{(t)}\cdots\bar{x}_n^{(t)}\bar{y}_n^{(t)}\bar{z}_n^{(t)}AU_1^{(t)}\cdots AU_{17}^{(t)}\right]^T$ to input the DBN.

4.3. Network Architecture

The DBNs structure of 2D-SADN and 3D-DGDN are similar. The first layer RBM is trained completely unsupervised for all DBNs. The biases and weights are randomly sampled from a normal distribution with $\mu = 0$, $\sigma = 0.01$. They are all updated after a full minibatch. Because the preprocessed data is larger in 2D-SADN, a penalty term is added to the weight and bias updates to obtain sparse representation. λ is fixed as 3, the sparsity parameter of bias is 0.1, and the learning rate is 0.02. The rest RBMs are also trained for 100 epochs using the same value used for training the first level RBM for all DBNs.

The hidden nodes number of every channel DBN is fixed to 8192-4096-2048-502 in 2D-SADN. The number of hidden nodes for 4DBN and 4DBN-AU from the first layer to 4-layer is selected over 3000-2048-1024-128 in 3D-DGDN. However, for AU-4DBN, the penultimate hidden layer with 1,024 nodes similar to 4DBN is then concatenated with AUs, and the resulting input serves as the visible layer of a top-level RBM with 150 hidden nodes (Gaussian–Bernoulli).

Our LSTM has two layers, one with 200 nodes and another with 64 nodes. We initialize the hidden states to zero and then use the current minibatch's final hidden states as the initial hidden state of the subsequent minibatch. The batch size is 50, and the training epoch is 50. The learning rate is set by grid search, and the momentum is 0.9.

The whole system is tested on the TensorFlow deep learning framework with a Xeon(R) CPU E7-4820 v4@2.00 GHz processor, 128 Gigabytes memory, and a Telsa M60 GPU, which can meet our computing needs.

TABLE 2 | The differences in valence and arouse dimension between the healthy and depressed group for the five stimuli with three emotional valences ($P < 0.1$).

Emotional dimension	Tasks	Subjects	Positive	Neutral	Negative
Valence	Film	Health	1.635	0.074	−1.534
		depression	−1.058	−0.036	−1.078
		P-value	**0.043**	2.428	**0.052**
	Question	Health	1.356	0.088	−0.744
		Depression	−0.273	0.333	−0.330
		P-value	0.601	0.565	1.035
	Reading	Health	0.947	0.260	0.829
		Depression	0.273	0.242	0.333
		P-value	0.233	0.137	0.531
	Expression figure	Health	1.084	0.205	−0.938
		Depression	−0.152	−0.198	−0.506
		P-value	**0.073**	0.960	**0.085**
	Scene figure	Health	0.874	0.110	−0.123
		Depression	−0.015	0.061	0.303
		P-value	0.125	1.531	0.211
arousal	Film	Health	1.058	−0.205	1.045
		Depression	−0.635	0.076	0.014
		P-value	**0.072**	0.151	**0.065**
	Question	Health	0.968	0.027	0.109
		Depression	−0.060	−0.030	0.151
		P-value	0.254	**0.096**	0.325
	Reading	Health	−0.810	0.137	0.164
		Depression	−0.182	−0.106	0.076
		P-value	0.222	0.115	0.177
	Expression figure	Health	1.008	0.205	0.233
		Depression	−0.014	0.333	0.106
		P-value	**0.098**	**0.042**	**0.087**
	Scene figure	Health	0.219	0.055	−0.068
		Depression	0.015	−0.091	−0.121
		P-value	0.146	**0.033**	**0.088**

The table's data are the statistical value of the difference between the pretest and posttest of the valence dimension and arouse dimension under different stimuli tasks.
Bold indicates significant difference.

4.4. Experimental Results

4.4.1. Qualitative Analysis of Stimulus Tasks

Previous studies have shown that depressed patients have less positive emotions and more negative emotions than healthy individuals, which indicates that depressed patients show sad when stimulated by positive or negative emotions (Delle-Vigne et al., 2014). Depressed patients will be less sensitive to emotional stimuli from the outside world; that is to say, it is difficult for depressed patients to elicit corresponding emotional feedback (Rottenberg, 2005). The differences in valence and arousal dimensions between the healthy and depressed groups for the five stimuli with three emotional valences were calculated, respectively, as shown in **Table 2**.

From **Table 2**, we can find that the absolute value of the healthy group's valence difference is generally greater than that of the depressed group in all stimulus tasks. The valence difference of the healthy group is basically consistent with emotional valence, which means that positive tasks stimulate joyful emotions, negative tasks stimulate sad emotions for the healthy group (the valence difference in positive stimuli is positive, and the valence difference in negative stimuli is negative), but for the depressed group, both positive and negative stimuli basically arouse sad emotions (the valence difference in positive and negative stimuli is negative). From the *T*-test values, it can be found that there is a significant difference in valence between the healthy group and depressed group under positive and negative stimulation, especially in the stimulation tasks of film clips and characters' facial expressions, as shown in bold.

From **Table 2**, we also can find that the arousal difference of the healthy group is almost greater than that of the depressed group in all stimulus tasks, and the arousal difference of the healthy group is basically positive, which means that the healthy group is more likely to be aroused than the depressed group. From the *T*-test values, it can be found that there is a significant difference in arousal between the healthy group and depressed group under positive and negative stimulation, especially in the stimulation tasks of film clips and characters' facial expressions, as shown in bold.

From **Table 2**, we can draw the following conclusions: positive and negative film clips and facial expression pictures are more likely to inspire significant differences between the healthy and depressed groups than the other three tasks. Moreover, the results reflected from **Table 2** are also consistent with paper (Delle-Vigne et al., 2014) and (Rottenberg, 2005). In order to more intuitively reflect the effectiveness of the experimental paradigm we designed, we draw comparison charts of the valence and arousal of the healthy group and depressed group before and after positive film clips stimulation, as shown in **Figure 8**. The healthy group was in a calm mood before watching the positive film clip, and the pleasure degree increased significantly after watching the film clip, which stimulated a happy mood. The depressed group felt a little sad before watching the film clip, but their mood became more and more depressed after watching the film clip, and the arousal degree did not change much. This is consistent with the characteristics of depression.

4.4.2. Determining the Number of Network Layers

We use three kinds of data in the whole framework, namely 2D images, 3D facial landmark points, and AUs. We first use 2D face images and 3D facial landmark points as input to train different deep DBNs for the five stimuli with three emotional valences, respectively. We use the validation set to test networks and find that the recognition accuracy of the both mainly increases with the number of layers, reaching the highest on the forth hidden layer using 2D images or 3D landmark points trained DBN models, but both of them subsequently lose recognition performance as the number of layers increases, as shown in **Figure 9**. Therefore, we regard 4-hidden-layer DBN as a benchmark model of the 3D-DGDN.

4.4.3. Global Performance Analysis

The accuracy of all proposed models for three emotional valences of the five stimuli was tested on the dataset, including all males and females. The experiment results are shown in **Table 3** and **Figure 10**. It can be seen from **Table 3** and **Figure 10** that the best performance among the three emotional valences of the five stimuli was obtained based on the joint model. In particular, the performance of 3D-DGDN was higher than the 2D-SADN in all tasks. In the process of data collection, it was found that depressed patients or subjects with depressive tendencies were more prone to hyperactivity, which would lead to changes in depth information. Therefore, we added time series information to depth information for modeling, which will obtain more discriminative features. However, the combined network produced good results. This indicates that the 2D-SADN

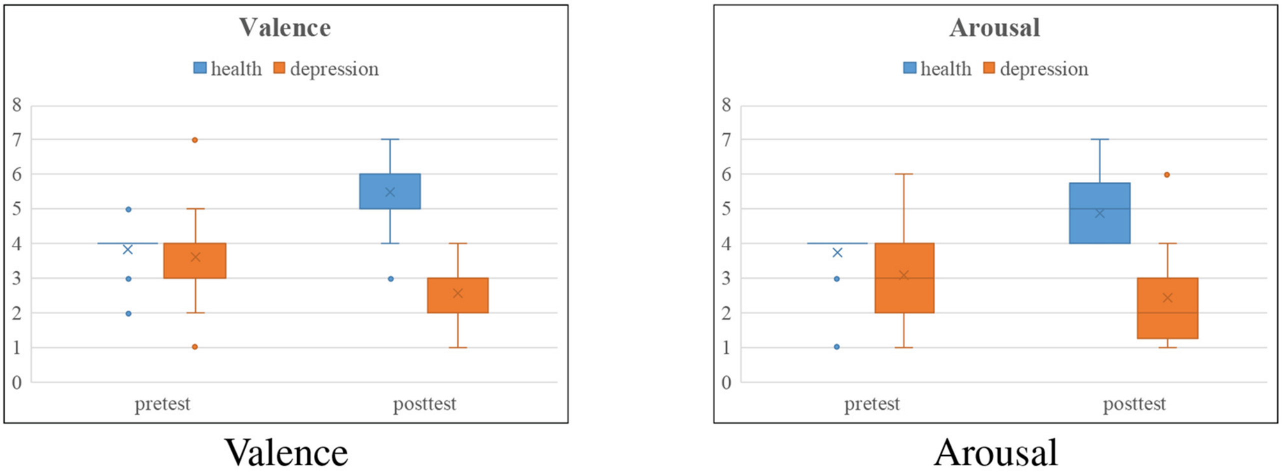


FIGURE 8 | Box comparison charts of the valence and arousal of the healthy group and depressed group before and after positive film clips stimulation.

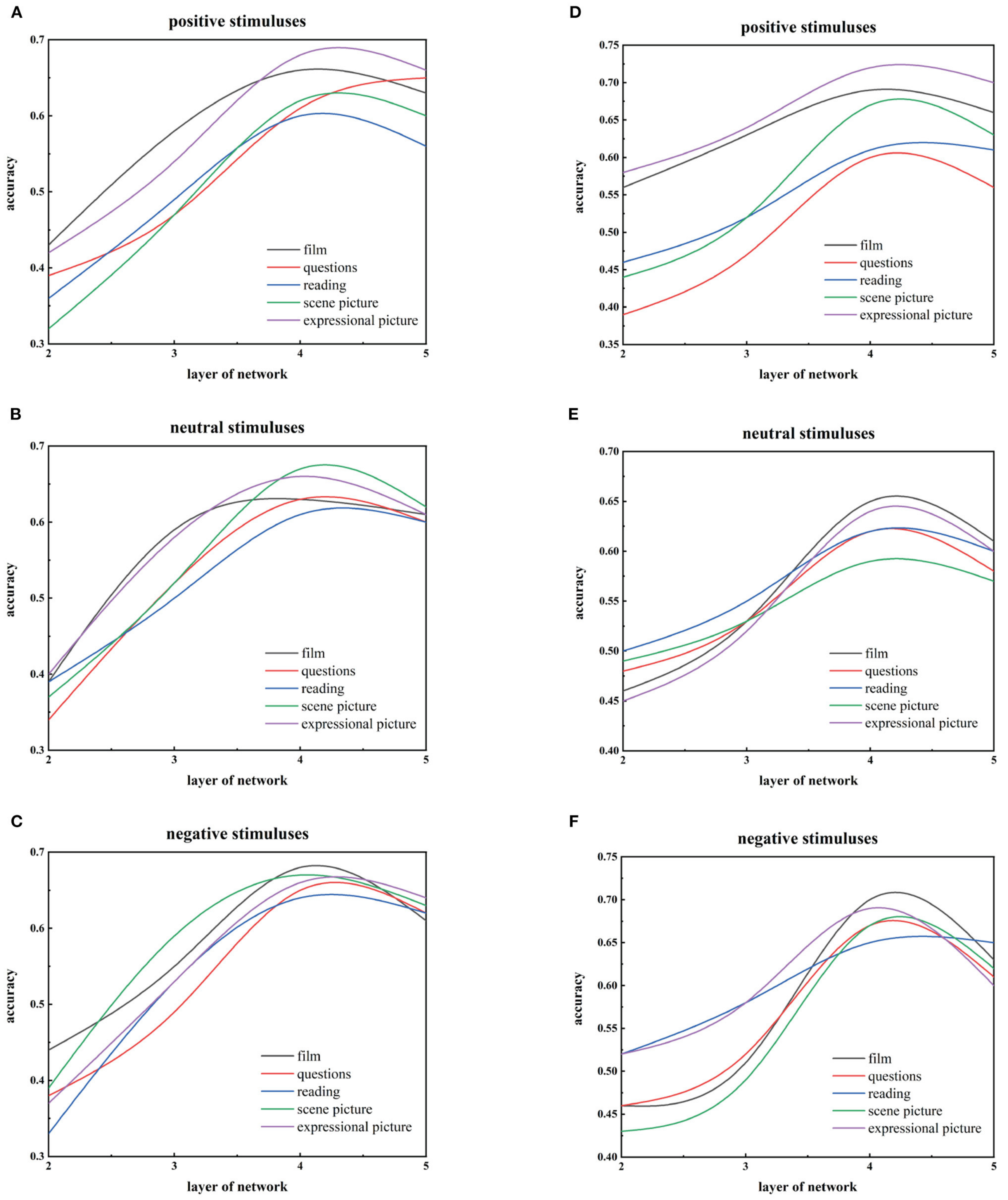


FIGURE 9 | Comparison of recognition performance with the different hidden levels on three emotional valences of the five stimuli for 3D facial points and 2D images. **(A–C)** are based on 3D facial points. **(D–F)** are based on 2D images.

TABLE 3 | The accuracy of all models for three emotional valences of five stimuli.

Type	Models	Positive	Neutral	Negative	Mean
Film	4DBN-AU	0.638	0.603	0.673	0.638
	AU-4DBN	0.693	0.635	0.701	0.676
	3D-DGDN	0.745	0.677	0.752	0.725
	2D-SADN	0.682	0.617	0.694	0.664
	Joint(2D-3D)	**0.798**	0.716	**0.807**	**0.774**
Question	4DBN-AU	0.605	0.593	0.592	0.597
	AU-4DBN	0.639	0.636	0.642	0.639
	3D-DGDN	0.687	0.659	0.693	0.680
	2D-SADN	0.618	0.632	0.647	0.632
	Joint(2D-3D)	0.702	0.683	0.713	0.699
Reading	4DBN-AU	0.572	0.583	0.601	0.585
	AU-4DBN	0.623	0.625	0.652	0.633
	3D-DGDN	0.668	0.658	0.694	0.673
	2D-SADN	0.583	0.613	0.635	0.610
	Joint(2D-3D)	0.711	0.697	0.712	0.707
Scene picture	4DBN-AU	0.617	0.538	0.608	0.588
	AU-4DBN	0.671	0.592	0.672	0.645
	3D-DGDN	0.716	0.651	0.724	0.697
	2D-SADN	0.623	0.613	0.668	0.635
	Joint(2D-3D)	0.747	0.707	0.752	0.735
Expression picture	4DBN-AU	0.659	0.591	0.635	0.628
	AU-4DBN	0.703	0.642	0.690	0.678
	3D-DGDN	0.729	0.683	0.751	0.721
	2D-SADN	0.684	0.657	0.668	0.700
	Joint(2D-3D)	**0.770**	0.725	**0.783**	**0.759**

Bold indicates a higher recognition rate.

network is a performance supplement to the 3D-DGDN network, and the two networks are complementary to each other.

It can be also seen that the recognition accuracy of positive and negative stimuli is higher than that of neutral stimulus, which is consistent with the emotional feedback theory of depressed patients compared with the healthy group, patients with depression have behavioral patterns such as weakened positive emotional feedback and enhanced negative emotional feedback. So in some cases, they have formed specific facial expressions such as reduced positive expressions and increased negative expressions (Babette et al., 2005). That is to say, they will not produce a larger change in expression compared with the normal population when facing the same stimulus. Therefore, there is a significant difference between positive and negative stimuli. Simultaneously, the accuracy of watching film clips is relatively higher in all positive or negative stimulus tasks and the highest recognition rates reach up to 0.798 and 0.807 for positive and negative stimulus, respectively, as shown in bold. The next high recognition rate is to view the expressional picture, as shown in bold. Because emotionally charged clips and images can, in principle, elicit an observable response (Pampouchidou et al., 2019). It is because that in order to eliminate the influence of unrelated facial movements on the facial expression behavior analysis of the participants, we only used the experimental data that the participants are completely prohibited from speaking in these two tasks. Relatively poor accuracies are obtained for questions and readings, which could be because facial expressions are associated with speech. When one feature is mixed with other factors, the purity expressed by this feature is not high.

4.4.4. Difference in Gender Analysis

The gender-dependent experiments are analyzed based on the integrated network. **Figure 11** shows the comparison of depression recognition based on gender difference under 95% confidence interval. From **Figure 11**, we can find that females' recognition accuracies are universally higher than that of males in three kinds of emotion valence, which explains that females are more likely to be aroused emotionally. According to WHO, evidence suggests that women are more prone than men to experience anxiety, depression, and somatic complaints—physical symptoms that cannot be explained medically[5]. We can also see that women are more likely to show the effects of positive stimulation than men. Among the three emotional valence tasks, the difference between female and male groups under the negative stimuluses are the smallest, which indicates that both female and male groups have higher accuracy and sensitivity under the negative stimuluses. In general, female are more emotionally aroused than male.

4.4.5. Comparative Analysis of Relevant Works

We compared our method's accuracy with the methods using the same data set, as shown in **Table 4**. Although our accuracy rate was slightly lower than that in previous studies (Li et al., 2018) (the selected samples are all major depressive disorder) in neutral stimulation tasks, we have improved the accuracy in the remaining tasks. Studies have shown that severely depressed individuals have lower emotional feedback to both positive and negative stimuli compared with healthy individuals (Nesse and Randolph, 2000). Therefore, our research results are more in line with the emotional experience of depressed patients. However, the performance has declined compared with our previous work (Guo et al., 2019). We will compare and summarize from the following points:

- *Data source:* The previous work only used the three-dimensional face point data collected by Kinect, although the point data can better handle some additional variables in the picture, such as illumination, angle, skin color, and so on, the videos (pictures) collected by the optical camera are still the main focus in actual application scenarios. Therefore, we combined the two data sources to make up for each other in this work.
- *Model:* The main framework of the DBN model was used in the two works. Considering that it was an extension of the previous work, the same data samples and the main model framework were used. However, this work further upgraded the model, using DBN to extract static facia appearance information and combined with the LSTM network to extract the dynamic information, and finally through a full connection

[5] Available at: https://www.who.int/life-course/news/commentaries/2015-intl-womens-day/.

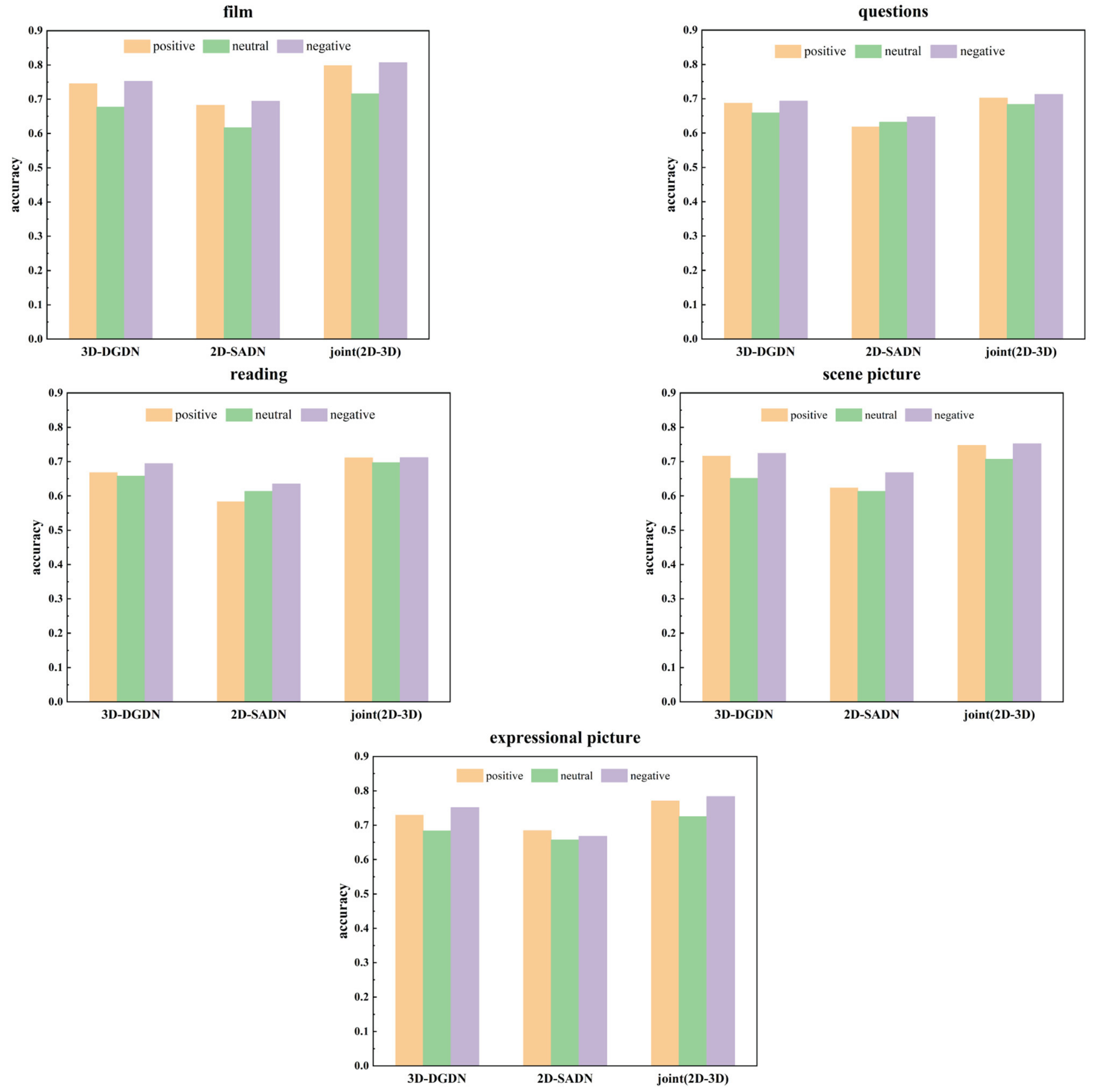


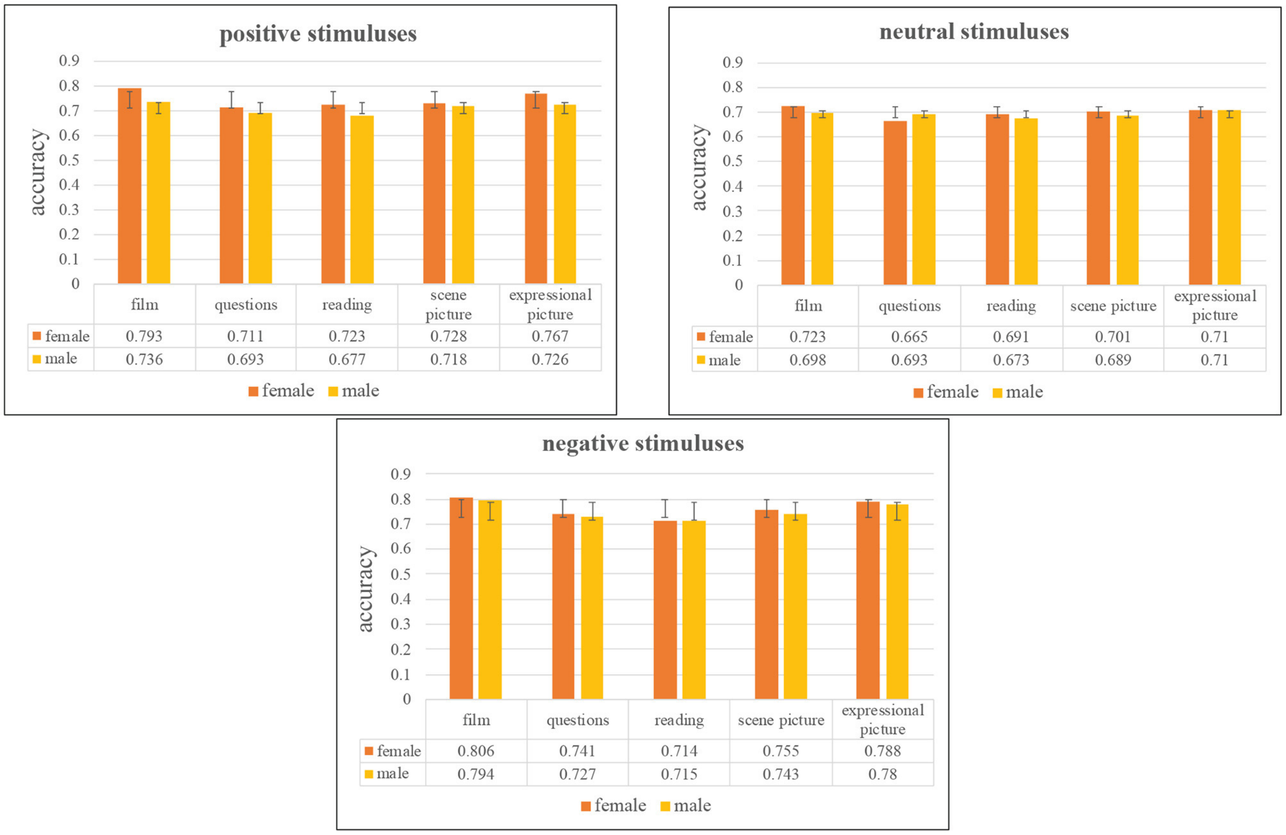
FIGURE 10 | Comparison of accuracy with the five stimuli among three networks.

to connect the two networks. The design of the model is complete and more prosperous than the previous work.

- *Practice:* We sampled frames by frames in the previous work, preprocessed 3D facial points data, and converted them into 200 ∗ 200 grayscale images. It took nearly a week to calculate the entire batch of data for 30 cycles. In this work, we sampled between frames and processed the original data directly. Although the network was more complicated than previous work, the calculation time was reduced. It only took 5 days to complete the whole batch of data for 100 cycles.

However, the classification accuracy has declined due to the following aspects.

First of all, the previous work results showed that visual stimuli classification effect, such as watching film clips and pictures, was better than the classification effect of language expression (here, mainly talking about changes in facial expressions). In our entire experiment, subjects completed every task and were asked to answer questions. The data used earlier included responses to questions on watching the film clips and seeing facial/scene expression tasks. In this work, to further verify



positive stimuluses

	film	questions	reading	scene picture	expressional picture
🟧 female	0.793	0.711	0.723	0.728	0.767
🟨 male	0.736	0.693	0.677	0.718	0.726

neutral stimuluses

	film	questions	reading	scene picture	expressional picture
🟧 female	0.723	0.665	0.691	0.701	0.71
🟨 male	0.698	0.693	0.673	0.689	0.71

negative stimuluses

	film	questions	reading	scene picture	expressional picture
🟧 female	0.806	0.741	0.714	0.755	0.788
🟨 male	0.794	0.727	0.715	0.743	0.78

FIGURE 11 | Comparison of accuracy on different gender under 95% confidence interval.

whether direct visual stimuli are more likely to elicit emotions in the depressed group, we selected only the data participants did not speak during watching the film clips and seeing facial/scene expression tasks. Coupled with the strategy of sampling between frames, the overall amount of data is far less than the previous work. For deep learning, the more considerable the amount of data, the better the trained model's performance.

Then, we converted 1,347 three-dimensional points into 200 ∗ 200 grayscale images in the preliminary work, which undoubtedly added some additional information for recognition. Whether this information played a role in the performance of the model could not be reasonably explained. In this work, we directly used the raw data without any additional information, so the final results are calculated from both the data and the model.

Besides, this paper's conclusion was obtained after 100 cycles based on 10-fold cross-validation, which is more stable than previous work.

In conclusion, on the basis of using the original data, the study in this paper significantly reduces the calculation time, while the accuracy rate is slightly lower than before, such as 0.02 for women and 0.01 for men. Based on the above points, this work is more reasonable and universal from three aspects of theory, model, and practice.

5. CONCLUSION AND FUTURE WORK

Depression is a common mental illness that can negatively affect people's mental health and daily life. In recent years, researchers have been looking for an objective evaluation method and quantitative indicators to identify depression objectively and effectively. Among them, the research of depression recognition based on facial expression behavior is a hot topic. In this paper, We first designed an experimental paradigm that can effectively stimulate the emotional differences between healthy and depressed groups and established a database for identifying depression. And then, we presented two deep network models that collaborate with each other. The first network was 2D-SADN, which is used to extract the static appearance features from images, and the second network was 3D-DGDN, which captures the dynamic geometry features of 3D facial landmark points and AUs from Kinect. We showed that the accuracy obtained by the 2D-SADN was lower than that of the 3D-DGDN, which may be because poor image quality and the DBN model cannot well retain the 2D information of an image. At last, we achieved the best recognition rates using the integrated deep network on the collected databases.

From the perspective of emotional stimulus materials, the experimental results also support this theory: apparent

TABLE 4 | Comparison of accuracy based on the same database.

Gender	Task	Accuracy								
		Positive stimulus			Neutral stimulus			Negative stimulus		
		This work	Previous work	(Li et al., 2018)	This work	Previous work	(Li et al., 2018)	This work	Previous work	(Li et al., 2018)
Female	Film clips	0.793	0.825	0.737	0.723	0.761	0.868	0.806	0.816	0.763
	Questions	0.711	0.761	0.649	0.665	0.705	0.737	0.741	0.765	0.675
	Readings	0.723	0.768	0.711	0.691	0.728	0.658	0.714	0.751	0.658
	Scene pictures	0.728	0.801	0.711	0.701	0.713	0.763	0.755	0.806	0.711
	Expression pictures	0.767	0.806	0.632	0.710	0.741	0.737	0.788	0.801	0.711
Male	Film clips	0.736	0.772	0.647	0.698	0.733	0.794	0.794	0.782	0.647
	Questions	0.693	0.738	0.725	0.693	0.728	0.735	0.727	0.726	0.667
	Readings	0.677	0.724	0.618	0.673	0.694	0.676	0.715	0.745	0.588
	Scene pictures	0.718	0.755	0.618	0.689	0.714	0.706	0.743	0.776	0.647
	Expression pictures	0.726	0.761	0.647	0.710	0.673	0.706	0.780	0.737	0.588

differences existed between the health and depressed groups for pleasant or unpleasant stimuli. Mostly, the accuracy of watching film clips and expressional pictures emotional stimulus tasks were generally high, but the accuracy of answering questions and reading texts is low. This is because the subjects recorded facial expressions while speaking in both the masks, with one feature being mixed with other factors. We will further investigate the experimental strategies to construct a more distinctively characteristic depression behavior database in future work. We will further analyze which of the two states of speech and non-speech information on discriminating depressed patients. Since CNN has shown superior performance on image classification/recognition problem, we aim to use the CNN-based methods to model depth information and video information. We also will try to use the state-of-the-art multimodal fusion methods to identify depression.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: Data involves privacy and has not been disclosed. Requests to access these datasets should be directed to Zhenyu Liu, liuzhenyu@lzu.edu.cn.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Tianshui Third People's Hospital and Lanzhou University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work. WG, HY, and BH were responsible for the entire study, including study concepts and study design. WG, ZL, and YX contributed to the experimental paradigm design. WG and YX were responsible for collecting data. WG wrote the manuscript. HY helped WG draft the manuscript and modify the important content. All authors read and approved the final manuscript.

FUNDING

This work was supported in part by the National Key Research and Development Program of China (Grant No. 2019YFA0706200), in part by the National Natural Science Foundation of China (Grant Nos. 61802159, 61632014, 61627808, 61210010, 11664036, 62067008), in part by the National Basic Research Program of China (973 Program, Grant No. 2014CB744600), in part by the Program of Beijing Municipal Science & Technology Commission (Grant No. Z171100000117005), in part by the High School Science and Technology Innovation Team Project of Gansu (Grant No. 2017C-03).

ACKNOWLEDGMENTS

The authors acknowledge all subjects and thank Director Wang and Director Ding from Tianshui Third People's Hospital for clinical diagnosis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.609760/full#supplementary-material

REFERENCES

Alghowinem, S. (2015). *Multimodal analysis of verbal and nonverbal behavior on the example of clinical depression* (Ph.D. thesis). The Australian National University Canberra, ACT.

Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Hyett, M., Parker, G., et al. (2018). Multimodal depression detection: fusion analysis of paralinguistic, head pose and eye gaze behaviors. *IEEE Trans. Affect. Comput.* 9, 478–490. doi: 10.1109/TAFFC.2016.2634527

Alghowinem, S., Goecke, R., Wagner, M., Parker, G., and Breakspear, M. (2013). "Eye movement analysis for depression detection," in *2013 IEEE International Conference on Image Processing* (Melbourne, VIC), 4220–4224.

Aly, S., Abbott, A. L., and Torki, M. (2016). "A multi-modal feature fusion framework for kinect-based facial expression recognition using dual kernel discriminant analysis (dkda)," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* (Lake Placid, NY), 1–10.

Anis, K., Zakia, H., Mohamed, D., and Jeffrey, C. (2018). "Detecting depression severity by interpretable representations of motion dynamics," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)* (Xi'an: IEEE), 739–745.

Babette, E., Heyn, K., Gebhard, R., and Bachmann, S. (2005). Facial expression of emotions in borderline personality disorder and depression. *J. Behav. Ther. Exp. Psychiatry* 36, 183–196. doi: 10.1016/j.jbtep.2005.05.002

Bhatia, S. (2016). "Multimodal sensing of affect intensity," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, ICMI '16 (New York, NY: Association for Computing Machinery), 567–571.

Bhatia, S., Hayat, M., Breakspear, M., Parker, G., and Goecke, R. (2017). "A video-based facial behaviour analysis approach to melancholia," in *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)* (Washington, DC), 754–761.

Carvalho, N., Laurent, E., Noiret, N., Chopard, G., Haffen, E., Bennabi, D., et al. (2015). Eye movement in unipolar depression and bipolar disorders: A systematic review of the literature. *Front. Psychol.* 6:1809. doi: 10.3389/fpsyg.2015.01809

Cohn, J. F., Kruez, T. S., Matthews, I., Yang, Y., Nguyen, M. H., Padilla, M. T., et al. (2009). "Detecting depression from facial actions and vocal prosody," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops* (Amsterdam), 1–7.

Delle-Vigne, D., Wang, W., Kornreich, C., Verbanck, P., and Campanella, S. (2014). Emotional facial expression processing in depression: data from behavioral and event-related potential studies. *Neurophysiol. Clin.* 44, 169–187. doi: 10.1016/j.neucli.2014.03.003

Demakakos, P., Cooper, R., Hamer, M., Oliveira, C., Hardy, R., and Breeze, E. (2015). The bidirectional association between depressive symptoms and gait speed: evidence from the english longitudinal study of ageing (ELSA). *PLoS ONE* 8:e68632. doi: 10.1371/journal.pone.0068632

Fischer, A., and Igel, C. (2012). "An introduction to restricted boltzmann machines," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications* (Buenos Aires: Springer), 14–36.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 580–587.

Gong, B., Wang, Y., Liu, J., and Tang, X. (2009). "Automatic facial expression recognition on a single 3d face by exploring shape deformation," in *Proceedings of the 17th ACM International Conference on Multimedia*, MM '09 (New York, NY: Association for Computing Machinery), 569–572. doi: 10.3969/j.issn.1000-6729.2011.01.011

Gong, X., Huang, Y., Wang, Y., and Luo, Y. (2011). Revision of the chinese facial affective picture system. *Chinese Ment. Health J.* 25, 40–46.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., and Schmidhuber, J. (2016). Lstm: a search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* 28, 2222–2232. doi: 10.1109/TNNLS.2016.2582924

Gross, J. J. and Levenson, R. W. (1995). Emotion elicitation using films. *Cogn. Emot.* 9, 87–108. doi: 10.1080/02699939508408966

Guo, W., Yang, H., and Liu, Z. (2019). "Deep neural networks for depression recognition based on facial expressions caused by stimulus tasks," in *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)* (Cambridge, MA: IEEE), 133–139.

Gupta, R., Malandrakis, N., Xiao, B., Guha, T., Van Segbroeck, M., Black, M., et al. (2014). "Multimodal prediction of affective dimensions and depression in human-computer interactions," in *AVEC'14* (Orlando, FL), 33–40.

He, L., Jiang, D., and Sahli, H. (2015). "Multimodal depression recognition with dynamic visual and audio cues," in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII) (Xi'an)*, 260–266.

He, L., Jiang, D., and Sahli, H. (2018). Automatic depression analysis using dynamic facial appearance descriptor and dirichlet process fisher encoding. *IEEE Trans. Multimedia* 21, 1476–1486. doi: 10.1109/TMM.2018.2877129

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

Jan, A., Meng, H., Gaus, Y. F. B. A., and Zhang, F. (2018). Artificial intelligent system for automatic depression level analysis through visual and vocal expressions. *IEEE Trans. Cogn. Dev. Syst.* 10, 668–680. doi: 10.1109/TCDS.2017.2721552

Jiang, H., Hu, B., Liu, Z., Yan, L., Wang, T., Liu, F., et al. (2017). Investigation of different speech types and emotions for detecting depression using different classifiers. *Speech Commun.* 90, 39–46. doi: 10.1016/j.specom.2017.04.001

Joshi, J., Dhall, A., Goecke, R., Breakspear, M., and Parker, G. (2012). "Neural-net classification for spatio-temporal descriptor based depression analysis," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)* (Tsukuba), 2634–2638.

Jung, H., Lee, S., Yim, J., Park, S., and Kim, J. (2015). "Joint fine-tuning in deep neural networks for facial expression recognition," in *2015 IEEE International Conference on Computer Vision (ICCV)* (Santiago), 2983–2991.

King, D. E. (2009). Dlib-ml: a machine learning toolkit. *J. Mach. Learn. Res.* 10, 1755–1758. doi: 10.1145/1577069.1755843

Kroenke, K., and Spitzer, R. L. (2002). The phq-9: a new depression diagnostic and severity measure. *Psychiatr. Ann.* 32, 509–521. doi: 10.3928/0048-5713-20020901-06

Lang, P. (1980). "Behavioral treatment and bio-behavioral assessment: computer applications," in *Technology in Mental Health Care Delivery Systems*, 119–137.

Leyvand, T., Meekhof, C., Wei, Y., Sun, J., and Guo, B. (2011). Kinect identity: technology and experience. *Computer* 44, 94–96. doi: 10.1109/MC.2011.114

Li, J., Liu, Z., Ding, Z., and Wang, G. (2018). "A novel study for mdd detection through task-elicited facial cues," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (Madrid: IEEE), 1003–1008.

Li, Y. (2014). *An ERP study of cognitive processing of subthreshold depression individuals on different stimulus emotional pictures* (Master's thesis). Beijing University of Chinese Medicine (Beijing).

Liu, M., Shan, S., Wang, R., and Chen, X. (2014). "Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition," in *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Columbus, OH).

McIntyre, G., Göcke, R., Hyett, M., Green, M., and Breakspear, M. (2009). "An approach for automatically measuring facial activity in depressed subjects," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops* (Amsterdam), 1–8.

McIntyre, G. J., et al. (2010). *The computer analysis of facial expressions: on the example of depression and anxiety* (Ph.D. thesis). The Australian National University. Canberra, ACT.

McPherson, A., and Martin, C. R. (2010). A narrative review of the beck depression inventory (bdi) and implications for its use in an alcohol-dependent population. *J. Psychiatr. Ment. Health Nurs.* 17, 19–30. doi: 10.1111/j.1365-2850.2009.01469.x

Melo, W. C. D., Granger, E., and Hadid, A. (2019). "Combining global and local convolutional 3d networks for detecting depression from facial expressions," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)* (Lille: IEEE), 1–8.

Meng, H., Huang, D., Wang, H., Yang, H., AI-Shuraifi, M., and Wang, Y. (2013). "Depression recognition based on dynamic facial and vocal expression features using partial least square regression," in *Proceedings of the 3rd ACM*

International Workshop on Audio/Visual Emotion Challenge, AVEC '13 (New York, NY: Association for Computing Machinery), 21–30.

Michalak, J., Troje, N. F., Fischer, J., Vollmar, P., Heidenreich, T., and Schulte, D. (2009). Embodiment of sadness and depression–gait patterns associated with dysphoric mood. *Psychosom. Med.* 71, 580–587. doi: 10.1097/PSY.0b013e3181a2515c

Morency, L.-P., Stratou, G., DeVault, D., Hartholt, A., Lhommet, M., Lucas, G., et al. (2015). "Simsensei demonstration: a perceptive virtual human interviewer for healthcare applications," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15 (Austin, TX: AAAI Press), 4307–4308.

Nasir, M., Jati, A., Shivakumar, P. G., Chakravarthula, S. N., and Georgiou, P. (2016). "Multimodal and multiresolution depression detection from speech and facial landmark features," in *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*, AVEC '16 (New York, NY: Association for Computing Machinery), 43–50.

National Collaborating Centre for Mental Health (2010). *Depression: The Treatment and Management of Depression in Adults*. British Psychological Society.

Nesse, and Randolph, M. (2000). Is depression an adaptation? *Arch. Gen. Psychiatry* 57, 14–20. doi: 10.1001/archpsyc.57.1.14

Ng, A. Y., and Jordan, M. I. (2002). "On discriminative vs. generative classifiers: a comparison of logistic regression and naive bayes," in *15th Annual Conference on Neural Information Processing Systems (NIPS)*, Vol. 14 (Vancouver, QC), 841–848.

Nicholas, C., Stefan, S., Jarek, K., Sebastian, S., Julien, E., and Thomas F, Q. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Commun.* 71, 10–49. doi: 10.1016/j.specom.2015.03.004

Ooi, K. E. B., Lech, M., and Allen, N. B. (2013). Multichannel weighted speech classification system for prediction of major depression in adolescents. *IEEE Trans. Biomed. Eng.* 60, 497–506. doi: 10.1109/TBME.2012.2228646

Pampouchidou, A., Marias, K., Tsiknakis, M., Simos, P., Yang, F., Lematre, G., et al. (2016a). "Video-based depression detection using local curvelet binary patterns in pairwise orthogonal planes," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Orlando), 3835–3838.

Pampouchidou, A., Marias, K., Tsiknakis, M., Simos, P., Yang, F., and Meriaudeau, F. (2015). "Designing a framework for assisting depression severity assessment from facial image analysis," in *2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)* (Kuala Lumpur), 578–583.

Pampouchidou, A., Simantiraki, O., Fazlollahi, A., Pediaditis, M., Manousos, D., Roniotis, A., et al. (2016b). "Depression assessment by fusing high and low level features from audio, video, and text," in *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*, AVEC '16 (New York, NY: Association for Computing Machinery), 27–34.

Pampouchidou, A., Simos, P. G., Marias, K., Meriaudeau, F., Yang, F., Pediaditis, M., et al. (2019). Automatic assessment of depression based on visual cues: a systematic review. *IEEE Trans. Affect. Comput.* 10, 445–470. doi: 10.1109/TAFFC.2017.2724035

Reddy, M. S. (2012). Depression - the global crisis. *Indian J. Psychol. Med.* 34, 201–203. doi: 10.4103/0253-7176.106011

Rottenberg, J. (2005). Mood and emotion in major depression. *Curr. Dir. Psychol. Sci.* 14, 167–170. doi: 10.1111/j.0963-7214.2005.00354.x

Rottenberg, J., Gross, J. J., and Gotlib, L. H. (2005). Emotion context insensitivity in major depressive disorder. *J. Abnorm. Psychol.* 114:627. doi: 10.1037/0021-843X.114.4.627

Schwartz, G. E., Fair, P. L., Salt, P., and Klerman, G. L. (1976). Facial expression and imagery in depression: an electromyographic study. *Psychosom. Med.* 38, 337–347. doi: 10.1097/00006842-197609000-00006

Sidorov, M., and Minker, W. (2014). "Emotion recognition and depression diagnosis by acoustic and visual features: a multimodal approach," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*, AVEC '14 (New York, NY: Association for Computing Machinery), 81–86.

Stratou, G., Scherer, S., Gratch, J., and Morency, L. P. (2014). Automatic nonverbal behavior indicators of depression and ptsd: the effect of gender. *J. Multimodal User Interfaces* 9, 17–29. doi: 10.1007/s12193-014-0161-4

Wang, P., Barrett, F., Martin, E., Milonova, M., Gur, R. E., Gur, R. C., et al. (2008). Automated video-based facial expression analysis of neuropsychiatric disorders. *J. Neurosci. Methods* 168, 224–238. doi: 10.1016/j.jneumeth.2007.09.030

Wen, L., Li, X., Guo, G., and Zhu, Y. (2015). Automated depression diagnosis based on facial dynamic analysis and sparse coding. *IEEE Trans. Inform. Forens. Secur.* 10, 1432–1441. doi: 10.1109/TIFS.2015.2414392

Williamson, J. R., Quatieri, T. F., Helfer, B. S., Ciccarelli, G., and Mehta, D. D. (2014). "Vocal and facial biomarkers of depression based on motor incoordination and timing," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*, AVEC '14 (New York, NY: Association for Computing Machinery), 65–72.

Winograd-Gurvich, C., Georgiou-Karistianis, N., Fitzgerald, P., Millist, L., and White, O. (2006). Ocular motor differences between melancholic and non-melancholic depression. *J. Affect. Disord.* 93, 193–203. doi: 10.1016/j.jad.2006.03.018

Yang, Y., Fairbairn, C., and Cohn, J. F. (2013). Detecting depression severity from vocal prosody. *IEEE Trans. Affect. Comput.* 4, 142–150. doi: 10.1109/T-AFFC.2012.38

Zhao, X., Huang, D., Dellandrea, E., and Chen, L. (2010). "Automatic 3d facial expression recognition based on a bayesian belief net and a statistical facial feature model," in *Proceedings of the 2010 20th International Conference on Pattern Recognition*, ICPR '10 (Istanbul: IEEE Computer Society), 3724–3727.

Zhou, X., Jin, K., Shang, Y., and Guo, G. (2020). Visually interpretable representation learning for depression recognition from facial images. *IEEE Trans. Affect. Comput.* 11, 542–552. doi: 10.1109/TAFFC.2018.2828819

Conflict of Interest: 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.

REVIEW
published: 09 July 2021
doi: 10.3389/fpsyg.2021.653112

Review: Posed vs. Genuine Facial Emotion Recognition and Expression in Autism and Implications for Intervention

Paula J. Webster[1], *Shuo Wang*[1]* *and Xin Li*[2]

[1]Department of Chemical and Biomedical Engineering, Rockefeller Neuroscience Institute, West Virginia University, Morgantown, WV, United States, [2]Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, United States

OPEN ACCESS

Edited by:
Anthony P. Atkinson,
Durham University, United Kingdom

Reviewed by:
Mikle South,
Brigham Young University,
United States
Michael K. Yeung,
Hong Kong Polytechnic University,
China

****Correspondence:***
Shuo Wang
shuo.wang@mail.wvu.edu

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

Received: *13 January 2021*
Accepted: *02 June 2021*
Published: *09 July 2021*

Citation:
Webster PJ, Wang S and Li X (2021)
Review: Posed vs. Genuine Facial
Emotion Recognition and
Expression in Autism and
Implications for Intervention.
Front. Psychol. 12:653112.
doi: 10.3389/fpsyg.2021.653112

Different styles of social interaction are one of the core characteristics of autism spectrum disorder (ASD). Social differences among individuals with ASD often include difficulty in discerning the emotions of neurotypical people based on their facial expressions. This review first covers the rich body of literature studying differences in facial emotion recognition (FER) in those with ASD, including behavioral studies and neurological findings. In particular, we highlight subtle emotion recognition and various factors related to inconsistent findings in behavioral studies of FER in ASD. Then, we discuss the dual problem of FER – namely facial emotion expression (FEE) or the production of facial expressions of emotion. Despite being less studied, social interaction involves both the ability to recognize emotions and to produce appropriate facial expressions. How others perceive facial expressions of emotion in those with ASD has remained an under-researched area. Finally, we propose a method for teaching FER [FER teaching hierarchy (FERTH)] based on recent research investigating FER in ASD, considering the use of posed vs. genuine emotions and static vs. dynamic stimuli. We also propose two possible teaching approaches: (1) a standard method of teaching progressively from simple drawings and cartoon characters to more complex audio-visual video clips of genuine human expressions of emotion with context clues or (2) teaching in a field of images that includes posed and genuine emotions to improve generalizability before progressing to more complex audio-visual stimuli. Lastly, we advocate for autism interventionists to use FER stimuli developed primarily for research purposes to facilitate the incorporation of well-controlled stimuli to teach FER and bridge the gap between intervention and research in this area.

Keywords: facial expression of emotion, emotion recognition, posed vs. genuine emotion, autism spectrum disorder, social deficits

INTRODUCTION

Individuals with autism spectrum disorder (ASD) often have difficulty interpreting and regulating their own emotions, understanding the emotions expressed by others, and labeling emotions based on viewing the faces of others (Harms et al., 2010; Uljarevic and Hamilton, 2013; Sheppard et al., 2016). These differences can contribute to social self-isolation by those with ASD either when others respond negatively if the person with ASD lacks a typical, socially expected response, or if the person with ASD chooses to socially isolate themselves to avoid possibly stressful interactions if they realize they struggle to recognize and respond appropriately to expressions of emotion by others (Jaswal and Akhtar, 2019).

Research investigating facial emotion recognition (FER) in ASD has primarily utilized static images composed of posed facial expressions (Pelphrey et al., 2007; Monk et al., 2010); however, more recent research has begun exploring the use of dynamic video with actors making posed facial expressions (Golan et al., 2015; Fridenson-Hayo et al., 2016; Simões et al., 2018). Few studies have utilized face stimuli of humans expressing genuine, spontaneous expressions of emotion, whether static or dynamic (Cassidy et al., 2014). This distinction is important because research has shown that the human brain, and artificial intelligence (AI) systems, process posed facial expressions differently compared to how spontaneous expressions of emotion are processed (Hess et al., 1989; Schmidt et al., 2006; Wang et al., 2015; Park et al., 2020).

Results have been mixed with most studies indicating that posed expressions of emotion being easier to recognize than those that are spontaneous (Naab and Russell, 2007); however, accuracy for FER may also depend on the specific emotion being evaluated (Faso et al., 2014; Sauter and Fischer, 2018). This may be due to the prototypical nature of posed expressions (e.g., most people show fewer teeth when they smile for posed pictures; Van Der Geld et al., 2008), whereas there is much more variability in genuine expressions of some feelings such as sadness (Krumhuber et al., 2019). Therefore, it has been proposed that the traditional use of posed facial expression stimuli in research may have artificially inflated behavioral measures of accuracy during emotion recognition tasks (Sauter and Fischer, 2018). Therefore, the historically prevalent use of posed facial expression stimuli in ASD research investigating FER may contribute to the mixed results seen in this research area.

How might these dissimilarities in posed vs. spontaneous facial expression stimuli be perceived differently by those with ASD? This review further argues that posed vs. genuine emotion is a critical factor that deserves more consideration when studying FER in ASD. We will first review the rich literature on the perception of posed facial expressions of emotion, highlighting the differences between ASD and control groups, though inconclusively. We will then discuss some recent research investigating how individuals with ASD differ from controls when asked to produce posed facial expressions of emotion and review the latest advances in the field of posed vs. spontaneous/genuine facial expressions and implications into autism research in terms of both perception and production of genuine facial expressions. Finally, based on these findings, we propose a method of teaching FER for individuals with ASD.

DIFFERENCES IN FER IN ASD

Autism studies investigating differences in understanding how others think or feel date back to as early as the 1970s (Langdell, 1978; Mesibov, 1984; Weeks and Hobson, 1987; Hobson et al., 1988; Ozonoff et al., 1990). In Langdell (1978) they found that adolescents with autism could identify schematically drawn happy and sad faces, but they demonstrated varying capability when sorting the faces just using the eye area. Another study (Hobson, 1986) provided further convincing evidence about the differences in the appraisal of facial expressions of emotion by children with autism suggesting that their failure to understand the emotional states of others might be related to their difficulty in recognizing the difference between particular emotions. However, due to different experimental designs (e.g., sorting, matching, and cross-modal), the interpretation of these early results is often debatable (Celani et al., 1999).

A more systematic study about the nature of early differences in social cognition in autism was conducted in Dawson et al. (2004) using high-density event-related potentials (ERPs). It was found that children with ASD, as young as 3 years of age, showed a disordered pattern of neural responses to emotional stimuli such as fearful vs. neutral facial expressions. More specifically, typically developing children demonstrated a larger early negative component and a negative slow wave to the fear than to the neutral, while children with autism did not show significant differences in both experiments. In contrast, the faster speed of early processing of the fear face among children with autism was associated with better performance on tasks assessing social attention such as social orienting, joint attention, and attention to distress. These findings have served as direct evidence for atypical psychological components involving emotion recognition among children with autism at a young age (3–4 years old).

To probe into the pathology of the underlying processes related to dysfunction in emotional and social cognition, it has been shown that amygdala dysfunction in ASD might contribute to a different ability to process social information (Adolphs et al., 2001). Varying face perception or emotion recognition in ASD might result from atypical fixations onto faces, which may, in turn, arise from amygdala dysfunction (Breiter et al., 1996; Baron-Cohen et al., 2000). This hypothesis is directly supported by evidence from both single-neuron recordings in the human amygdala (Rutishauser et al., 2013) and neuroimaging studies (Dalton et al., 2005; Kliemann et al., 2012). Given the critical role of the amygdala in emotion processing (Adolphs, 2008), more systematic studies will be needed to reveal whether the amygdala has a different response for posed vs. genuine emotions. Further studies using visual scanning/eye-tracking (Pelphrey et al., 2002), or functional neuroimaging (Dalton et al., 2005; Pelphrey et al., 2005), have shown abnormal activity in patients with ASD. Even with the enhanced emotional salience of facial stimuli, a positron emission tomography (PET) study showed that adults with ASD demonstrated lower activity in the fusiform cortex than typically developing (TD) controls and differed from the TD group within other brain regions (Hall et al., 2003). This line of research was further extended into the identification of differences in key components of human face processing systems that might contribute to the differences in processing facial expressions of emotion (Pelphrey and Carter, 2008).

Unlike previous studies employing more simplistic stimuli (e.g., the face stimulus as an exemplar of a given emotion, "100% expression"), subtle differences in FER were considered (Law Smith et al., 2010; Black et al., 2020). Using stimuli that incrementally morphed the expression between a neutral face and the posed expression, they found that adolescents and young adults with ASD were less accurate at identifying basic emotional expressions of disgust, anger, and surprise. In a follow-up study (Kennedy and Adolphs, 2013), adults with ASD were found to give ratings that were significantly less sensitive to a given emotion and less reliable across repeated testing. Therefore, an overall decreased specificity in emotion perception suggests a subtle but specific pattern of differences in facial emotion perception among those with ASD. Along this line of research, significant differences were found between males and females with ASD for emotion recognition but not for self-reported empathy recognition (Sucksmith et al., 2013). Most recently, a gender-biased study showed that differences in FER in *females* with autism might not be attributed to ASD but instead to their co-occurring alexithymia (difficulty describing one's own emotions and those of others; Ola and Gullon-Scott, 2020). Thus, consideration for future FER studies is to recruit significant numbers of male and female participants with ASD and consider sex as a factor in the analysis.

We note that there have been several excellent review articles about research findings of FER in ASD (Harms et al., 2010; Bons et al., 2011; Nuske et al., 2013; Uljarevic and Hamilton, 2013). In Harms et al. (2010), demographic and experiment-related factors are addressed to account for inconsistent findings in behavioral studies of FER in ASD. Future studies of FER in ASD suggested by Harms et al. (2010) include the incorporation of longitudinal designs to examine the developmental trajectory of FER and behavioral and brain imaging paradigms that include young children. In Uljarevic and Hamilton (2013), a formal meta-analytic study has shown that recognition of happiness was only marginally modified in ASD, but recognition of fear was marginally worse than recognition of happiness. In Nuske et al. (2013), it was found that (1) emotion-processing differences might not be *universal* to all individuals with ASD and are not *specific* to ASD; and (2) the specific pattern of emotion-processing strengths and weaknesses observed in ASD, involving difficulties with processing social vs. nonsocial, and complex versus simple emotional information, appears to be *unique* to ASD (Tang et al., 2019). It is also worth noting the "double empathy problem" described (Milton, 2012). It was found that just like people with ASD have difficulty interpreting the facial emotions of TDs, TD people have just as much difficulty understanding people with autism. Such a "double" perspective has profound implications for ASD service providers because differences in neurology could lead to differences in sociality. A more recent study (Milton and Sims, 2016) has demonstrated a need for less focus on remediation for patients with autism. Instead, it advocated for focusing on limiting social isolation as a more constructive solution. The most recent study (Crompton et al., 2020) has shown that peer-to-peer information transfer concerning autism is more effective than information transfer between persons with and without autism.

Given the finding that FER differences are not strictly applicable to those with ASD (Nuske et al., 2013), several studies have been conducted to compare differences in FER in ASD with other neurological disorders. In Wong et al. (2012), emotion recognition abilities are examined for three groups of children aged 7–13 years: high functioning autism (HFA), social phobia (SP), and TD. Although no evidence was found for negative interpretation biases in children with HFA or SP, children with HFA were found to detect mild affective expressions less accurately than TD peers suggesting subtle changes in emotion expression are more difficult for those with ASD. In Sachse et al. (2014), a similar study was conducted with adolescents and adults with HFA, schizophrenia (SZ), and TD to identify convergent and divergent mechanisms between ASD and SZ. It was found that individuals with SZ were comparable to TD in all emotion recognition measures, but the basic visuoperceptual abilities of the SZ individuals were reduced. By contrast, the HFA group was more affected in recognizing basic and complex emotions when compared to both SZ and TD. As reported in Sachse et al. (2014), group differences between SZ and ASD remained but only for recognizing complex emotions after taking facial identity recognition into account. Such experimental results suggest that (1) there is an SZ subgroup with predominantly paranoid symptoms that do not show problems in FER but visuoperceptual differences only; and (2) no shared FER difference was found for paranoid SZ and ASD, implying differential cognitive underpinnings of ASD and SZ about FER.

A study by Lundqvist (2015) directly links sensory abnormality with social dysfunction of ASD – for example, hyper-responsiveness to touch mediated social dysfunction in ASD, and the tactile sensory system is foundational for social functioning in ASD. There is also evidence that social functioning in those with ASD is impacted by sensory dysregulation in multiple sensory modalities that arise early in the progression of the disorder (Thye et al., 2018). This meta-analysis suggests an early intervention that targets sensory abnormalities and social differences, considering the critical role ASD sensory processing differences play in social interactions. In another systematic review and meta-analysis (Zhou et al., 2018), quantitative comparisons of sensory temporal acuity were made between healthy controls and two clinical groups (ASD and SZ). They revealed a consistent difference in multisensory temporal integration in ASD and SZ, which may be associated with differences in social communication. Finally, studying differential patterns of visual sensory alternation using neuroimaging (Martínez et al., 2019) has shown that SZ and ASD participants demonstrated similar FER and motion sensitivity differences, but significantly different visual processing contributed to FER group differences. This data would suggest that FER differences are not unique to ASD.

DIFFERENCES IN FACIAL EMOTION EXPRESSION IN ASD

It has been hypothesized that in ASD, both FER and FEE are affected, contributing importantly to social differences and difficulty in relationship formation (Manfredonia et al., 2019). By contrast, fewer studies about individuals with ASD have been devoted to FEE than FER in the published literature. In an early study of imitation and expression of facial affect (Loveland et al., 1994), the production of elicited/posed affective expressions is more difficult for individuals with ASD than for patients with Down's syndrome of similar chronological age, mental age, and IQ. In Begeer et al. (2008), four aspects of emotional competence (expression, perception, responding, and understanding) are reviewed for children and adolescents with ASD. It was found that different emotional competence in ASD was highly dependent on age, context, and intelligence. In another unique study (Faso et al., 2014), the dual problem of FER and FEE were studied, namely how facial expressivity by those with ASD is perceived by others. It was reported that facial expressions of emotion by participants with ASD were regarded as more intense and less natural than expressions by the TD group. Surprisingly, ASD expressions were also identified with greater accuracy by TD judges due primarily to the category of angry expressions. The above findings collectively suggest differences, instead of a reduced ability, in facial expressivity among individuals with ASD. Those differences do not necessarily hinder the accuracy of emotion recognition by others but may affect the quality of social interactions between ASD and TD, as demonstrated in a recent study (Sasson et al., 2017).

In Volker et al. (2009), each participant was photographed after being prompted to enact a facial expression from one of six basic emotions – happiness, sadness, anger, fear, surprise, and disgust. It was reported that children with HFA were significantly less adept at enacting sadness, and their expressions were dramatically odder compared to controls. However, no significant differences were found for anger and fear; and even more surprisingly, the ASD group demonstrated somewhat greater skills at enacting surprise and disgust. More recently, a systematic study (Brewer et al., 2016) investigated TD and ASD participants' ability to recognize facial expressions of emotion produced by TD and ASD actors posing basic emotions. With three designed posing conditions, this study aimed to determine whether potential group differences were due to (1) atypical cognitive representations of emotion; (2) affected the understanding of the communicative value of expressions; or (3) poor proprioceptive feedback. They found that expressions posed by participants with ASD were not recognized as well by TD and ASD participants as expressions posed by TD posers. Subsequently, a computational approach was used in Guha et al. (2018) to study the details of facial expressions for children with HFA. This study aimed to uncover subtle characteristics of facial expressions by analyzing localized facial dynamics and found differences in the eye region. Finally, in a meta-analysis (Trevisan et al., 2018), it was found that participants with ASD display facial expressions less frequently and for less amount time. Meanwhile, participants with ASD are less likely to share facial expressions with others or automatically mimic the expressions. These observations have partially inspired the design of an intervention system for young children with ASD, as we will elaborate later.

POSED VS. GENUINE FACIAL EXPRESSIONS OF EMOTION

Multiple databases of face stimuli have been developed for FER research (Jia et al., 2020). These databases include static images of computer-generated human faces that can be titrated to modify facial expressions or include static or dynamic images of real human faces containing posed and spontaneous facial expressions (Cassidy et al., 2015). More recently, there has arisen a question in the emotion recognition field regarding whether there is a difference between how the human brain perceives and processes emotions that are posed (artificially generated) compared to those that are genuine (spontaneously generated). One study found that adults are much more accurate at labeling emotions when the facial expression is posed than when it is spontaneous (Krumhuber et al., 2019). In this study, they also used facial recognition software to label the emotions and found the software to be more accurate than the human participants at FER for the posed emotions; however, the accuracy dropped for AI and the human participants to similar levels when the expressions of emotion were spontaneous. It was thought that this result was due to the fact that posed expressions showed more prototypical facial features of the emotions (e.g., downturned mouth and furled brow for sadness) enabling both humans and AI to learn and recognize the posed emotions with higher accuracy. Spontaneous emotional expressions have subtle, but substantial differences compared to posed expressions of emotion, with changes in small muscles and less prototypical facial expressions (Kim and Huynh, 2017). Few studies have compared FER for posed and genuine FEs with mixed results. Here, we will first highlight a few existing studies on posed vs. genuine facial expressions of emotion for ASD and then discuss our envisioned future directions along this line of research.

Recent studies had revealed differences in the literature when processing posed vs. genuine facial expressions of emotion (Pelphrey et al., 2007). There are prototypical signs exhibited for some expressions of emotion, while genuine expressions of the same emotion are more complex and harder to interpret. For example, the expression of sadness when posed includes an out-turned lower lip, though spontaneous expressions of sadness are much more highly variable and often do not include this prototypical expression (Kim and Huynh, 2017). The class of smile expressions has received special attention regarding posed vs. genuine distinction (Blampied, 2008; Boraston et al., 2008). In Blampied (2008), the sensitivity of children with ASD was compared against that of age and sex-matched control children to the different emotions underlying posed vs. genuine smiles. It was found that individuals with ASD are often less sensitive to the differences between posed and genuine smiles than TD participants. Toward deeper reasoning about this difference, it was hypothesized that experience during development viewing the eye region of a face is critical to identifying genuine smiles from posed ones. In a related study (Boraston et al., 2008), the reduced ability to discriminate genuine from posed smiles for adults with ASD is attributed to reduced eye contact. It was also found that the individuals with ASD who were more affected in recognition of genuine smiles also had more severe social interaction differences. In a recent review of studies using eye-tracking (ET) and electroencephalography (EEG) to explore FER in ASD (Black et al., 2017), they report that differences in ET and EEG result from differences in facial emotion processing that arise from functional differences in the social brain.

Evaluating posed and evoked facial expressions of emotion from adults with ASD has been studied (Faso et al., 2014). It was reported that ASD expressions were rated as more intense and less natural than TD expressions. Meanwhile, the naturalness ratings of evoked expressions were positively associated with identification accuracy for TD but not individuals with ASD. These findings collectively highlight differences in facial expressivity among ASD that do not hinder emotion recognition accuracy but may affect the quality of social interaction. Along this line of research, it has also been found that just like the failure of ASD recognize the facial expressions of TD (no matter posed or spontaneous), TD individuals also find it difficult to recognize autistic emotional expressions (Brewer et al., 2016). More recently, it has been found that neurotypical peers are less willing to interact with those with autism based on thin slice judgments (Sasson et al., 2017), and first impressions for intellectually able adults with ASD improve with diagnostic disclosure and increased autism understanding of the part of peers (Sasson and Morrison, 2019).

Considering the differences in TD accuracy for posed and spontaneous FEs, it would stand to reason that differentiating these types of stimuli in autism interventions targeting FER should be considered. Next, we propose a progressive intervention strategy inspired by research investigating posed vs. genuine expressions of emotion.

IMPLICATIONS FOR ASD INTERVENTION

While FER differences in individuals with ASD may not be universal, they are highly prevalent, and thus FER is often specifically taught as part of the autism curriculum of a child (Ayres and Robbins, 2005). Interventions have been developed that explicitly teach individuals with ASD to recognize specific emotions in others and themselves with mixed results (for a review, see Berggren et al., 2018). Stimuli for FER interventions can vary widely and may include static or dynamic images of the six basic emotions (i.e., sad, happy, angry, afraid, disgust, and surprise) as well as complex emotions, such as jealousy, that are more difficult to recognize and may require the use of contextual clues (Baron-Cohen et al., 2009). The basic goal of teaching FER to those with ASD is to help them better understand others and foster communication and social interactions (core difference areas in ASD). Previous works (Gordon et al., 2014) have focused on how to train children with ASD to produce happy and anger expressions with a computer game ("FaceMaze"). Recently, technology-based learning tools have been designed to help ASD preschoolers with FER and emotional understanding (Boccanfuso et al., 2016; Zhang et al., 2019).

Additionally, based on the observation that happiness is the easiest among the six basic emotions for encoding and decoding by humans, a computer-based tutoring system called *SmileMaze* (Cockburn et al., 2008) was designed to improve the FEE production skills of children with ASD in a dynamic and interactive format. The Computer Expression Recognition Toolbox (CERT) in *SmileMaze* is capable of automatically detecting frontal faces from a video stream and encoding each frame into 37 continuous features, including six basic facial expressions as well as 30 facial action units (AUs) as defined by the Facial Action Coding System (Ekman, 1997). Such a computational approach notably targets those characteristics in ASD that are distinct from those in TD children, which are often difficult to detect by direct visual inspection. The combination of FEE training and computer vision systems leads to the most recent work (White et al., 2018) – an automated, game-like system based on the Kinect 3D sensing technology developed by Microsoft. It has been reported that youth with ASD preferred to interact with the system more than their TD peers. Such a discovery seems to suggest that new technology-based interventions (e.g., 3D avatar-based digital twin; Wang et al., 2019), music-based therapeutic methods (Wagener et al., 2020), and computer-based recognition of posed vs. spontaneous facial expressions (Mavadati et al., 2016), have good potential in remediation of transdiagnostic processes such as FER and FEE in ASD and possibly in other disorders with facial emotion processing differences such as SZ, traumatic brain injury, and stroke. It has recently been reported in Keating and Cook (2021) that individuals with autism have difficulties recognizing neurotypical facial expressions and vice versa. TD and ASD individuals might exhibit expressive differences, but individuals with autism tend to display less frequent expressions that are rated as lower in quality by TD observers. Such observation suggests that future research should investigate what specifically is different about the facial expressions produced by ASD and TD individuals (e.g., how dynamic aspects of expressions affect emotion recognition).

Considering the scientific literature outlined in this review on FER in ASD and differences between posed and genuine facial expressions of emotion discussed above, we propose a hierarchical teaching method as part of an intervention to teach FER to individuals with ASD that considers the increased difficulty in processing more complex FER stimuli (Nuske et al., 2013). We propose three aspects for consideration when teaching FER: (1) whether the image is simple (drawings and cartoons) or complex (includes human faces or life-like artificially generated faces); (2) whether the image is static or dynamic [audio-visual (AV)]; and (3) in complex images, whether the expression of emotion is posed or genuine. Those three aspects collectively take previous findings in the literature of FER/FEE in ASD into consideration and introduce a new sequential approach toward posed vs. genuine. Compared with previous approaches such as SmileMaze (Cockburn et al., 2008) and FaceMaze (Gordon et al., 2014), ours distinguishes them by emphasizing hierarchical learning and covering more facial expressions.

We propose two possible approaches for teaching FER/FEE:

Approach (1) Teaching FER/FEE Progressively: This strategy is based on the previous finding that happiness and sadness are the least affected in ASD, but fear, surprise, and disgust are more impacted in ASD. Starting with simple, static images that include basic drawings and cartoon characters and then progressing step-wise to more complex static images with photos of human faces and expressions that are posed and genuine, and then to dynamic AV images using a life-like avatar of the therapist or child conversing in real-time as a transition between static and dynamic images of real people, and finally, real-world AV videos that contain context clues and genuine expressions of emotion. While using photos of real faces constitutes a more natural stimulus and may positively impact generalizability, the simplicity of the hand-drawn images may make them a better place to begin teaching emotions for some individuals with ASD (Sasson et al., 2008). In this vein, similar to standard Applied Behavior Analysis (ABA) methods, once they have mastered an emotion at the hand-drawn image level, it may be beneficial to move to the next level of complexity and target cartoon characters that the child enjoys. Theoretically, intervention would then move to the inclusion of stimuli with real human faces with posed emotions because posed photos are easier for typically developed individuals to label. Ultimately, using real human face stimuli with genuine, spontaneous expressions of emotion (static or dynamic) would be the ultimate target since they may be more difficult to interpret (Hanley et al., 2013). Images of the child undergoing intervention that shows him/her expressing these emotions could also be included and analysis of their facial expressions.

Approach (2) Teaching FER/FEE in a Field of Images: Alternatively, since individuals with ASD often have difficulty generalizing what they have learned in many areas, including FER (Berggren et al., 2018) and FEE (White et al., 2018), it may be best, to begin with, multiple images of a specific emotion to teach a child (e.g., in a field of drawn images, cartoon characters, and posed and random static photos of human faces expressing a target emotion). Teaching skills to individuals with ASD in a field of stimuli has been proposed previously based on the finding that repeatedly using limited stimuli increases the rigidity of thinking and reduces generalizability (Harris et al., 2015). Thus, in Approach 2, we propose to begin by teaching FER in a field that contains static images that are both simple and complex of posed and genuine expressions of a target emotion and then progress to dynamic AV FER stimuli that may contain more context clues and incorporates multisensory integration to facilitate learning (Sasson, 2006). While teaching in a field may take longer to master, research shows it may reduce learned rigidity of thought and improve generalizability. Finally, incorporating these stimuli into games that are enjoyable to play (see above referenced FER/FEE interventions), and could be customized so that the interventionist can select the images at each level of FER/FEE functioning, could facilitate facial emotion training in some individuals with ASD.

While the intrinsic social motivations of a child may not significantly impact how FER/FEE is taught (Garman et al., 2016), delivering the stimuli in a fun and intrinsically motivating way could improve generalizability (Baron-Cohen et al., 2009). A feasibility study was conducted by White et al. (2018) of their system developed to teach FEE to children with ASD. The system provides critical feedback to the child *via* computer analysis of the facial expression a child made in response to a cue. Such a system could be used in conjunction with a FER/FEE training program since the ability of a child to recognize their own emotions may likely facilitate FER/FEE learning and thus may be a framework upon which recognition of others' emotions can be built (Manfredonia et al., 2019; Ola and Gullon-Scott, 2020). Additionally, avatars can be created to interact in real-time with a child and may provide an added opportunity for a person with ASD to initiate conversations of their own accord, as has been seen at Disney World where children with ASD willingly interact with an avatar of Crush the turtle from the movie *Nemo* (Carter et al., 2014). Regarding the dynamic AV stimuli, since multisensory integration has been shown to enhance our ability to learn new information (Shams and Seitz, 2008), incorporating auditory input with visual input may facilitate the ability for individuals with ASD to learn emotion recognition, especially at the more complex levels of FER/FEE as in, where the stimuli would be considered the most complex (real-world, AV, and genuine expressions of emotion). Consideration should be given to the level of functioning of an individual in face/emotion processing and learning style when determining where to begin and whether to teach progressively (Approach 1) or to teach in a field (Approach 2) of static images and then progress to dynamic AV videos.

Additionally, the scientific community has developed multiple datasets of face stimuli for research purposes to investigate how FER/FEE is perceived in TD, ASD, and other disorders (for a review of FER databases, see Jia et al., 2020). These stimuli have static and dynamic expressions of emotion that are often well titrated (morphed levels between two emotions), but these stimuli are generally not known to autism therapists and are not utilized by them for teaching FER/FEE. Thus, the availability of face stimuli for teaching is often dependent upon the funds available to an interventionist. Therapists have been very creative and find free face stimuli to use when teaching their students, which can benefit children when various images are used. However, this can be time-consuming and costly, especially if therapists must purchase images from different datasets to acquire a set of images for teaching a specific emotion. Therefore, we propose that interventionists take advantage of the variety of FER datasets that include both posed and genuine expressions of emotion and dynamic videos of facial expressions of emotions.

Many FER/FEE databases have been developed using the six basic emotions that were found to be universal (Ekman, 1970) and the Face Action Coding System (FACS) that breaks down movements of muscles in the face used to make expressions of emotion into AUs (Ekman, 1997). These same AUs are the primary measures for facial expressions used by entities like Disney to animate characters to make their facial movements more realistic. Thus, using stimuli to teach FER/FEE in those with ASD that has incorporated realistic portrayals of human emotions (e.g., Disney characters) and analyses of human expressions of emotion based on these same measurements of facial micromovements (Leo et al., 2019) would bring full circle the application of the research investigating this critical aspect of human existence to help those who struggle in this area. Lastly, avatar software developed by companies like ObEN can benefit FER intervention by enabling the creation of life-like avatars of a therapist or of the person with ASD,[1] which may help individuals with ASD to transition between static images and dynamic real-world videos that contain context clues and possibly help them better understand their own expressions of emotion.

The proposed FETH method requires research to investigate the merits of teaching FER/FEE serially (Approach 1) or teaching in a field of images at different levels of complexity to improve generalizability (Approach 2). Regardless, a more refined FER/FEE intervention based on current scientific outcomes has far-reaching implications for children and adults with ASD and other disorders where FER/FEE difficulties can significantly hinder social interactions, including SZ, stroke, and traumatic brain injury.

[1] https://oben.me

AUTHOR CONTRIBUTIONS

PW, SW, and XL contributed to the research, analysis, and writing of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by an NSF CAREER Award (BCS-1945230), Air Force Young Investigator Program Award (FA9550-21-1-0088), Dana Foundation Clinical Neuroscience Award, ORAU Ralph E. Powe Junior Faculty Enhancement Award, West Virginia University (WVU), and WVU PSCoR Program (to SW), and an NSF grant (OAC-1839909) and the WV Higher Education Policy Commission grant (HEPC.dsr.18.5; to XL).

ACKNOWLEDGMENTS

The authors want to thank two reviewers for improving the literary presentation of this work and Ms. Amber Li for helping the preparation of the references.

REFERENCES

Adolphs, R. (2008). Fear, faces, and the human amygdala. *Curr. Opin. Neurobiol.* 18, 166–172. doi: 10.1016/j.conb.2008.06.006

Adolphs, R., Sears, L., and Piven, J. (2001). Abnormal processing of social information from faces in autism. *J. Cogn. Neurosci.* 13, 232–240. doi: 10.1162/089892901564289

Ayres, A. J., and Robbins, J. (2005). *Sensory Integration and the Child: Understanding Hidden Sensory Challenges*. Los Angeles: Western Psychological Services.

Baron-Cohen, S., Golan, O., and Ashwin, E. (2009). Can emotion recognition be taught to children with autism spectrum conditions? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3567–3574. doi: 10.1098/rstb.2009.0191

Baron-Cohen, S., Ring, H. A., Bullmore, E. T., Wheelwright, S., Ashwin, C., and Williams, S. C. R. (2000). The amygdala theory of autism. *Neurosci. Biobehav. Rev.* 24, 355–364. doi: 10.1016/S0149-7634(00)00011-7

Begeer, S., Koot, H. M., Rieffe, C., Meerum Terwogt, M., and Stegge, H. (2008). Emotional competence in children with autism: diagnostic criteria and empirical evidence. *Dev. Rev.* 28, 342–369. doi: 10.1016/j.dr.2007.09.001

Berggren, S., Fletcher-Watson, S., Milenkovic, N., Marschik, P. B., Bölte, S., and Jonsson, U. (2018). Emotion recognition training in autism spectrum disorder: a systematic review of challenges related to generalizability. *Dev. Neurorehabil.* 21, 141–154. doi: 10.1080/17518423.2017.1305004

Black, M. H., Chen, N. T. M., Iyer, K. K., Lipp, O. V., Bölte, S., Falkmer, M., et al. (2017). Mechanisms of facial emotion recognition in autism spectrum disorders: insights from eye tracking and electroencephalography. *Neurosci. Biobehav. Rev.* 80, 488–515. doi: 10.1016/j.neubiorev.2017.06.016

Black, M. H., Chen, N. T., Lipp, O. V., Bölte, S., and Girdler, S. (2020). Complex facial emotion recognition and atypical gaze patterns in autistic adults. *Autism* 24, 258–262. doi: 10.1177/1362361319856969

Blampied, M. (2008). Are children with Autism Spectrum Disorder sensitive to the different emotions underlying posed and genuine smiles? MS Thesis. University of Canterbury.

Boccanfuso, L., Barney, E., Foster, C., Ahn, Y.A., Chawarska, K., Scassellati, B., et al. (2016). "Emotional robot to examine different play patterns and affective responses of children with and without ASD," in *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*; March, 2016; 19–26.

Bons, D., Scheepers, F. E., Rommelse, N. N. J., and Buitelaar, J. K. (2011). "Motor, emotional, and cognitive empathic abilities in children with autism and conduct disorder," in *Proceedings of the ACM International Conference Proceeding Series*; August, 2011; 109–113.

Boraston, Z. L., Corden, B., Miles, L. K., Skuse, D. H., and Blakemore, S.-J. (2008). Brief report: perception of genuine and posed smiles by individuals with autism. *J. Autism Dev. Disord.* 38, 574–580. doi: 10.1007/s10803-007-0421-1

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., et al. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron* 17, 875–887. doi: 10.1016/S0896-6273(00)80219-6

Brewer, R., Biotti, F., Catmur, C., Press, C., Happé, F., Cook, R., et al. (2016). Can neurotypical individuals read autistic facial expressions? Atypical production of emotional facial expressions in Autism Spectrum Disorders. *Autism Res.* 9, 262–271. doi: 10.1002/aur.1508

Carter, E. J., Williams, D. L., Hodgins, J. K., and Lehman, J. F. (2014). Are children with autism more responsive to animated characters? A study of interactions with humans and human-controlled avatars. *J. Autism Dev. Disord.* 44, 2475–2485. doi: 10.1007/s10803-014-2116-8

Cassidy, S., Mitchell, P., Chapman, P., and Ropar, D. (2015). Processing of spontaneous emotional responses in adolescents and adults with autism spectrum disorders: effect of stimulus type. *Autism Res.* 8, 534–544. doi: 10.1002/aur.1468

Cassidy, S., Ropar, D., Mitchell, P., and Chapman, P. (2014). Can adults with autism spectrum disorders infer what happened to someone from their emotional response? *Autism Res.* 7, 112–123. doi: 10.1002/aur.1351

Celani, G., Battacchi, M. W., and Arcidiacono, L. (1999). The understanding of the emotional meaning of facial expressions in people with autism. *J. Autism Dev. Disord.* 29, 57–66. doi: 10.1023/A:1025970600181

Cockburn, J., Bartlett, M., Tanaka, J., Movellan, J., Pierce, M., and Schultz, R. (2008). "SmileMaze: A Tutoring System in Real-Time Facial Expression Perception and Production for Children with Autism Spectrum Disorder," in *ECAG 2008 Workshop: Facial and Bodily Expressions for Control and Adaptation of Games*; September, 2008; 3–9.

Crompton, C. J., Ropar, D., Evans-Williams, C. V., Flynn, E. G., and Fletcher-Watson, S. (2020). Autistic peer-to-peer information transfer is highly effective. *Autism* 24, 1704–1712. doi: 10.1177/1362361320919286

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., et al. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nat. Neurosci.* 8, 519–526. doi: 10.1038/nn1421

Dawson, G., Webb, S. J., Carver, L., Panagiotides, H., and McPartland, J. (2004). Young children with autism show atypical brain responses to fearful versus neutral facial expressions of emotion. *Dev. Sci.* 7, 340–359. doi: 10.1111/j.1467-7687.2004.00352.x

Ekman, P. (1970). Universal facial expressions of emotion. *Cal. Ment. Health* 8, 151–158.

Ekman, R. (1997). *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. USA: Oxford University Press.

Faso, D. J., Sasson, N. J., and Pinkham, A. E. (2014). Evaluating posed and evoked facial expressions of emotion from adults with Autism Spectrum Disorder. *J. Autism Dev. Disord.* 45, 75–89. doi: 10.1007/s10803-014-2194-7

Fridenson-Hayo, S., Berggren, S., Lassalle, A., Tal, S., Pigat, D., Bölte, S., et al. (2016). Basic and complex emotion recognition in children with autism: cross-cultural findings. *Mol. Autism.* 7, 1–11. doi: 10.1186/s13229-016-0113-9

Garman, H. D., Spaulding, C. J., Webb, S. J., Mikami, A. Y., Morris, J. P., and Lerner, M. D. (2016). Wanting it too much: an inverse relation between social motivation and facial emotion recognition in autism spectrum disorder. *Child Psychiatry Hum. Dev.* 47, 890–902. doi: 10.1007/s10578-015-0620-5

Golan, O., Sinai-Gavrilov, Y., and Baron-Cohen, S. (2015). The Cambridge mindreading face-voice battery for children (CAM-C): complex emotion recognition in children with and without autism spectrum conditions. *Mol. Autism* 6, 1–9. doi: 10.1186/s13229-015-0018-z

Gordon, I., Pierce, M. D., Bartlett, M. S., and Tanaka, J. W. (2014). Training facial expression production in children on the autism spectrum. *J. Autism Dev. Disord.* 44, 2486–2498. doi: 10.1007/s10803-014-2118-6

Guha, T., Yang, Z., Grossman, R. B., and Narayanan, S. S. (2018). A computational study of expressive facial dynamics in children with autism. *IEEE Trans. Affect. Comput.* 9, 14–20. doi: 10.1109/TAFFC.2016.2578316

Hall, G. B. C., Szechtman, H., and Nahmias, C. (2003). Enhanced salience and emotion recognition in autism: a PET study. *Am. J. Psychiatry* 160, 1439–1441. doi: 10.1176/appi.ajp.160.8.1439

Hanley, M., McPhillips, M., Mulhern, G., and Riby, D. M. (2013). Spontaneous attention to faces in Asperger syndrome using ecologically valid static stimuli. *Autism* 17, 754–761. doi: 10.1177/1362361312456746

Harms, M. B., Martin, A., and Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. *Neuropsychol. Rev.* 20, 290–322. doi: 10.1007/s11065-010-9138-6

Harris, H., Israeli, D., Minshew, N., Bonneh, Y., Heeger, D. J., Behrmann, M., et al. (2015). Perceptual learning in autism: over-specificity and possible remedies. *Nat. Neurosci.* 18, 1–4. doi: 10.1038/nn.4129

Hess, U., Kappas, A., McHugo, G. J., Kleck, R. E., and Lanzetta, J. T. (1989). An analysis of the encoding and decoding of spontaneous and posed smiles: the use of facial electromyography. *J. Nonverbal Behav.* 13, 121–137. doi: 10.1007/BF00990794

Hobson, R. P. (1986). The autistic child's appraisal of expressions of emotion. *J. Child Psychol. Psychiatry* 27, 321–342. doi: 10.1111/j.1469-7610.1986.tb01836.x

Hobson, R. P., Ouston, J., and Lee, A. (1988). What's in a face? The case of autism. *Br. J. Psychol.* 79, 441–453. doi: 10.1111/j.2044-8295.1988.tb02745.x

Jaswal, V. K., and Akhtar, N. (2019). Being versus appearing socially uninterested: challenging assumptions about social motivation in autism. *Behav. Brain Sci.* 1–84. doi: 10.1017/S0140525X18001826 [Epub ahead of print]

Jia, S., Wang, S., Hu, C., Webster, P., and Li, X. (2020). Detection of genuine and posed facial expressions of emotion: databases and methods. *Front. Psychol.* 11:580287. doi: 10.3389/fpsyg.2020.580287

Keating, C. T., and Cook, J. L. (2021). Facial expression production and recognition in autism spectrum disorders: a shifting landscape. *Psychiatr. Clin.* 44, 125–139. doi: 10.1016/j.psc.2020.11.010

Kennedy, D. P., and Adolphs, R. (2013). Perception of emotions from facial expressions in high-functioning adults with autism. *Neuropsychologia* 50, 3313–3319. doi: 10.1016/j.neuropsychologia.2012.09.038

Kim, Y. G., and Huynh, X.-P. (2017). "Discrimination between Genuine Versus Fake Emotion Using Long-Short Term Memory with Parametric Bias and Facial Landmarks," in *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW)*; October, 2017; Venice, 3065–3072.

Kliemann, D., Dziobek, I., Hatri, A., Baudewig, J., and Heekeren, H. R. (2012). The role of the amygdala in atypical gaze on emotional faces in autism spectrum disorders. *J. Neurosci.* 32, 9469–9476. doi: 10.1523/JNEUROSCI.5294-11.2012

Krumhuber, E. G., Küster, D., Namba, S., Shah, D., and Calvo, M. G. (2019). Emotion recognition from posed and spontaneous dynamic expressions: human observers versus machine analysis. *Emotion* 21, 447–451. doi: 10.1037/emo0000712

Langdell, T. (1978). Recognition of faces: an approach to the study of autism. *J. Child Psychol. Psychiatry* 19, 255–268. doi: 10.1111/j.1469-7610.1978.tb00468.x

Law Smith, M. J., Montagne, B., Perrett, D. I., Gill, M., and Gallagher, L. (2010). Detecting subtle facial emotion recognition deficits in high-functioning autism using dynamic stimuli of varying intensities. *Neuropsychologia* 48, 2777–2781. doi: 10.1016/j.neuropsychologia.2010.03.008

Leo, M., Carcagnì, P., Distante, C., Mazzeo, P. L., Spagnolo, P., Levante, A., et al. (2019). Computational analysis of deep visual data for quantifying facial expression production. *Appl. Sci.* 9:4542. doi: 10.3390/app9214542

Loveland, K. A., Tunali-Kotoski, B., Pearson, D. A., Brelsford, K. A., Ortegon, J., and Chen, R. (1994). Imitation and expression of facial affect in autism. *Dev. Psychopathol.* 6, 433–444. doi: 10.1017/S0954579400006039

Lundqvist, L. O. (2015). Hyper-responsiveness to touch mediates social dysfunction in adults with autism spectrum disorders. *Res. Autism Spectr. Disord.* 9, 13–20. doi: 10.1016/j.rasd.2014.09.012

Manfredonia, J., Bangerter, A., Manyakov, N. V., Ness, S., Lewin, D., Skalkin, A., et al. (2019). Automatic recognition of posed facial expression of emotion in individuals with Autism Spectrum Disorder. *J. Autism Dev. Disord.* 49, 279–293. doi: 10.1007/s10803-018-3757-9

Martínez, A., Tobe, R., Dias, E. C., Ardekani, B. A., Veenstra-VanderWeele, J., Patel, G., et al. (2019). Differential patterns of visual sensory alteration underlying face emotion recognition impairment and motion perception deficits in Schizophrenia and Autism Spectrum Disorder. *Biol. Psychiatry* 86, 557–567. doi: 10.1016/j.biopsych.2019.05.016

Mavadati, M., Sanger, P., and Mahoor, M. H. (2016). "Extended DISFA Dataset: investigating posed and spontaneous facial expressions," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*; June, 2016; 1452–1459.

Mesibov, G. B. (1984). Social skills training with verbal autistic adolescents and adults: a program model. *J. Autism Dev. Disord.* 14, 395–404. doi: 10.1007/BF02409830

Milton, D. E. (2012). On the ontological status of autism: the "double empathy problem". *Disabil. Soc.* 27, 883–887. doi: 10.1080/09687599.2012.710008

Milton, D., and Sims, T. (2016). How is a sense of well-being and belonging constructed in the accounts of autistic adults? *Disabil. Soc.* 31, 520–534. doi: 10.1080/09687599.2016.1186529

Monk, C. S., Weng, S. J., Wiggins, J. L., Kurapati, N., Louro, H. M. C., Carrasco, M., et al. (2010). Neural circuitry of emotional face processing in autism spectrum disorders. *J. Psychiatry Neurosci.* 35, 105–114. doi: 10.1503/jpn.090085

Naab, P. J., and Russell, J. A. (2007). Judgments of emotion from spontaneous facial expressions of new Guineans. *Emotion* 7, 736–744. doi: 10.1037/1528-3542.7.4.736

Nuske, H. J., Vivanti, G., and Dissanayake, C. (2013). Are emotion impairments unique to, universal, or specific in autism spectrum disorder? A comprehensive review. *Cognit. Emot.* 27, 1042–1061. doi: 10.1080/02699931.2012.762900

Ola, L., and Gullon-Scott, F. (2020). Facial emotion recognition in autistic adult females correlates with alexithymia, not autism. *Autism* 24, 2021–2034. doi: 10.1177/1362361320932727

Ozonoff, S., Pennington, B. F., and Rogers, S. J. (1990). Are there emotion perception deficits in young autistic children? *J. Child Psychol. Psychiatry* 31, 343–361. doi: 10.1111/j.1469-7610.1990.tb01574.x

Park, S., Lee, K., Lim, J. A., Ko, H., Kim, T., Lee, J. I., et al. (2020). Differences in facial expressions between spontaneous and posed smiles: automated method by action units and three-dimensional facial landmarks. *Sensors* 20:1199. doi: 10.3390/s20041199

Pelphrey, K. A., and Carter, E. J. (2008). Brain mechanisms for social perception: from autism and typical development. *Ann. N. Y. Acad. Sci.* 1145, 283–299. doi: 10.1196/annals.1416.007

Pelphrey, K. A., Morris, J. P., and McCarthy, G. (2005). Neural basis of eye gaze processing deficits in autism. *Brain* 128, 1038–1048. doi: 10.1093/brain/awh404

Pelphrey, K. A., Morris, J. P., McCarthy, G., and Labar, K. S. (2007). Perception of dynamic changes in facial affect and identity in autism. *Soc. Cogn. Affect. Neurosci.* 2, 140–149. doi: 10.1093/scan/nsm010

Pelphrey, K., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., and Piven, J. (2002). Visual scanning of faces in Autism. *J. Autism Dev. Disord.* 32, 249–261. doi: 10.1023/A:1016374617369

Rutishauser, U., Tudusciuc, O., Wang, S., Mamelak, A. N., Ross, I. B., and Adolphs, R. (2013). Single-neuron correlates of atypical face processing in autism. *Neuron* 80, 887–899. doi: 10.1016/j.neuron.2013.08.029

Sachse, M., Schlitt, S., Hainz, D., Ciaramidaro, A., Walter, H., Poustka, F., et al. (2014). Facial emotion recognition in paranoid schizophrenia and autism spectrum disorder. *Schizophr. Res.* 159, 509–514. doi: 10.1016/j.schres.2014.08.030

Sasson, N. J. (2006). The development of face processing in autism. *J. Autism Dev. Disord.* 36, 381–394. doi: 10.1007/s10803-006-0076-3

Sasson, N. J., Faso, D. J., Nugent, J., Lovell, S., Kennedy, D. P., and Grossman, R. B. (2017). Neurotypical peers are less willing to interact with those with Autism based on thin slice judgments. *Sci. Rep.* 7, 1–10. doi: 10.1038/srep40700

Sasson, N. J., and Morrison, K. E. (2019). First impressions of adults with autism improve with diagnostic disclosure and increased autism knowledge of peers. *Autism* 23, 50–59. doi: 10.1177/1362361317729526

Sasson, N. J., Turner-Brown, L. M., Holtzclaw, T. N., Lam, K. S., and Bodfish, J. W. (2008). Children with autism demonstrate circumscribed attention during passive viewing of complex social and nonsocial picture arrays. *Autism Res.* 1, 31–42. doi: 10.1002/aur.4

Sauter, D. A., and Fischer, A. H. (2018). Can perceivers recognise emotions from spontaneous expressions? *Cognit. Emot.* 32, 504–515. doi: 10.1080/02699931.2017.1320978

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

Shams, L., and Seitz, A. R. (2008). Benefits of multisensory learning. *Trends Cogn. Sci.* 12, 411–417. doi: 10.1016/j.tics.2008.07.006

Sheppard, E., Pillai, D., Wong, G. T. L., Ropar, D., and Mitchell, P. (2016). How easy is it to read the minds of people with autism spectrum disorder? *J. Autism Dev. Disord.* 46, 1247–1254. doi: 10.1007/s10803-015-2662-8

Simões, M., Monteiro, R., Andrade, J., Mouga, S., França, F., Oliveira, G., et al. (2018). A novel biomarker of compensatory recruitment of face emotional imagery networks in autism spectrum disorder. *Front. Neurosci.* 12:791. doi: 10.3389/fnins.2018.00791

Sucksmith, E., Allison, C., Baron-Cohen, S., Chakrabarti, B., and Hoekstra, R. A. (2013). Empathy and emotion recognition in people with autism, first-degree relatives, and controls. *Neuropsychologia* 51, 98–105. doi: 10.1016/j.neuropsychologia.2012.11.013

Tang, J. S., Chen, N. T., Falkmer, M., Bölte, S., and Girdler, S. (2019). Atypical visual processing but comparable levels of emotion recognition in adults with autism during the processing of social scenes. *J. Autism Dev. Disord.* 49, 4009–4018. doi: 10.1007/s10803-019-04104-y

Thye, M. D., Bednarz, H. M., Herringshaw, A. J., Sartin, E. B., and Kana, R. K. (2018). The impact of atypical sensory processing on social impairments in autism spectrum disorder. *Dev. Cogn. Neurosci.* 29, 151–167. doi: 10.1016/j.dcn.2017.04.010

Trevisan, D. A., Hoskyn, M., and Birmingham, E. (2018). Facial expression production in autism: a meta-analysis. *Autism Res.* 11, 1586–1601. doi: 10.1002/aur.2037

Uljarevic, M., and Hamilton, A. (2013). Recognition of emotions in autism: a formal meta-analysis. *J. Autism Dev. Disord.* 43, 1517–1526. doi: 10.1007/s10803-012-1695-5

Van Der Geld, P., Oosterveld, P., Bergé, S. J., and Kuijpers-Jagtman, A. M. (2008). Tooth display and lip position during spontaneous and posed smiling in adults. *Acta Odontol. Scand.* 66, 207–213. doi: 10.1080/00016350802060617

Volker, M. A., Lopata, C., Smith, D. A., and Thomeer, M. L. (2009). Facial encoding of children with high-functioning autism spectrum disorders. *Focus Autism Other Dev. Disabil.* 24, 195–204. doi: 10.1177/1088357609347325

Wagener, G. L., Berning, M., Costa, A. P., Steffgen, G., and Melzer, A. (2020). Effects of emotional music on facial emotion recognition in children with Autism Spectrum Disorder (ASD). *J. Autism Dev. Disord.* 1–10. doi: 10.1007/s10803-020-04781-0 [Epub ahead of print]

Wang, R., Chen, C. F., Peng, H., Liu, X., Liu, O., and Li, X. (2019). Digital Twin: acquiring high-fidelity 3D avatar from a single image. ArXiv [Preprint].

Wang, S., Wu, C., He, M., Wang, J., and Ji, Q. (2015). Posed and spontaneous expression recognition through modeling their spatial patterns. *Mach. Vis. Appl.* 26, 219–231. doi: 10.1007/s00138-015-0657-2

Weeks, S. J., and Hobson, R. P. (1987). The salience of facial expression for autistic children. *J. Child Psychol. Psychiatry* 28, 137–152. doi: 10.1111/j.1469-7610.1987.tb00658.x

White, S. W., Abbott, L., Wieckowski, A. T., Capriola-Hall, N. N., Aly, S., and Youssef, A. (2018). Feasibility of automated training for facial emotion expression and recognition in autism. *Behav. Ther.* 49, 881–888. doi: 10.1016/j.beth.2017.12.010

Wong, N., Beidel, D. C., Sarver, D. E., and Sims, V. (2012). Facial emotion recognition in children with high functioning autism and children with social phobia. *Child Psychiatry Hum. Dev.* 43, 775–794. doi: 10.1007/s10578-012-0296-z

Zhang, S., Xia, X., Li, S., Shen, L., Liu, J., Zhao, L., et al. (2019). Using technology-based learning tool to train facial expression recognition and emotion understanding skills of Chinese preschoolers with autism spectrum disorder. *Int. J. Dev. Disabil.* 65, 378–386. doi: 10.1080/20473869.2019.1656384

Zhou, H., Cai, X., Weigl, M., Bang, P., Cheung, E. F. C., and Chan, R. C. K. (2018). Multisensory temporal binding window in autism spectrum disorders and schizophrenia spectrum disorders: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* 86, 66–76. doi: 10.1016/j.neubiorev.2017.12.013

Conflict of Interest: 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.

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: frontiersin.org/about/contact

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