scientific reports

OPEN

Sensing emotional valence and arousal dynamics through automated facial action unit analysis

Junyao Zhang[1], Wataru Sato[2]✉, Naoya Kawamura[1], Koh Shimokawa[2], Budu Tang[1] & Yuichi Nakamura[3]

Information about the concordance between dynamic emotional experiences and objective signals is practically useful. Previous studies have shown that valence dynamics can be estimated by recording electrical activity from the muscles in the brows and cheeks. However, whether facial actions based on video data and analyzed without electrodes can be used for sensing emotion dynamics remains unknown. We investigated this issue by recording video of participants' faces and obtaining dynamic valence and arousal ratings while they observed emotional films. Action units (AUs) 04 (i.e., brow lowering) and 12 (i.e., lip-corner pulling), detected through an automated analysis of the video data, were negatively and positively correlated with dynamic ratings of subjective valence, respectively. Several other AUs were also correlated with dynamic valence or arousal ratings. Random forest regression modeling, interpreted using the SHapley Additive exPlanation tool, revealed non-linear associations between the AUs and dynamic ratings of valence or arousal. These results suggest that an automated analysis of facial expression video data can be used to estimate dynamic emotional states, which could be applied in various fields including mental health diagnosis, security monitoring, and education.

Keywords Facial action units, Emotional valence/arousal dynamics, Automated video analysis, Machine learning, SHapley Additive exPlanation

Understanding the concordance between dynamic emotional experiences and objective signals is pivotal in deciphering human emotions. Previous psychological studies have shown that emotional experiences fluctuate from moment to moment[1–4], and are linked to mental health and psychological disorders[5]. Because emotional experiences are private, and the continuous assessment of emotional experiences is difficult while conducting other tasks[6], dynamic sensing of emotions using objective signals can be useful, laying the foundation for practical applications in fields ranging from mental health diagnostics to enhanced interpersonal communication.

Several psychophysiological studies have shown that the dynamics of emotional valence can be reliably estimated by recording facial electromyography (EMG) from the corrugator supercilii muscle (related to frowning) and zygomatic major muscle (related to smiling)[7–10]. For example, one study[9] measured changes in emotional valence ratings and EMG activity of the corrugator supercilii and zygomatic major muscles while participants watched emotional films. The results showed that the EMG activity of the corrugator supercilii and zygomatic major muscles was negatively and positively correlated with the dynamic valence ratings, respectively. Such results suggest associations between facial EMG signals and the dynamic experiences of emotional valence.

However, whether facial actions based on video data and analyzed without electrodes can be used to sense emotion dynamics remains unknown. As recording facial EMG requires placing electrodes on the face, which may change the appearance of the participant[11], non-contact recording is more desirable in practical situations. Some studies have shown correspondence between facial EMG signals recorded from the brow and cheek muscles and brow and cheek actions analyzed from video data[12]. However, the facial muscle activities detected using facial EMG signals are subtle and invisible[13]. To the best of our knowledge, no study has investigated the associations

[1]Graduate School of Informatics, Kyoto University, Yoshida-Honmachi, Sakyo, Kyoto 606-8507, Japan. [2]Psychological Process Research Team, Guardian Robot Project, RIKEN, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan. [3]Academic Center for Computing and Media Studies, Kyoto University, Yoshida-Honmachi, Sakyo, Kyoto 606-8507, Japan. ✉email: wataru.sato.ya@riken.jp

between visually detectable facial actions and dynamic emotional ratings. However, a recent study acquired facial expression video data while participants remembered personal events varying in emotional valence and arousal levels[14]. Automated analysis of the Facial Action Coding System (FACS)[15,16] revealed that the action units (AUs) of brow lowering (i.e., AU 04) and lip-corner pulling (i.e., AU 12) associated with memory-based emotional experiences are also negatively and positively associated with valence. Thus, we hypothesized that AUs 04 and 12, detected through an automated analysis of video data, would be negatively and positively associated with dynamic ratings of subjective valence, respectively.

Additionally, whether automated AU analysis results have non-linear associations with subjective valence or arousal dynamics remains untested. A recent study applied non-linear machine learning (ML) models, including random forest (RF) regression, to dynamic valence ratings and facial EMG data[17]. The researchers reported that the ML analysis had better predictive performance in terms of the valence ratings (i.e., higher positive correlations between actual and predicted valence ratings) than linear analysis. Furthermore, SHapley Additive exPlanation (SHAP) analysis, which aids interpretation of ML models given their black box nature[18], revealed non-linear associations between subjective valence ratings and facial EMG activity. Thus, we hypothesized that ML modeling and SHAP analysis would reveal the non-linear associations between dynamic ratings of valence or arousal and automated AU data.

To investigate these hypotheses, we recorded videos of participants' faces and obtained dynamic valence and arousal ratings while they observed emotional films. We presented five film clips, categorically labeled as anger, sadness, neutral, contentment, and amusement, which reportedly show linear and quadratic relationships with subjective valence and arousal ratings, respectively[9]. First, the participants viewed the clips and provided one-shot subjective valence and arousal ratings. Throughout this process, videos capturing full-face views of the participants were continuously recorded; these were subsequently analyzed using validated software to automatically extract the AU intensities[20]. Following the initial presentation, the film clips were shown to the participants twice more. During these latter viewings, participants recalled and dynamically rated their emotional experience during the first viewing in terms of valence or arousal using a slider-type affect rating dial[21]. This cued-recall approach was employed to acquire two different types of dynamic ratings (i.e., valence and arousal) that were difficult to simultaneously assess during the first viewing. Furthermore, online ratings were not used because online monitoring of subjective experiences can affect the naturalness of facial or subjective emotional responses[22,23]. Additionally, because previous studies have demonstrated strong positive correlations between cued-recall and online ratings for emotional films[7,24], we expected that the cued-recall ratings would be characterized by subjective emotional dynamics comparable with those of online ratings. We analyzed the correlations of second-by-second dynamic valence or arousal ratings with AU intensities across time; the data of all film conditions were concatenated, as in a previous EMG study[9]. In addition to AUs 04 and 12, we explored 18 other AUs that could be automatically coded by the software. We also performed RF regression modeling and SHAP analysis of the relationships between the dynamic ratings and the AUs.

Results

Subjective ratings

The one-shot and dynamic ratings were evaluated as indicators of subjective emotion elicitation (Table 1 and Fig. 1). One-shot ratings reflect the overall emotional experience, and dynamic ratings reflect emotional states on a moment-to-moment basis. The group mean dynamic valence and arousal ratings in Fig. 1 indicate that the emotional film clips elicited dynamic changes in subjective emotional experience. Planned contrasts confirmed that the one-shot ratings and mean dynamic ratings acquired during film presentation reflected the linear and quadratic patterns of the valence and arousal ratings across films, respectively, as expected (one-shot valence: $F[1,22] = 227.82$, $p < 0.001$, $\eta^2_p = 0.91$; one-shot arousal: $F[1,22] = 109.04$, $p < 0.001$, $\eta^2_p = 0.83$; dynamic valence: $F[1,22] = 167.47$, $p < 0.001$, $\eta^2_p = 0.88$; dynamic arousal: $F[1,22] = 155.29$, $p < 0.001$, $\eta^2_p = 0.88$).

AUs

To illustrate the differences in facial expressions elicited by different emotional stimuli, two different visualization methods are used in Fig. 2 to show how the mean values and standard errors (*SE*s) of specific facial AUs varied over time during the viewing of the five films.

	One-shot				Dynamic			
	Valence		Arousal		Valence		Arousal	
Film emotion	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Anger	2.00	0.23	7.26	0.34	4.26	0.23	6.40	0.26
Sadness	3.52	0.29	7.13	0.34	3.57	0.29	6.27	0.25
Neutral	4.26	0.27	3.87	0.42	4.37	0.22	3.82	0.32
Contentment	6.09	0.28	4.17	0.25	5.84	0.21	4.43	0.22
Amusement	7.61	0.23	7.50	0.19	6.72	0.19	6.52	0.22

Table 1. Mean [with standard error (*SE*)] one-shot and dynamic ratings of valence and arousal across five emotional films.

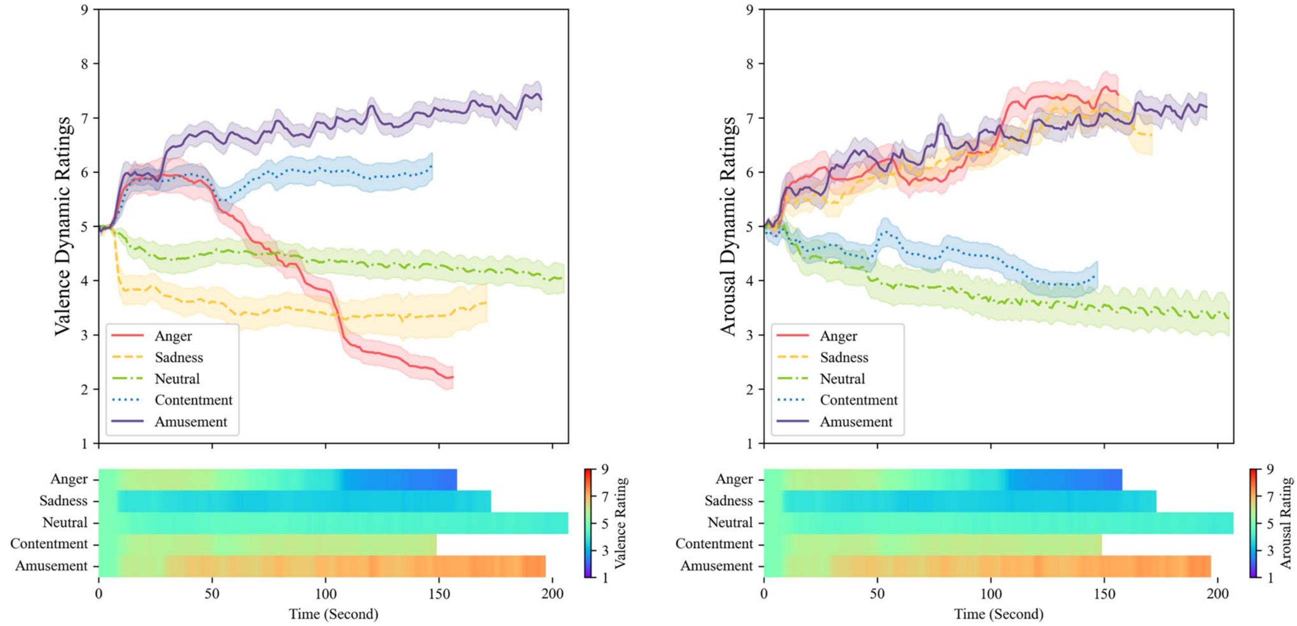

Fig. 1. The mean second-by-second subjective dynamic ratings of valence (left) and arousal (right). In the upper panels, solid-colored lines represent the mean values, and shaded areas of the same color indicate the standard error. In the lower panels, redder hues in the spectrograms correspond to higher levels of valence/arousal, and more purple hues correspond to lower levels of valence/arousal.

Correlations between subjective ratings and AUs

Pearson's correlation coefficients were calculated between the valence/arousal ratings and AUs summed across films for each participant as measures of intra-individual subjective–facial associations (Fig. 3). First, based on our research interests, we conducted a priori analyses of the associations of valence ratings with AUs 04 and 12. The r-values, summed across films, were subjected to one-sample t-tests against zero after Fisher's z transformation (Table 2). The results indicated a significant negative correlation between valence and AU 04 ($t[22] = 2.16$, $p = 0.042$, $d = 0.45$) and a significant positive correlation between valence and AU 12 ($t[22] = 8.55$, $p < 0.001$, $d = 1.78$).

Next, we conducted exploratory analyses of the associations between the valence or arousal ratings and all 20 analyzed AUs using Hotelling's one-sample T^2 test and follow-up univariate one-sample t-tests. Hotelling's one-sample T^2 test revealed a significant association between the valence ratings and AUs ($T^2[3, 20] = 4610.67$, $p = 0.008$, $\eta_p^2 = 0.99$). Follow-up univariate tests indicated significant negative correlations for AUs 01, 04, 15, and 18, and significant positive correlations for AUs 06, 07, 09, 12, 20, and 25 ($t[22] > 2.14$, $p < 0.05$, $d > 0.44$).

Similarly, Hotelling's one-sample T^2 test revealed a significant association between arousal ratings and AUs ($T^2[3, 20] = 13{,}109.04$, $p = 0.002$, $\eta_p^2 = 1.00$). Follow-up univariate tests indicated a significant negative correlation for AU 43 and significant positive correlations for AUs 02, 06, 07, 09, 12, and 17 ($t[22] > 2.34$, $p < 0.03$, $d > 0.48$).

Figure 4 illustrates the AUs showing significant correlations with valence or arousal ratings.

ML analysis for the associations between subjective ratings and AUs

To evaluate the predictive performance of the RF and linear models, we calculated correlation coefficients between the actual and predicted values. Leave-one-out cross-validation was employed, where the data from one participant served as the evaluation dataset and the data of the other participants comprised the training dataset, which was used to train the models, a correlation coefficient was computed for each individual participant serving as the evaluation dataset. The mean ± SE correlation coefficients between the actual and predicted valence and arousal ratings in the RF model were 0.42 ± 0.05, and 0.29 ± 0.06, respectively, compared with 0.43 ± 0.06 and 0.28 ± 0.07, respectively, in the linear model. One-sample t-tests revealed that all models had correlation coefficients significantly greater than zero ($t[22] = 6.56$, 4.01, 7.65, and 5.38, all $ps < 0.001$, $d = 1.37$, 0.84, 1.60, and 1.12 for the RF-valence, RF-arousal, linear-valence, and linear-arousal models, respectively). Paired t-tests indicated no significant differences between the RF and linear models in valence ($t[22] = 0.21$, $p = 0.827$, $d = 0.05$) or arousal ($t[22] = 0.22$, $p = 0.839$, $d = 0.04$).

SHAP tools were applied to quantify and visually depict nonlinear associations in the RF model. The absolute mean SHAP values, indicating feature importance in the RF model, are shown in Fig. 5. The results suggested that AUs 06, 12, 09, and 04, and AUs 43, 09, 06, and 07, were the four most important features for predicting valence and arousal, respectively.

The SHAP dependency plots in Fig. 6 show representative relationships between the AUs and SHAP values for changes in valence/arousal ratings. Both simple linear and more complex non-linear associations are shown (e.g., step-like and gradual changes in the valence—AU 12 and valence—AU 04 relationships, respectively).

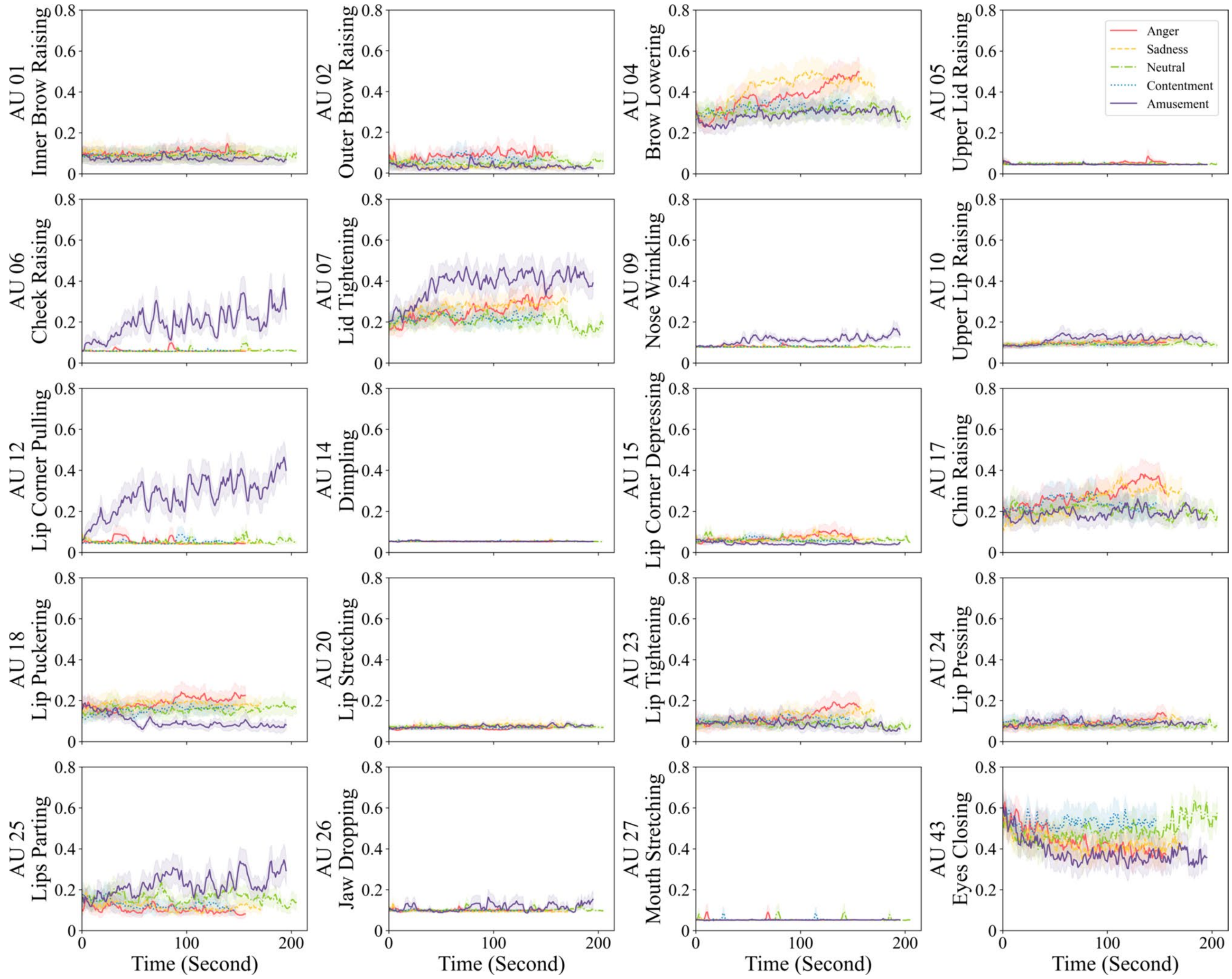

Fig. 2. Group mean action unit (AU) intensities associated with the viewing of five films. The upper panels show line charts highlighting the differences in activation levels of the same AUs across different films, emphasizing how emotional responses varied with the content viewed. The solid lines represent the mean values, and the shaded area of the same color indicates the standard error. The lower panels show spectrograms highlighting the differences in activation levels of the different AUs while watching the same film, revealing the complexity of the emotional reactions elicited by a single film. A redder hue indicates higher AU intensity, and a more purple hue indicates lower intensity.

Discussion

We confirmed our hypothesis that emotional valence ratings were negatively and positively associated with facial AUs 04 (brow lowering) and 12 (lip-corner pulling), respectively, as detected through automated video data analysis. These results are consistent with several previous studies that reported associations between dynamic valence ratings and EMG activity recorded from the corrugator supercilii and zygomatic major muscles[7–10]. However, it is unknown whether such electrical facial muscle activity was visible, as facial EMG can detect invisible muscle activity[13]. These results are also consistent with previous findings that AUs 04 and 12 are associated with emotional valence states elicited by remembering personal emotional events[14], although dynamic ratings of subjective valence were not obtained. Extending these findings, the present study is the first to report that automated video analysis of AUs 04 and 12 is associated with subjective valence ratings.

In addition to these two AUs, our exploratory analyses revealed associations of other AUs with dynamic valence ratings, including AUs 01 (inner brow raising), 15 (lip corner depressing), and 18 (lip puckering) (negative associations) and AUs 06 (cheek raising), 07 (lid tightening), 09 (nose wrinkling), 20 (lip stretching), and 25 (lips parting) (positive associations). These results align with those of previous studies examining the associations between categories of subjective emotions and AUs[27,28], despite those studies not obtaining subjective valence ratings. For example, one study acquired videos of facial expressions among Japanese participants and performed an automated AU analysis, similar to our approach, and found that AUs 06 and 07 were activated by happy facial expressions elicited by emotional scenarios[29]. Another study involved manual FACS analysis while participants produced emotional facial expressions in response to scenarios, showing that AUs 01 and 15 were activated during sad experiences[30]. Our study replicated these results using a film presentation paradigm

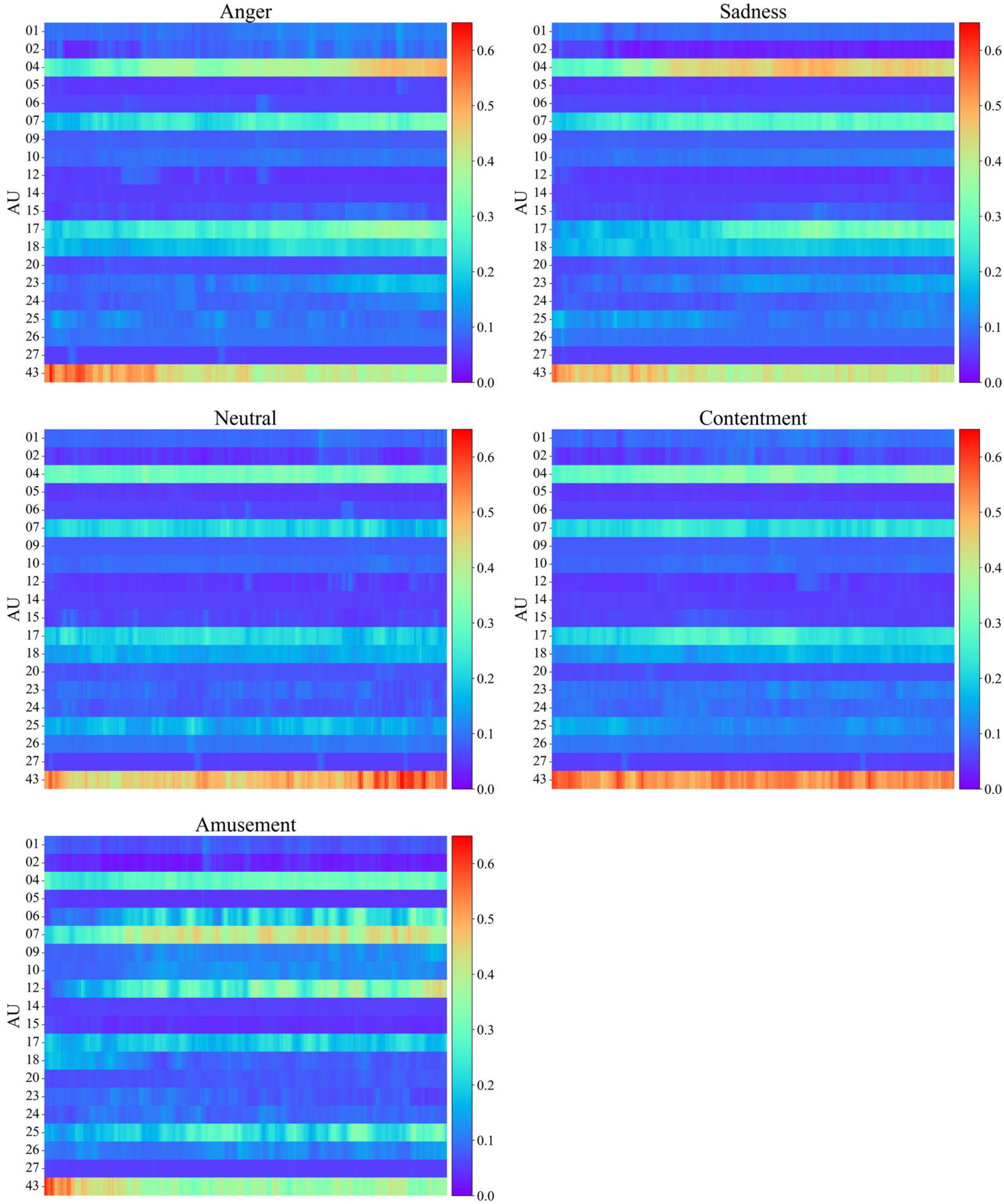

Fig. 2. (continued)

to induce spontaneous emotional responses. Taken together, our findings suggest that changes in AU activation patterns may reflect dynamic subjective experiences.

Furthermore, our exploratory analyses revealed associations between dynamic arousal ratings and several AUs, including AUs 02 (outer brow raising), 06 (cheek raising), 07 (lid tightening), 09 (nose wrinkling), 12 (lip-corner pulling), 17 (chin raising), and 43 (eyes closing). The positive associations between subjective arousal ratings and AU 12 were consistent with the results of a previous study that analyzed AUs videotaped while participants remembered personal emotional events[14]. Another study reported that the intensity of AUs 02 and

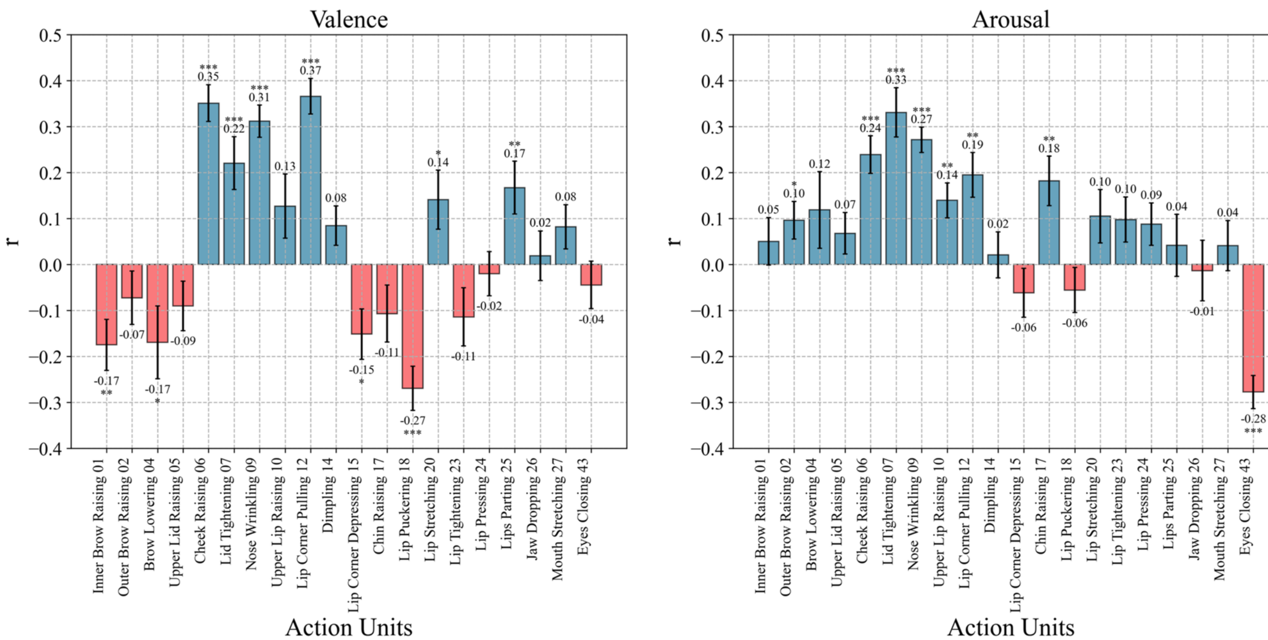

Fig. 3. Mean (standard error) Pearson's correlation coefficients between second-by-second dynamic valence (left) or arousal ratings (right) and action units across time (concatenated for all film conditions). Coefficients indicating positive and negative associations are shown as red and blue bars, respectively. The significance of the correlations is denoted by asterisks: *$p<0.05$, **$p<0.01$, and ***$p<0.001$.

Action unit	Description	Valence			Arousal		
		t	*p*	*d*	*t*	*p*	*D*
01	Inner brow raising	**3.09**	**0.005**	**0.65**	1.02	0.319	0.21
02	Outer brow raising	1.29	0.212	0.27	**2.35**	**0.028**	**0.49**
04	Brow lowering	**2.16**	**0.042**	**0.45**	1.27	0.216	0.27
05	Upper lid raising	1.7	0.102	0.36	1.52	0.144	0.32
06	Cheek raising	**8.35**	**< 0.001**	**1.74**	**5.54**	**< 0.001**	**1.16**
07	Lid tightening	**3.85**	**< 0.001**	**0.80**	**5.85**	**< 0.001**	**1.22**
09	Nose wrinkling	**8.54**	**< 0.001**	**1.78**	**9.07**	**< 0.001**	**1.89**
10	Upper lip raising	1.67	0.109	0.35	**3.63**	**0.002**	**0.76**
12	Lip corner pulling	**8.55**	**< 0.001**	**1.78**	**3.73**	**0.001**	**0.78**
14	Dimpling	1.96	0.063	0.41	0.35	0.728	0.07
15	Lip corner depressing	**2.79**	**0.011**	**0.58**	1.19	0.247	0.25
17	Chin raising	1.75	0.094	0.37	**3.32**	**0.003**	**0.69**
18	Lip puckering	**5.27**	**< 0.001**	**1.10**	1.08	0.294	0.22
20	Lip stretching	**2.14**	**0.044**	**0.45**	1.77	0.091	0.37
23	Lip tightening	1.78	0.089	0.37	2.00	0.058	0.42
24	Lip pressing	0.46	0.65	0.10	1.91	0.069	0.40
25	Lips parting	**2.94**	**0.008**	**0.61**	0.53	0.598	0.11
26	Jaw dropping	0.35	0.727	0.07	0.37	0.713	0.08
27	Mouth stretching	1.7	0.103	0.36	0.74	0.468	0.15
43	Eyes closing	0.9	0.38	0.19	**6.95**	**< 0.001**	**1.45**

Table 2. Results of one-sample *t*-tests (two-tailed) of the correlation coefficients between second-by-second subjective valence or arousal ratings and action units (concatenated for all film conditions). Significant results ($p<0.05$) are in bold.

43 was associated with higher and lower arousal, respectively, as recognized in the facial expressions of virtual agents[31]. Interestingly, previous psychophysiological studies proposed that while subjective valence can be assessed using facial EMG, estimating subjective arousal requires recording autonomic nervous system activity, including electrodermal activity[32,33]. Together with previous findings, our data suggest that facial actions may reflect the dynamics of arousal experiences.

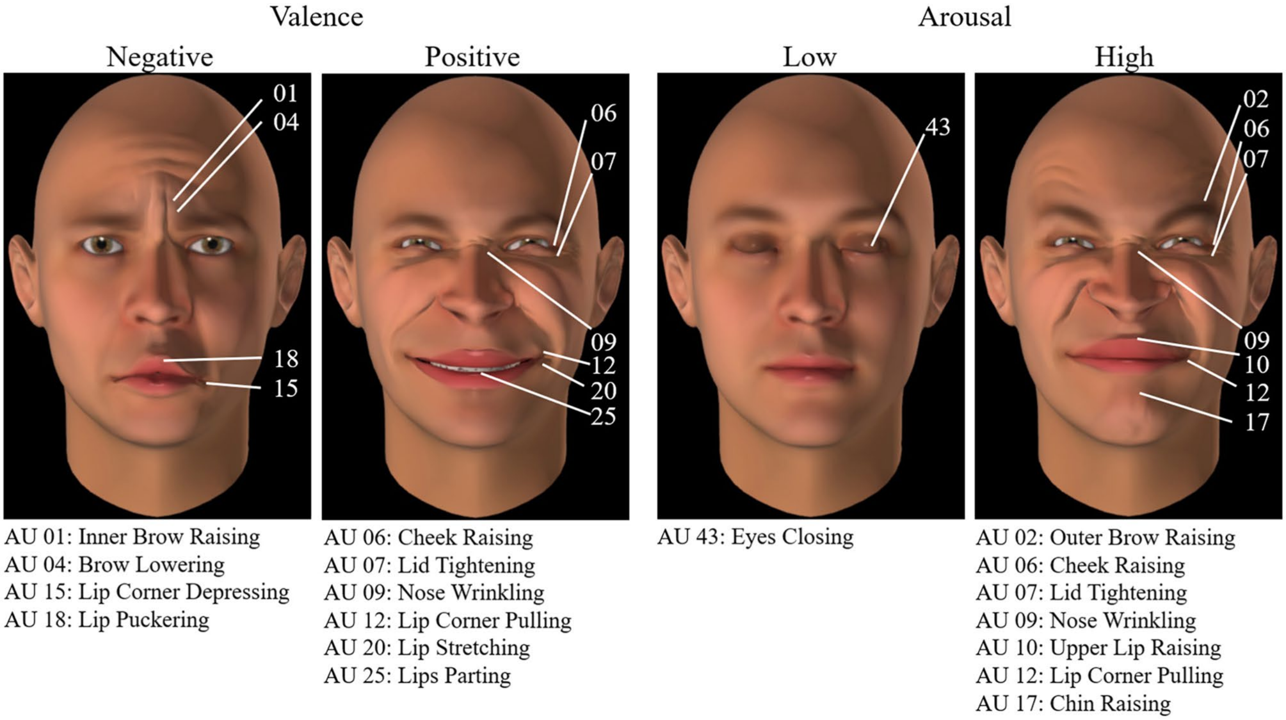

Fig. 4. Illustrations of the action units (AUs) that were significantly correlated with valence or arousal ratings. The images were created using FACSGen 2.0[25,26].

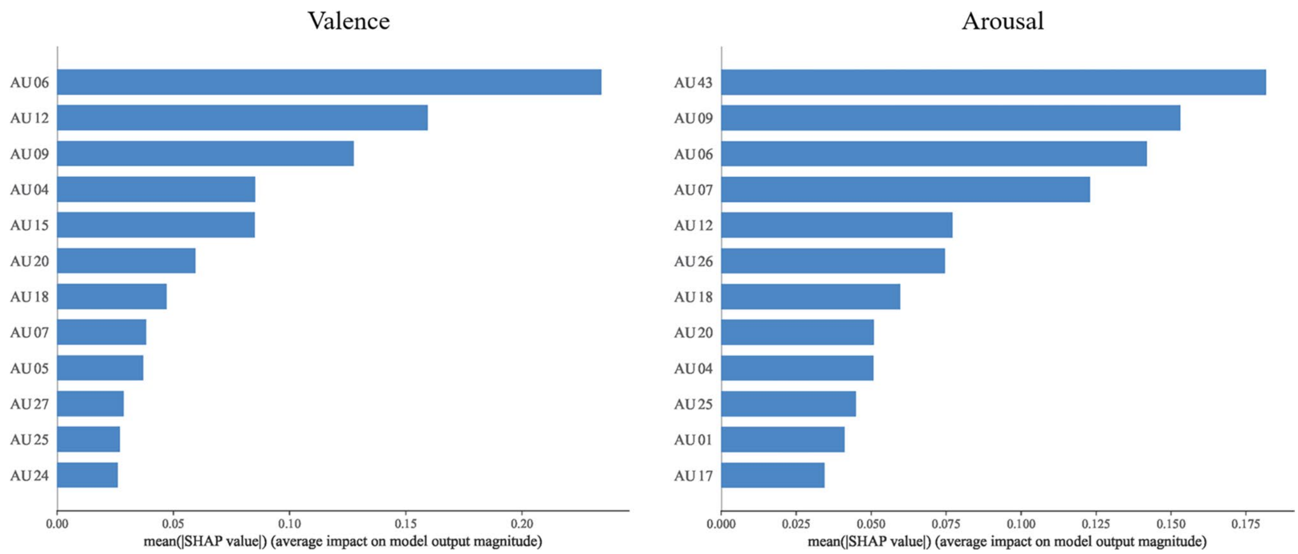

Fig. 5. Mean absolute SHapley Additive exPlanation (SHAP) values for each action unit (AU) in the random forest regression model, showing the effect of each feature on the subjective valence (left) and arousal (right) rating predictions.

Our ML modeling results also revealed some non-linear associations between subjective valence/arousal ratings and AUs. The SHAP dependency plots (Fig. 6) revealed both simple linear and more complex non-linear associations between the emotional dynamics and AUs. These results are in line with those of previous studies in which video analysis of the encoding of emotional facial expressions revealed non-linear AU trajectories[34,35], although the relationships with subjective emotional dynamics were not assessed. Our results revealed step-like and gradual relationships of subjective valence dynamics with AU 12 (lip-corner pulling) and AU 04 (brow lowering), respectively. These results are consistent with the relationships between dynamic valence ratings and EMG signals recorded from the zygomatic major and corrugator supercilii muscles. These data suggest that cheek and brow actions are robustly associated with valence dynamics in unique, non-linear ways that can be detected via video analysis. However, predictive performance was comparable between linear and non-linear ML

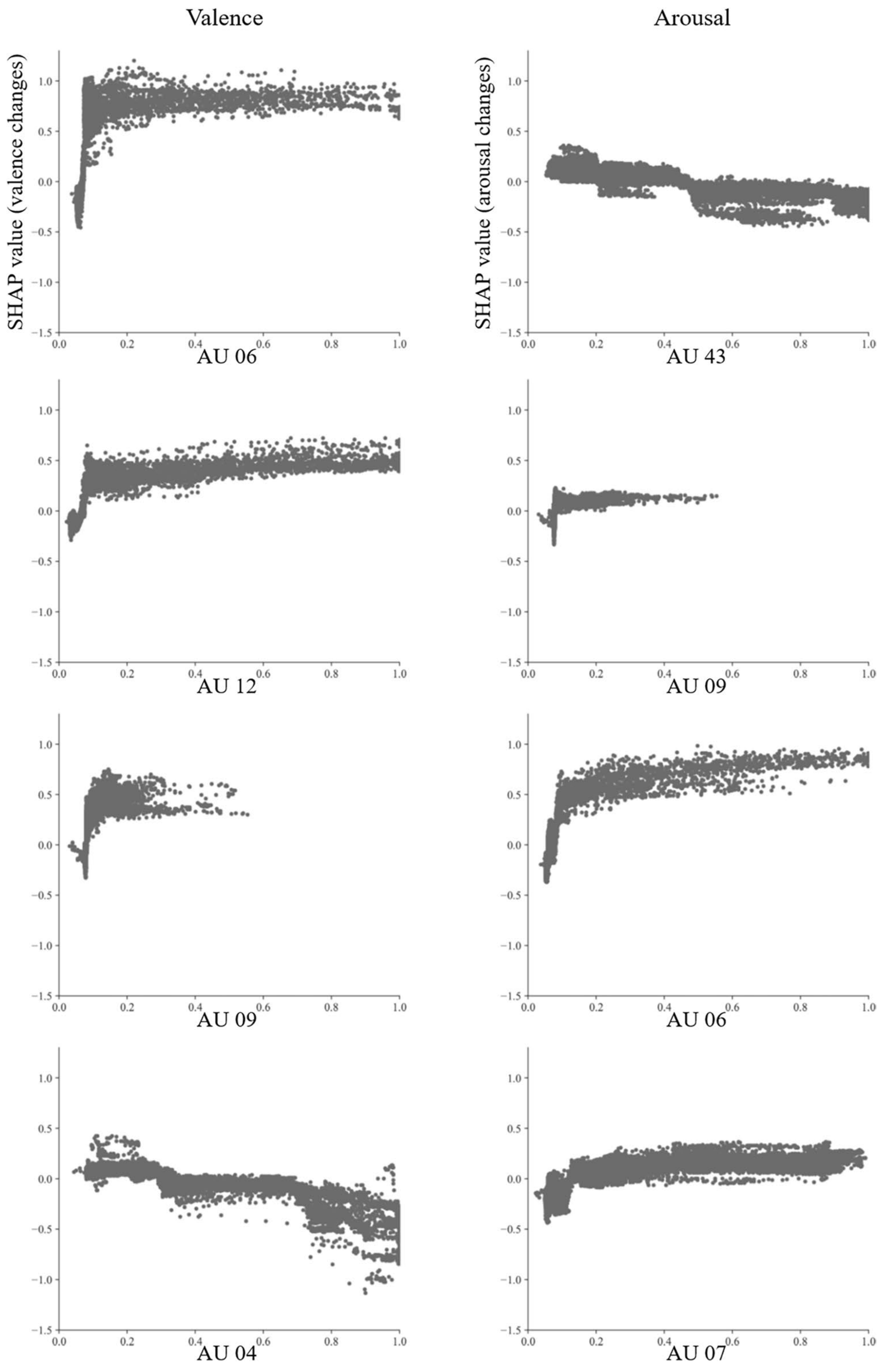

Fig. 6. SHapley Additive exPlanation (SHAP) dependency plots showing relationships between predicted valence (left) or arousal (right) ratings and facial action units (AUs). The AUs shown are the four most important features for predicting valence and arousal. The SHAP values demonstrate how the valence and arousal ratings change as the AUs are inputted.

models. The feature importance plots (Fig. 5) obtained via SHAP analysis also showed that the AUs important for predicting subjective dynamic valence/arousal ratings were similar between linear and non-linear models.

The results suggest that the relationships between emotional valence/arousal ratings and AUs may primarily be linear. However, our sample size was relatively small and may not have been sufficient for ML modeling; thus, further ML modeling is warranted.

The present results have practical implications, showing that subjective valence and arousal ratings can be estimated via automated AU analyses of video data without the need for contact electrodes. Video analysis is useful for evaluating many facial actions, as well as overall facial expressions, providing a more comprehensive and detailed analysis than facial EMG recordings. These analyses could be used for emotion sensing in several crucial areas, including mental health diagnostics[36–39], security monitoring[40,41], and education[42–44]. Mental health professionals are occasionally asked to monitor patients' emotional states remotely for early identification of a disturbance and timely intervention, which improves treatment efficacy. Precise and non-invasive facial expression analysis enhances the capability to identify suspicious behavior in public and sensitive environments, thereby supporting safety protocols without impinging on personal privacy. This innovative approach enables instructors to evaluate student engagement and emotional state in real time, facilitating the development of teaching approaches that are responsive to the emotional and educational requirements of learners.

Several limitations of the present study may have affected the findings. First, AU 04 (brow lowering) was continuously active while viewing all of the films (see Fig. 2). We speculate that this could be attributed to the recording environment, in which participants were illuminated by a single overhead light. This lighting from above may have cast shadows between the eyebrows, leading to consistent detection of AU 04 activation; observed activations of facial AUs may be affected by external lighting conditions as well as actual changes in facial expressions. Enhancing the lighting on either side of the participant may be a viable solution to minimize shadow-induced activations in future studies. Second, our facial action analysis included only 20 AUs due to the software settings[20], even though human FACS coders can evaluate more AUs[15,16]. Therefore, it is possible that we overlooked some associations between subjective emotional dynamics and other AUs. For example, a previous study[45] testing the relationship between emotional category recognition and AUs found a positive association between fear recognition and AU 16 (lower lip depressor), which was not included in our analysis. Further research will be needed to analyze more AUs using other software or methods. Third, we analyzed correlations between AU intensities and dynamic ratings across time using all film condition data as in a previous EMG study[9], so it is still unclear whether the correlations obtained herein can be generalized or apply only to specific emotional types. We conducted preliminary analyses and found that almost all of the significant rating–AU correlations reported above were maintained after controlling for the main effect of stimulus valence and the interaction between AU and stimulus valence (see "Methods"). However, our experimental design was insufficient to draw definitive conclusions because we presented only one or two film stimuli with negative, neutral, or positive valence, thus precluding discrimination between the effects of emotional and non-emotional film factors. To overcome this issue, future research should present multiple film stimuli inducing each type of emotion. Fourth, the generalizability of the findings to different social situations remains untested; this is an important matter for future research. Research indicates that facial expressions play an important role in communication in social situations[46,47], which is relevant to the association between subjective emotional experience and facial actions found in this study. Fifth, our sample included only Japanese participants. A previous study showed that, although Japanese and American participants exhibited nearly identical facial expressions in response to negative films when they were alone, Japanese participants masked their negative facial expressions in the presence of another person[48]. Although our participants watched emotional films alone, the experimental setup may have been implicitly experienced as a social situation[49], thus reducing their facial actions. Future research should include participants from different cultural backgrounds. Finally, we did not assess the participants' default emotional states, familiarity with the film contents, personality traits, or facial features, which may affect emotional experience and AU activation. Studies with a more detailed assessment process would deepen understanding of the relationships between subjective emotional dynamics and facial actions.

In conclusion, this study demonstrated associations between emotional valence/arousal ratings and facial AUs through automated video analysis and provided the first evidence that automated video analysis can reveal the dynamics of subjective emotional valence/arousal. Facial AU activity exhibited both simple linear and complex non-linear relationships. Different AUs vary in their sensitivity to emotional stimuli, and thus also in their response to changes in emotions. For example, some AUs, such as AU 04 (brow lowering), 06 (cheek raising), and 12 (lip-corner pulling) are extremely sensitive to slight changes in emotion, showing significant activation in response to subtle changes in facial expressions when emotions are finely tuned. In contrast, other AUs may only show significant activation when there is a stronger emotional stimulus. Therefore, when analyzing and interpreting AU activity data, it is crucial to consider these dynamics, and non-linear relationships, to accurately characterize people's emotional responses. Despite several methodological limitations, our study provides support for facial expression-based emotional monitoring and personalized interventions. Future research should improve the experimental methods and explore non-linear relationships between dynamic emotional states and AUs to more accurately understand and predict these states. This approach will further advance the application and development of facial expression analysis technology for use in various critical sectors.

Methods

Participants

We enrolled 23 healthy Japanese adults [11 women and 12 men; mean ± standard deviation (*SD*) age = 22.0 ± 2.6 years]. The sample size was determined through a priori power analysis conducted using G*Power software (ver. 3.1.9.2)[50], based on a previous study that recorded facial EMG of the corrugator supercilii and zygomatic major muscles and obtained dynamic valence ratings using similar experimental procedures[9]. Analysis of subjective facial associations using a two-step procedure with one-sample *t*-tests (two-tailed) was planned. An effect size d of 0.55 was estimated based on the weak subjective-facial association; with an α level of 0.05 and a power ($1-\beta$) of 0.8, the power analysis showed that > 22 participants were needed. Data from two additional participants, though tested, were excluded due to technical problems with the video acquisition system. The participants were recruited via advertisements at Kyoto University and were compensated in book coupons corresponding to a value of 4000 Japanese yen. The inclusion criteria were as follows: willingness to participate in subjective and physiological measurements; normal or corrected-to-normal vision without the use of glasses; Japanese as the first language; and no neurological or psychiatric issues. The exclusion criterion was previous experience of participating in experiments using the emotional film clips employed in our study. All participants provided written informed consent after a thorough explanation of the procedures. This study was approved by the RIKEN Ethics Committee. The experiment was conducted following the institutional ethical guidelines and the Declaration of Helsinki.

Apparatus

In the one-shot rating session, experimental events were managed using Presentation software (Neurobehavioral Systems, Berkeley, CA, USA) running on an HP Z200 SFF Windows computer (Hewlett-Packard Japan, Tokyo, Japan). The software presented films and response displays, recorded the participants' ratings, and provided digital trigger pulses synchronized with film onset. Visual stimuli were displayed on a 19-inch cathode ray tube monitor (HM903D-A; Iiyama, Tokyo, Japan) with a 100-Hz refresh rate and 1024 × 768-pixel resolution. A digital web camera (HD1080 P; Logicool, Tokyo, Japan) was placed above the monitor for video recording. In addition, A655sc infrared thermal imaging cameras (FLIR Systems, Wilsonville, OR, USA) were used to acquire the facial thermal images; these data are not reported here. For the cued-recall dynamic valence and arousal ratings, we used PsychoPy software (v2023.2.3; Open Science Tools Ltd., Nottingham, UK) running on a MacBook Air laptop (M2; Apple, Cupertino, CA, USA).

Stimuli

Five films were utilized to evoke a spectrum of emotions: "Cry Freedom" (highly negative, anger), "The Champ" (moderately negative, sadness), "Abstract Shapes" (neutral), "Wild Birds Of Japan" (moderately positive, contentment), and "M-1 Grand Prix The Best 2007–2009" (highly positive, amusement). Gross & Levenson[49] and Sato et al.[7] developed the first three and the latter two film stimuli, respectively, and their effectiveness in eliciting the target emotions was validated in several previous studies of Japanese samples[7–10,19]. The mean ± *SD* duration of these film stimuli was 175.8 ± 22. 2 s (anger, 157 s; sadness, 172 s; neutral, 206 s; contentment, 148 s; amusement, 196 s). Two additional films were used for the practice trials, i.e., scenes from "Silence of the Lambs"[10] and "Colour Bars" from Gross & Levenson[51]. The stimulus resolution was 640 × 480 pixels, corresponding to visual angles of approximately 25.5° and 11°.

Procedure

The experiments were completed on an individual basis. Upon arrival at the laboratory, participants were briefed on the overall procedure. The study commenced with a one-shot rating session for the acquisition of video data, followed by a dynamic rating session.

The one-shot rating session was conducted in a soundproof, electrically shielded room (Science Cabin; Takahashi Kensetsu, Tokyo, Japan). The room had a ceiling-mounted light (EFG25ED/19H; Panasonic, Tokyo, Japan) that remained on throughout the experiment. The room temperature was maintained at 23.5–24.5 °C, monitored, and recorded using a TR-76Ui data logger (T&D, Matsumoto, Japan). A previous study showed that participants in warmer conditions had higher arousal scores for specific picture categories compared with those in neutral or cool conditions, and the valence score for warmer conditions was lower than that for neutral conditions[52]. Without consideration of energy consumption, the Japanese government recommends an indoor temperature of 20°C in winter. However, Japanese people are accustomed to slightly higher temperatures in the range of 22–28 °C[53]. Therefore, before starting each experiment, we set the indoor temperature to 24 °C using central air conditioning. We also asked the participants to remove their outerwear, and to indicate their level of comfort with the room temperature while they completed the study forms. Adjustments were made based on the participants' feedback to ensure that the temperature did not become too high or too low during the experiment, as this could have affected the results. All participants reported feeling comfortable at around 24 °C, leading us to designate 23.5–24.5 °C as the neutral temperature range in our experiments. Each participant was seated comfortably on a chair that was fixed such that the participant's face was approximately 0.77 m from the monitor. The digital web camera was located above the monitor, in alignment with the front edge of the monitor.

In the one-shot rating session, after two practice films, participants were presented with the five experimental films (presented pseudo-randomly). In each trial, a fixation point was displayed for 1 s, followed by a white screen for 10 s and then the film. The screen reverted to white for another 10 s post-film. Subsequently, the affect grid[54] was presented, allowing participants to provide emotional valence and arousal ratings using a 9-point scale (Fig. 7). The participants were instructed to focus on the fixation point, watch the film, and then provide their subjective valence and arousal ratings by pressing number keys 1–9 on the keyboard. Following their responses, the screen turned black during the inter-trial interval, which varied randomly between 24 and 30 s. Video data were continuously acquired throughout all of the trials using LabChart Pro v8.0 software (ADInstruments, Dunedin, New Zealand). The software simultaneously recorded digital signals (triggers) associated with film onset. Upon completion of the one-shot rating trials, participants exited the soundproof room and rested at a desk in the laboratory.

No videos were recorded during the dynamic rating session. The same films, in the same order, as viewed in the soundproof room, were presented twice more on separate occasions. In each trial, participants viewed each

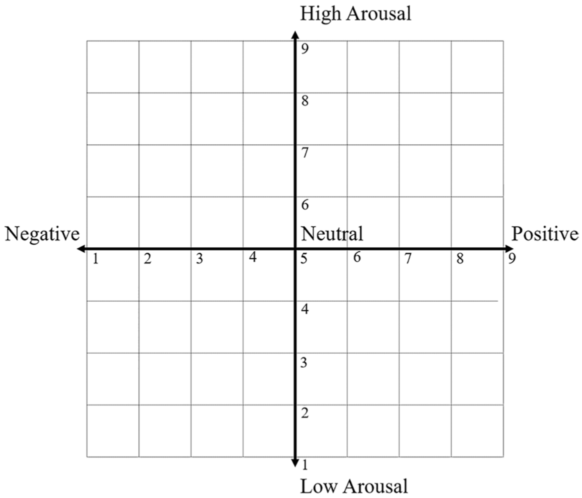

Fig. 7. Response display for the one-shot ratings of valence and arousal.

film stimulus on a laptop, accompanied by horizontal and vertical 9-point scales for the valence and arousal ratings, respectively. The participants were instructed to recall their initial emotional response and indicate their subsequent emotional state by adjusting the slider on a touchpad (Fig. 8). No specific frequency was required with respect to rating changes. The participants' fingers were continuously in contact with the touchpad to ensure that the software could automatically update the data in real time. Participants first provided valence ratings for all five films, followed by arousal ratings after a short break. Previous studies have shown a strong positive correlation between cued-recall dynamic ratings and online dynamic ratings of emotional films[7,24].

Data analysis

Video data were analyzed using FaceReader 9 (Noldus Information Technology, Wageningen, The Netherlands). We used this software because of its validated AU coding peformance[20,55]. First, the software detected faces in frames based on the Viola–Jones algorithm[56]. Next, it constructed three-dimensional face models based on the active appearance method[57] in combination with deep artificial neural network classification[58]. Finally, by using an artificial neural network trained on a large database of AUs, the software quantified the intensities of 20 AUs (01, 02, 04, 05, 06, 07, 09, 10, 12, 14, 15, 17, 18, 20, 23, 24, 25, 26, 27, and 43). The East Asian template was used. The start and end times for each film stimulus were determined based on digital triggers synchronized with the film onset and stimulus duration, respectively. The video data for each participant were segmented into five clips based on the start and end times. One-shot ratings were exported directly from the log files of Presentation

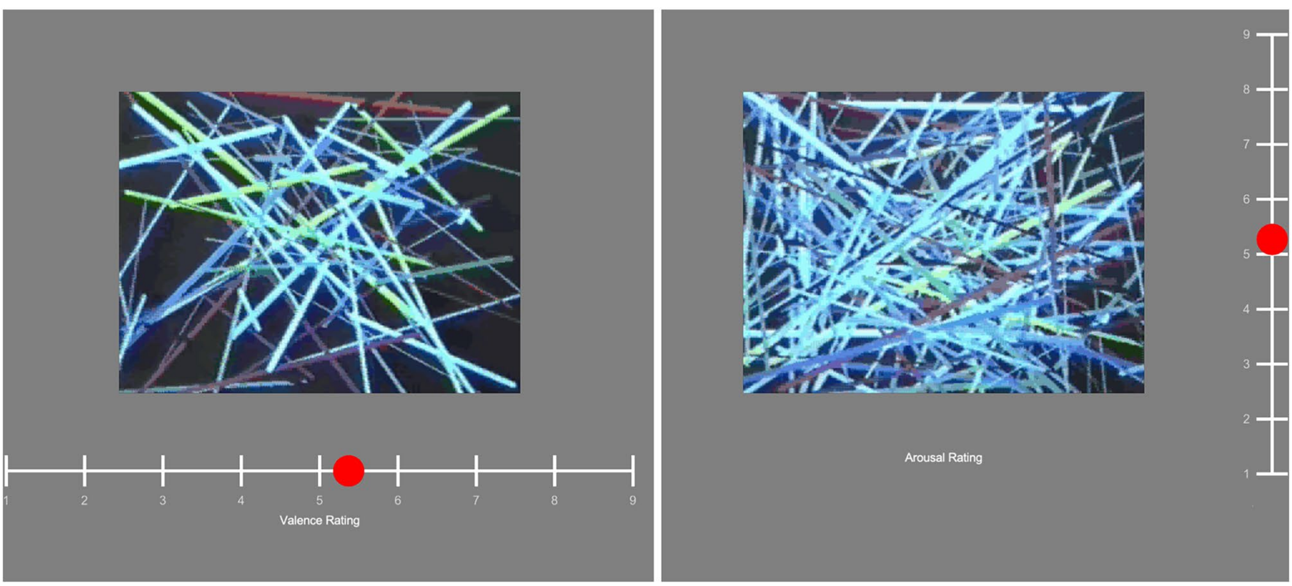

Fig. 8. Response displays for the dynamic ratings of valence (left) and arousal (right).

software. Dynamic ratings were exported from the log files of PsychoPy software, where they were originally recorded on a frame-by-frame basis. We used the codeOnlineRating function in PsychoPy software to record dynamic valence and arousal ratings while participants watched the films. With regard to technical details, component setup was as follows: codeOnlineRating has fixed parameters, including "Before Experiment," "Begin Experiment," "Begin Routine," "Each Frame," "End Routine," and "End Experiment". It also featured frame-by-frame recording: we wrote code for the "Each Frame" parameter to ensure that, whenever participants updated their rating (i.e., changed the slider position), the system would update the rating, with the data saved every three frames. Each data entry event included the current rating and the timestamp (in milliseconds), as well as the monitor refresh rate; the recording of frames is tied to the monitor's refresh rate, which PsychoPy cannot control. Our monitor's refresh rate was 60 Hz, such that there were 60 frame updates per second. Consequently, we saved 20 frames of data per second. To ensure alignment with the AU intensity data, the ratings were converted to a second-by-second format by calculating the mean values of all frames within each second. Subsequently, each participant's AU intensity data, along with their continuous valence and arousal ratings, were arranged in a fixed film stimulus order (anger, sadness, neutral, contentment, and amusement) and stored in CSV files for subsequent analysis.

Statistical analysis

Data analysis was conducted using Python (version 3.11.4), JASP 0.14.1[59], MATLAB 2020a (MathWorks, Natick, MA, USA), and the Hotelling T^2 function[60]. Results were considered significant at $p < 0.05$.

The one-shot ratings and mean dynamic ratings during film presentation (five ratings for each participant) were subjected to repeated-measures trend analyses. A previous study showed that the five film clips (anger, sadness, neutral, contentment, and amusement) used in this study exhibited linear and quadratic relationships with subjective valence and arousal ratings, respectively[9]. To confirm this, the linear and quadratic natures of the valence and arousal ratings were assessed across films.

We conducted preliminary analyses to explore differences between the one-shot ratings and mean dynamic ratings. The valence and arousal ratings were subjected to a 2 (rating type: one-shot and dynamic) × 5 (film: anger, sadness, neutral, contentment, and amusement) repeated-measures analysis of variance. The two-way interactions were evaluated for the linear and quadratic relationships for valence and arousal ratings, respectively. Regarding the valence ratings, the results showed significant main effects of rating type (dynamic > one-shot; $F[1,22] = 12.10$, $p = 0.002$, $\eta^2_p = 0.36$) and film ($F[4, 88] = 66.04$, $p < 0.001$, $\eta^2_p = 0.75$), and a significant rating type × film interaction ($F[4, 88] = 45.34$, $p < 0.001$, $\eta^2_p = 0.67$). Trend analysis revealed that the linear relationship was significantly different between the rating types (one-shot > dynamic; $F[1,22] = 140.09$, $p < 0.001$, $\eta^2_p = 0.86$). Similarly, for the arousal ratings, we found significant main effects of rating type (one-shot > dynamic; $F[1,22] = 22.12$, $p < 0.001$, $\eta^2_p = 0.50$) and film ($F[4, 88] = 45.10$, $p < 0.001$, $\eta^2_p = 0.67$), and a significant rating type × film interaction ($F[4, 88] = 6.61$, $p < 0.001$, $\eta^2_p = 0.23$). Trend analysis revealed that the quadratic relationship was significantly different between the rating types (one-shot > dynamic; $F[1,22] = 9.10$, $p = 0.006$, $\eta^2_p = 0.29$). These data indicate greater differences in valence and arousal ratings between films for one-shot versus mean dynamic ratings. We speculate that the results could be explained by memory biases in one-shot retrospective emotional ratings (i.e., the peak-end rule[61]).

To assess the individual-level linear associations between the subjective dynamic valence/arousal ratings and AUs, Pearson's product-moment correlation coefficients (r-values) were calculated for each participant. Individual-level r-values, after normalization using Fisher transformation, were subjected to a two-stage random-effects analysis performed at the group level[62]. First, we conducted a priori analyses of the associations of valence ratings with AUs 04 and 12, based on our research interests, as described in the Introduction. The z-transformed r-values were analyzed using one-sample t-tests (two-tailed), as in a previous study[9]. Next, we conducted exploratory analyses of the associations of the valence and arousal ratings with all 20 analyzed AUs. We performed multivariate analysis using Hotelling's one-sample T^2 test, which is a multivariate generalization of a one-sample t-test, to control the experiment-wise type I error rate[63,64]. We conducted univariate one-sample t-tests (two-tailed) for the follow-up analyses. To visually illustrate the AUs that showed significant correlations with valence or arousal ratings, artificial facial images were created using FACSGen 2.0[25,26].

We conducted preliminary analyses to determine whether the correlations between the dynamic ratings and AUs were specific to particular types of stimulus valence, although our experimental design (only one or two films for negative, neutral, and positive stimuli) was not appropriate to draw definitive conclusions. Linear mixed-effects models were constructed, including the second-by-second dynamic ratings of valence or arousal as the dependent variable (879 data points for each participant). The AU intensities were the effect-of-interest fixed-effect independent variables; these were analyzed together with the effect-of-no-interest fixed-effect covariates of stimulus valence (negative, neutral, and positive) and the interaction between AU intensity and stimulus valence. Random by-participant intercepts were added as per standard repeated-measures analyses. Beta estimates of the AU intensities were evaluated with the degrees of freedom calculated using Satterthwaite's approximation. The results (Supplementary Table 1) confirmed that all significant associations of valence and arousal ratings with AU intensities reported above were also significant in these analyses, except for the valence—AU 20, arousal—AU 9, and arousal—AU 10 associations. These data suggest that almost none of the reported dynamic rating–AU correlations were restricted to specific types of stimulus valence.

ML modeling

The ML modeling was similar to that conducted in a previous study analyzing the relationship between facial EMG and dynamic ratings[17]. First, the data were segmented into 10-s sections, and the mean values were calculated as the features. The RandomForestRegressor in the Python library scikit-learn was used for RF regression modeling. Finally, the RF model incorporated 20 decision trees, with a maximum depth of 6; the other parameters took their default values. For comparison, linear multiple regression analyses were conducted using scikit-learn. The Pearson's correlation coefficients between the model predictions and the actual values for each participant were calculated as the model evaluation index. Leave-one-out cross-validation was employed, where data from one participant served as the evaluation dataset, and the data from the other participants comprised the training dataset used to train the models. The indices were subjected to statistical analysis, including one-sample *t*-tests (two-tailed) contrasting with zero and paired *t*-tests comparing the RF and linear models. For the SHAP analysis, we computed SHAP values for each instance[18]. We calculated the absolute mean SHAP values for the obtained hyperquantities; this allowed us to assess the significance of each input feature. Finally, we plotted scatterplots of the relationships between input features and their SHAP values for each instance.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 7 May 2024; Accepted: 19 August 2024
Published online: 22 August 2024

References

1. Rozin, A., Rozin, P. & Goldberg, E. The feeling of music past: How listeners remember musical affect. *Music Percept.* **22**, 15–39 (2004).
2. Schäfer, T., Zimmermann, D. & Sedlmeier, P. How we remember the emotional intensity of past musical experiences. *Front. Psychol.* **5**, 911 (2014).
3. Strijbosch, W. *et al.* From experience to memory: On the robustness of the peak-and-end-rule for complex, heterogeneous experiences. *Front. Psychol.* **10**, 1705 (2019).
4. Strijbosch, W. *et al.* When the parts of the sum are greater than the whole: Assessing the peak-and-end-theory for a heterogeneous, multi-episodic tourism experience. *J. Destin. Mark. Manag.* **20**, 100607 (2021).
5. Houben, M., Van Den Noortgate, W. & Kuppens, P. The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychol. Bull.* **141**, 901–930 (2015).
6. Sato, W. Advancements in sensors and analyses for emotion sensing. *Sensors* **24**, 4166 (2024).
7. Sato, W., Kochiyama, T. & Yoshikawa, S. Physiological correlates of subjective emotional valence and arousal dynamics while viewing films. *Biol. Psychol.* **157**, 107974 (2020).
8. Sato, W. *et al.* Emotional valence sensing using a wearable facial EMG device. *Sci. Rep.* **11**, 5757 (2021).
9. Sato, W. & Kochiyama, T. Exploration of emotion dynamics sensing using trapezius EMG and fingertip temperature. *Sensors* **22**, 6553 (2022).
10. Saito, A., Sato, W. & Yoshikawa, S. Altered emotional mind-body coherence in older adults. *Emotion* **24**, 15–26 (2024).
11. Huang, C. N., Chen, C. H. & Chung, H. Y. The review of applications and measurements in facial electromyography. *J. Med. Biol. Eng.* **25**, 15–20 (2005).
12. Hsu, C. T. & Sato, W. Electromyographic validation of spontaneous facial mimicry detection using automated facial action coding. *Sensors* **23**, 9076 (2023).
13. Tassinary, L. G. & Cacioppo, J. T. Unobservable facial actions and emotion. *Psychol. Sci.* **3**, 28–33 (1992).
14. Namba, S. *et al.* Development of the RIKEN database for dynamic facial expressions with multiple angles. *Sci. Rep.* **13**, 21785 (2023).
15. Ekman, P. & Friesen, W. V. *Facial action coding system (FACS): A Technique for the Measurement of Facial Action* (Consulting Psychologists Press, 1978).
16. Ekman, P., Friesen, W. V. & Hager, J. V. *Facial Action Coding System*. 2nd ed. (Research Nexus eBook, 2002).
17. Kawamura, N., Sato, W., Shimokawa, K., Fujita, T. & Kawanishi, Y. Machine learning-based interpretable modeling for subjective emotional dynamics sensing using facial EMG. *Sensors* **24**, 1536 (2024).
18. Lundberg, S. M. & Lee, S. I. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17).* 4768–4777 (2017).
19. Sato, W., Noguchi, M. & Yoshikawa, S. Emotion elicitation effect of films in a Japanese sample. *Soc. Behav. Pers.* **35**, 863–874 (2007).
20. Lewinski, P., den Uyl, T. M. & Butler, C. Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *J. Neurosci. Psychol. Econ.* **7**, 227–236 (2014).
21. Ruef, A. M. & Levenson, R. W. Continuous measurement of emotion: The affect rating dial. In *Handbook of Emotion Elicitation and Assessment* (eds. Coan, J. A. & Allen, J. J. B.). 286–297 (Oxford University Press, 2007).
22. Rosenberg, E. L. & Ekman, P. Coherence between expressive and experiential systems in emotion. *Cogn. Emot.* **8**, 201–229 (1994).
23. Nielsen, L. & Kaszniak, A. W. Conceptual, theoretical, and methodological issues in inferring subjective emotion experience: Recommendations for researchers. In *Handbook of Emotion Elicitation and Assessment*. (eds. Coan, J. A. & Allen, J. J. B.). 361–375 (Oxford University Press, 2007).
24. Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H. & Gross, J. J. The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion* **5**, 175–190 (2005).
25. Roesch, E. *et al.* FACSGen: A tool to synthesize emotional facial expressions through systematic manipulation of facial action units. *J. Nonverbal Behav.* **35**, 1–16 (2011).
26. Krumhuber, E. G., Tamarit, L., Roesch, E. B. & Scherer, K. R. FACSGen 2.0 animation software: Generating three-dimensional FACS-valid facial expressions for emotion research. *Emotion* **12**, 351–363 (2012).
27. Reisenzein, R., Studtmann, M. & Horstmann, G. Coherence between emotion and facial expression: Evidence from laboratory experiments. *Emot. Rev.* **5**, 16–23 (2013).
28. Durán, J. I., Reisenzein, R. & Fernández-Dols, J.-M. Coherence between emotions and facial expressions: A research synthesis. In *The Science of Facial Expression* (eds. Fernández-Dols, J.-M. & Russell, J. A.). 107–129 (Oxford University Press, 2017).
29. Sato, W., Hyniewska, S., Minemoto, K. & Yoshikawa, S. Facial expressions of basic emotions in Japanese laypeople. *Front. Psychol.* **10**, 259 (2019).
30. Gosselin, P., Kirouac, G. & Doré, F. Y. Components and recognition of facial expression in the communication of emotion by actors. *J. Pers. Soc. Psychol.* **68**, 83–96 (1995).
31. Boukricha, H., Wachsmuth, I., Hofstätter, A. & Grammer, K. Pleasure-arousal-dominance driven facial expression simulation. In *3rd International Conference on Affective Computer Intelligence on Interacting Workshops*. 1–7 (2009).
32. Cacioppo, J. T., Berntson, G. G. & Klein, D. J. What is an emotion? The role of somatovisceral afference, with special emphasis on somatovisceral "illusions". In *Emotion Social Behaviour* (Clark, M. S. Ed.). 63–98 (1992).

33. Lang, P. J., Bradley, M. M. & Cuthbert, B. N. Emotion, motivation, and anxiety: Brain mechanisms and psychophysiology. *Biol. Psychiatry* **44**, 1248–1263 (1998).
34. Pantic, M. & Patras, I. Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Trans. Syst. Man. Cybern. B Cybern*. **36**, 433–449 (2006).
35. Dobs, K. *et al.* Quantifying human sensitivity to spatio-temporal information in dynamic faces. *Vis. Res.* **100**, 78–87 (2014).
36. Muhammad, G., Alsulaiman, M., Amin, S. U., Ghoneim, A. & Alhamid, M. F. A facial-expression monitoring system for improved healthcare in smart cities. *IEEE Access* **5**, 10871–10881 (2017).
37. Fei, Z. *et al.* Deep convolution network based emotion analysis towards mental health care. *Neurocomputing* **388**, 212–227 (2020).
38. Lee, Y. S. & Park, W. H. Diagnosis of depressive disorder model on facial expression based on fast R-CNN. *Diagnostics* **12**(2), 317 (2022).
39. Rodosky, S. E. *et al.* Facial expressions in adolescent–parent interactions and mental health: A proof-of-concept study. *Emotion* **23**, 2110–2115 (2023).
40. Shakya, S., Sharma, S. & Basnet, A. Human behavior prediction using facial expression analysis. In *International Conference on Computing Communication on Automation* (*ICCCA*), pp. 399–404 (2016).
41. Sajjad, M. *et al.* Raspberry Pi assisted facial expression recognition framework for smart security in law-enforcement services. *Inf. Sci.* **479**, 416–431 (2019).
42. Saneiro, M., Santos, O. C., Salmeron-Majadas, S. & Jesus, G. B. Towards emotion detection in educational scenarios from facial expressions and body movements through multimodal approaches. *Sci. World J.* **2014**, 484873 (2014).
43. Xu, R., Chen, J., Han, J., Tan, L. & Xu, L. Towards emotion-sensitive learning cognitive state analysis of big data in education: Deep learning-based facial expression analysis using ordinal information. *Computing* **102**, 765–780 (2020).
44. Li, X., Yue, R., Jia, W., Wang H. & Zheng Y. Recognizing students' emotions based on facial expression analysis. In 11th *International Conference on Information Technology Medical Education (ITME)*, pp. 96–100 (2021).
45. Hyniewska, S., Sato, W., Kaiser, S. & Pelachaud, S. Naturalistic emotion decoding from facial action sets. *Front. Psychol.* **9**, 2678 (2019).
46. Crivelli, C. & Fridlund, A. J. Facial displays are tools for social influence. *Trends Cogn. Sci.* **22**, 388–399 (2018).
47. Keltner, D. & Kring, A. Emotion, social function, and psychopathology. *Rev. Gen. Psychol.* **2**, 320–342 (1998).
48. Ekman, P. & Friesen, W. V. *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues* (Prentice-Hall Inc, 1975).
49. Fridlund, A. J. *et al.* Audience effects on solitary faces during imagery: Displaying to the people in your head. *J. Nonverbal Behav.* **14**, 113–137 (1990).
50. Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175–191 (2007).
51. Gross, J. J. & Levenson, R. W. Emotion elicitation using films. *Cogn. Emot.* **9**, 87–108 (1995).
52. Hojatmadani, M., Shepard, S., Salomon, K. & Reed, K. A controlled study on evaluation of thermal stimulation influence on affective measures of uninformed individuals. arXiv: 2311.12989 (2023).
53. Khadka, S. *et al.* Study on winter comfort temperature in mixed mode and HVAC office buildings in Japan. *Energies* **15**, 7331 (2022).
54. Russell, J. A., Weiss, A. & Mendelsohn, G. A. Affect grid: A single-item scale of pleasure and arousal. *J. Pers. Soc. Psychol.* **57**, 493–502 (1989).
55. Dupré, D., Krumhuber, E.G., Küster, D. & McKeown, G.J. A performance comparison of eight commercially available automatic classifiers for facial affect recognition. *PLoS One* **15**, e0231968 (2020).
56. Viola, P. & Jones, M. Rapid object detection using a boosted cascade of simple features. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*. **5**, 183–195 (2001).
57. Cootes, T. & Taylor, C. *Statistical Models of Appearance for Computer Vision*. (University of Manchester, 2000).
58. Gudi, A. *Recognizing Semantic Features in Faces Using Deep Learning*. arXiv:1512.00743v2 (2015).
59. JASP Team. *JASP* (*Version 0.14.1*) (computer software) (2020).
60. Trujillo-Ortiz, A. Hotelling T^2. *MATLAB Central File Exchange*. https://www.mathworks.com/matlabcentral/fileexchange/2844-hotellingt2 (2024).
61. Fredrickson, B. L. & Kahneman, D. Duration neglect in retrospective evaluations of affective episodes. *J. Pers. Soc. Psychol.* **65**, 45–55 (1993).
62. Holmes, A. & Friston, K. J. Generalisability, random effects & population inference. *Neuroimage* **7**, S754 (1998).
63. Hummel, T. J. & Sligo, J. R. Empirical comparison of univariate and multivariate analysis of variance procedures. *Psychol. Bull.* **76**, 49–57 (1971).
64. Rencher, A. C. *Methods of Multivariate Analysis* 2nd edn. (Wiley, 2002).

Acknowledgements
The authors thank Masaru Usami for his technical support.

Author contributions
Conceived and designed the experiments: JZ, WS, KS, BT, and YN. Performed the experiments: JZ. Analyzed the data: JZ, WS, and NK. Wrote the paper: JZ, WS, NK, KS, BT, and YN.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-70563-8.

Correspondence and requests for materials should be addressed to W.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.