RESEARCH-ARTICLE

Caught by Surprise, Caught by Culture: Bridging Facial Expression's Recognition and Interpretation of Surprise Across Cultures

HA EUN JANG, Yonsei University, Seoul, South Korea

YEONSUN YANG, Daegu Gyeongbuk Institute of Science and Technology, Daegu, South Korea

JEAN Y SONG, Yonsei University, Seoul, South Korea

Open Access Support provided by:

Yonsei University

Daegu Gyeongbuk Institute of Science and Technology

Published: 22 March 2026

Citation in BibTeX format

IUI '26: 31st International Conference on Intelligent User Interfaces
March 23 - 26, 2026
Paphos, Cyprus

Conference Sponsors:
SIGCHI
SIGAI

https://doi.org/10.1145/3742413.3789169
ISBN: 9798400719844

Caught by Surprise, Caught by Culture: Bridging Facial Expression's Recognition and Interpretation of Surprise Across Cultures

Ha Eun Jang
Department of Innovation
Yonsei University
Incheon, Republic of Korea
j.haeun@yonsei.ac.kr

Yeonsun Yang
Electrical Engineering and Computer Science
DGIST
Daegu, Republic of Korea
diddustjs98@dgist.ac.kr

Jean Y Song*
Humanities, Arts, and Social Sciences,
Underwood International College
Yonsei University
Incheon, Republic of Korea
jeansong@yonsei.ac.kr

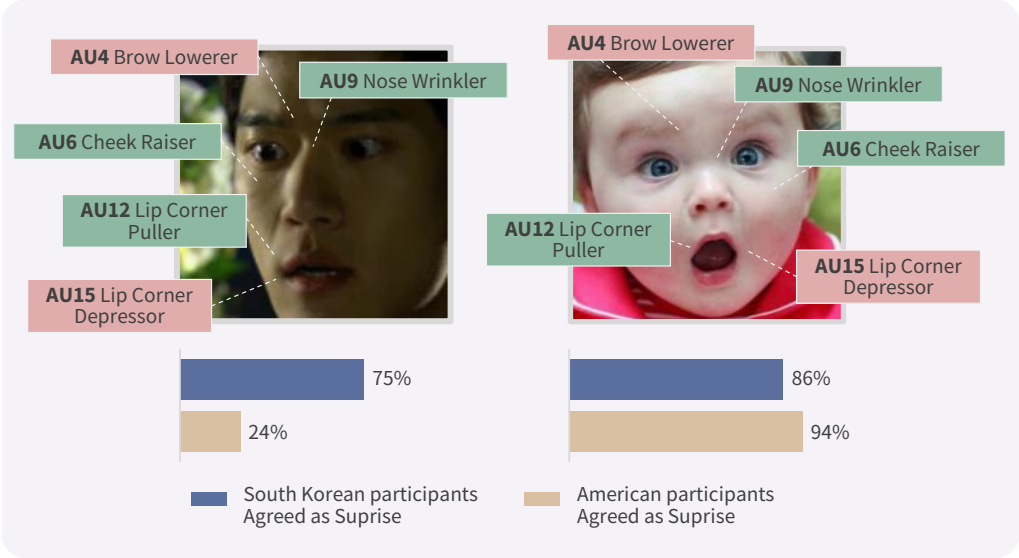

Figure 1: Examples of facial expression stimuli used in our study. The left image was labeled as *surprise* by 75% of South Korean participants and 24% of American participants, whereas the right image was labeled by 86% and 94%, respectively. Facial muscle activations (*i.e.*, Action Units) related to both positive and negative valence are marked in each face. These examples illustrate how people from different cultural backgrounds can recognize the same facial expression differently.

Abstract

Facial expressions are powerful signals of human emotion, shaping both human–human and human–computer interaction. As interactive technologies, from adaptive interfaces to emotion-aware agents, become more pervasive, systems are increasingly expected to recognize and respond to users' emotions naturally. But what if a system misreads your face? Such misinterpretation is particularly likely when cultural differences in emotion perception are overlooked. This problem may be compounded by the fact that most facial emotion recognition (FER) models are trained on datasets that reflect the norms of a particular cultural group that assume universality, limiting their reliability in multicultural contexts. Surprise, in particular, is an emotion whose valence can be either positive or negative depending on context, making it a critical case for investigating cultural bias in FER. To address this, we examined how cultural background shapes the recognition and valence interpretation of surprise facial expressions among South Korean (N=36) and American (N=34) participants. Participants labeled 200 facial expressions (surprise and fear), rated their perceived valence, and described personal experiences of surprise. Results show that South Korean-labeled surprise expressions exhibited stronger negative Action Unit (AU) activation and lower valence ratings, whereas American-labeled ones showed more balanced or positive facial cues. Qualitative accounts further revealed that South Koreans framed surprise as tense or socially cautious, while Americans viewed it as open and situationally flexible. These findings bridge recognition and interpretation in cross-cultural emotion research

*Corresponding author.

IUI '26, Paphos, Cyprus

ACM ISBN 979-8-4007-1984-4/26/03
https://doi.org/10.1145/3742413.3789169

and highlight the need for culturally adaptive FER systems that can interpret ambiguous emotions like surprise more inclusively.

CCS Concepts

• **Human-centered computing → User centered design**.

Keywords

Facial Emotion Recognition, Cultural Differences, Emotion Label Datasets

ACM Reference Format:

Ha Eun Jang, Yeonsun Yang, and Jean Y Song. 2026. Caught by Surprise, Caught by Culture: Bridging Facial Expression's Recognition and Interpretation of Surprise Across Cultures. In *31st International Conference on Intelligent User Interfaces (IUI '26), March 23–26, 2026, Paphos, Cyprus.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3742413.3789169

1 Introduction

Facial expressions are a primary channel through which people convey emotional states, making their accurate recognition essential for understanding human affect. Consequently, facial emotion recognition (FER) has become not only a topic of interest in psychology, but also an increasingly important area in HCI and AI [9, 50, 65]. Understanding facial expressions holds strong potential for diverse applications ranging from healthcare [31, 71] to social robotics [4, 72], and FER technologies are already being widely deployed in affective computing and user-centered design [1, 26, 43].

However, inaccurate assumptions of universality in facial expressions and their interpretations can lead to biased outcomes, particularly when cultural variation is overlooked. It is therefore critical that FER datasets and models reflect cultural differences. Indeed, a number of recent studies have reported cross-cultural variation in FER [18, 53, 74], suggesting that people from different cultural backgrounds do not always recognize facial expressions in the same way. Yet, most large-scale datasets still fail to capture this diversity. For example, AffectNet [54], the largest publicly available in-the-wild FER dataset, can be considered culturally biased, as its annotations were produced by groups with relatively homogeneous cultural profiles, mainly consisting of presumably westernized annotators. This bias can be particularly problematic for ambiguous emotions such as *surprise*, as people from different cultures may recognize or interpret it differently. To identify preliminary evidence of these cultural biases within established benchmarks, we conducted a pilot study examining cross-cultural annotation patterns in commonly used FER datasets. The pilot suggested cross-cultural variation in how facial expressions are annotated, motivating a study design that systematically examines the underlying processes involved in facial expression understanding.

Informed by this pilot study, we focus specifically on the cultural differences in the recognition and interpretation of *surprise*. Unlike emotions with clear positive or negative valence, *surprise* can be experienced in both ways [59]: it can be pleasant, as in receiving an unexpected gift, or unpleasant, as in hearing sudden bad news. This ambiguity makes *surprise* a theoretically rich but empirically underexplored target for cross-cultural FER applications. Despite its relevance, prior research has rarely examined how people from different cultures visually recognize and emotionally evaluate surprise expressions. To address this gap, we investigate how the same facial expressions of surprise are differently recognized and interpreted between two different cultures, combining computational and psychological perspectives. We define **recognition** as the process of categorizing external facial cues into emotion labels, and **interpretation** as the internal evaluation of the emotional meaning or valence associated with those cues.

Prior studies have often approached these aspects separately—focusing either on Action Unit(AUs, or anatomically defined facial muscle movements [17])-based categorizing of facial expressions [20, 66] or on the psychological interpretation of emotions [51, 60]—thus overlooking how perceptual and evaluative processes jointly shape cross-cultural emotion perception. To bridge this gap, we analyze (1) cross-cultural recognition tendencies based on labeling patterns and valence-oriented AU activation, and (2) how these recognition patterns relate to differences in the emotional interpretation of surprise, examined through both quantitative valence ratings and qualitative narrative accounts. The AU analysis focuses on expressions labeled as *surprise*, using a valence-informed categorization of AUs derived from prior work [81], while the interpretation analysis integrates statistical and thematic approaches to capture affective and contextual nuances. Our research was guided by two research questions (RQs):

- RQ1: What are the cross-cultural differences in the **recognition** of surprise expressions?
- RQ2: How do these recognition differences relate to cross-cultural variations in the emotional valence and **interpretation** of surprise?

We address these research questions through an online survey with 36 South Korean and 34 American participants. Participants (1) labeled 200 facial expression images from AffectNet (*surprise* and *fear*) to examine recognition patterns, (2) rated the perceived valence of each image on a seven-point scale, and (3) described two of their most recent surprising experiences. This mixed approach allowed us to capture both visual recognition patterns and emotional interpretations of surprise between two different cultural groups. Specifically, the analyses were organized around two dimensions: **recognition** (RQ1) and **interpretation** (RQ2). For RQ1, we quantitatively examined cross-cultural differences in labeling tendencies and valence-oriented AU activation. For RQ2, we combined quantitative valence ratings with qualitative analyses of participants' written accounts to understand how each culture emotionally interpreted and contextualized surprise.

By bridging the gap between computational (AU analysis, labeling patterns) and psychological perspectives (valence ratings, qualitative accounts), we offer concrete contributions to the HCI community by identifying where and how these differences manifest. We believe that these results are imperative for designing user-centered AI-driven affective systems, enabling the creation of explainable AI that moves beyond universal, shallow recognition by presenting both the recognized emotion and its potential cultural valence, thereby fostering greater user trust and nuanced, context-aware interaction.

In summary, the main contributions of our paper is as following:

- A cross-cultural investigation that integrates AU-based analysis of facial expressions with valence-based interpretation

to explain how *surprise* is recognized and understood across cultures.
- Empirical evidence demonstrating that the same facial expressions of *surprise* are categorized and emotionally evaluated in culturally distinct ways by South Korean and American participants.
- An exploratory discussion of potential cultural biases in existing FER datasets, highlighting the need for more culturally diverse emotion annotations.
- Design implications for HCI, outlining directions for developing culturally adaptive FER systems and emotion-aware interfaces that support inclusive, multicultural interaction.

2 Backgrounds and Related Work

To situate our work within the existing literature, we reviewed related research on cultural variation in FER and the challenges it poses for FER-based systems. We also examined research on the psychology of surprise as a culturally sensitive emotion that offers a lens for studying cultural interpretation.

2.1 Importance of Understanding Cross-Cultural Variation in FER

Facial emotion recognition (FER) plays an important role in building emotion-aware intelligent agents and affective computing systems. The fields of HCI and human-AI interaction (HAI) have adopted FER models to recognize and interpret human emotions from facial expressions in various domains, including healthcare [31, 57], e-learning [32, 77], driver-assistance [40], security [68], and other well-being applications [6, 8, 75]. Recent advances in large language models (LLMs) and sensing technologies have improved FER systems by integrating contextual cues with multimodal inputs, including text, speech, body gestures, and sensory data, thereby leveraging richer information to enhance model performance [14, 21, 42, 73, 76]. For example, Nepal *et al.* [57] introduced an approach to detect depression using facial images automatically captured by a smartphone's front-facing camera, trained with image-based features (facial cues and image characteristics). Yi *et al.* [78] developed FER-based emotion tracking tool using laptop cameras and explored user's experiences and perceptions in daily life. Nahulanthran *et al.* [56] proposed an explanation method that utilizes facial action units (AUs) to better interpret FER model predictions.

Despite the promise of FER technologies, their deployment in daily contexts remains limited by dataset bias and model misalignment arising from cross-cultural variation [7]. Research in psychology and neuroscience has argued that basic emotional facial expressions (*i.e.*, happiness, sadness, fear, surprise, anger, and disgust) are universally recognized [16, 19], and most FER machine learning datasets are built on these basic emotions [2, 27, 44, 54]. However, recognition rates for these emotions are not consistent among cultures, particularly between Western and Eastern populations [30, 49, 52]. For instance, Matsumoto [48] compared American and Japanese participants' recognition of universal facial expressions and found that, although both groups could identify basic emotions, Japanese participants showed lower agreement for negative emotions. Similarly, recent findings showed that *fear* perceived by Western observers for non-masked faces was often interpreted as *surprise* by East Asian participants for both masked and non-masked faces, revealing systematic cross-cultural differences in emotion perception [55].

Such perceptual differences are likely to extend to the process of data annotation itself, as annotators inevitably interpret facial emotions through their own cultural lens. While previous studies have actively discussed racial and demographic biases in the composition of facial expression datasets [12, 13], little attention has been paid to the cultural background of annotators themselves. Despite evidence that emotion perception differs across cultures, major FER datasets were annotated by groups with relatively homogeneous cultural profiles. For instance, AffectNet [54] was labeled by 12 American university students, and similar Western annotator recruitment practices can be found in datasets such as IEMOCAP [5] or CK+ [37] datasets.

To address these limitations, a growing body of research investigates emotional perception gaps across cultures and offers insights for crowdsourcing and debiasing approaches [7, 45, 80], underscoring that cultural variation should be considered in the design of future FER systems. Inspired by this line of research, we aim to provide the first mixed-methods analysis to explicitly link visual recognition discrepancies with emotional interpretation differences for an ambiguous emotion like *surprise* between two distinct cultures. Given its ambivalent nature [61], examining how surprise is recognized and interpreted across cultures is crucial for developing fair and context-aware FER systems.

2.2 Surprise as a Culturally Sensitive Emotion in Cross-Cultural FER

Whereas most other basic emotions identified by Ekman and Friesen [16] are associated with a clear-cut valence, as they are experienced as either positive or negative, the case of surprise is not as straightforward [59, 61]. A surprise can feel delightful when walking into a room full of friends shouting "Happy Birthday," yet deeply unsettling upon hearing sudden bad news about a loved one—the same emotion, but an entirely different feeling. Large-scale dimensional studies have consistently placed surprise near the neutral point of the valence axis, characterized by high arousal but ambiguous positivity or negativity [25, 67]. In this sense, surprise is ambiguously valenced, experienced as either positive or negative depending on situational cues and the perceiver's background [59]. Empirical evidence also shows that the valence of surprised faces is shaped by environmental [38, 58] and cultural contexts [39], highlighting its context sensitivity. This makes surprise uniquely informative for studying how people from different cultures interpret emotionally ambiguous expressions, providing insight into the interplay between recognition and interpretation in facial expressions.

Despite extensive cross-cultural research on FER, most studies have examined cultural differences across basic emotions as a whole [34, 35, 55, 63, 74]. For instance, Jack *et al.* [35] challenged the universality hypothesis by reconstructing how Western and East Asian observers mentally represent the six basic emotions, revealing overlapping representations—especially for surprise, fear, disgust, and anger—among Eastern participants. These findings demonstrate that even canonical emotional expressions vary across cultural contexts.

Only a limited number of studies have examined specific emotions in detail rather than addressing basic emotions as a whole. In particular, cross-cultural FER research has rarely concentrated on this single ambiguous emotion, despite the cultural dependency of surprise discussed earlier. When such attempts exist, they tend either to analyze the cognitive and affective characteristics of the emotion within psychology, rather than investigating them through FER [51, 60], or to distinguish surprise from morphologically and cognitively similar emotions such as fear, without addressing cultural variation [28, 29, 66]. In contrast, the present study centers on surprise itself to address this gap, bridging recognition and interpretation processes to understand why cultural differences emerge in how this ambiguous emotion is perceived and evaluated.

To better contextualize cultural interpretation of ambiguous emotions, we consider surprise's close relationship with fear—a clear-valence emotion with which surprise shares notable morphological and cognitive similarity [11, 28, 48, 66]. By jointly examining these two related emotions, we aim to understand not merely how recognition differs across cultures, but why such differences emerge.

3 Method

In our main study, we systematically investigate our research questions through an online survey with 36 South Korean and 34 American participants. The survey consists of three main tasks: facial expression labeling (*surprise* or *fear*), valence rating, and surprise recall, each addressing different aspects of cross-cultural **recognition** and **interpretation**. Before introducing the detailed procedure and rationale for each task in this main survey study, we summarize our preliminary findings from a pilot study using existing labeling datasets, which informed the design of our main study.

3.1 Primary Evidence of FER Dataset Bias

We performed a pilot study using two widely used FER datasets: AffectNet [54] and the Combined Japanese and Caucasian Facial Expressions of Emotion (JACFEE) and Neutral Faces (JACNeuF) [47].

AffectNet consists of facial expression images annotated with eight emotion categories (basic emotions [15] and neutral) by 12 American university students. To examine potential cultural bias in these annotations, we recruited 59 South Korean university students (38 male, 21 female; age $M = 21.08$ and $SD = 2.71$) through a university mailing list. We randomly selected 28 facial expression images from AffectNet for each emotion category, totaling 224 images. We compared the aggregated majority labels provided by South Korean participants with the original annotations. We measured the inter-annotator agreement between South Korean participants and the original American annotators using Fleiss' kappa [24]. As shown in Figure 2, the results showed a moderate average agreement of 0.411. However, agreement varied across emotions, with surprise showing the lowest value (0.183), indicating only slight agreement.

We additionally looked into the Combined JACFEE and JACNeuF dataset, which consists of posed facial expression images annotated with eight emotion categories (basic emotions and neutral) based on AUs by experts. We recruited 275 South Korean participants (aged from 18 to 59) by snowball sampling and online advertising. We used all 64 facial expression images from JACFEE and JACNeuF, with eight images per emotion category. As a result, the recognition

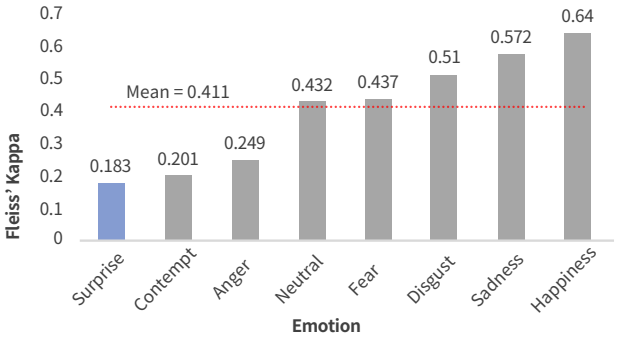

Figure 2: The plot shows the level of agreement between the emotion labels generated in our pilot study by South Korean university students and the original labels created by American annotators for randomly sampled 224 facial expression images from the AffectNet dataset. The level of agreement was measured using Fleiss' kappa. In the plot, the red horizontal line represents the mean value (0.411).

patterns by South Korean participants were largely aligned with the original AU-based labels (Figure 3). However, recognition of specific emotions (*i.e.*, neutral, surprise, disgust, and anger) was distributed across multiple emotion categories. In particular, approximately 32% of the images labeled as surprise by South Korean participants were originally intended to represent fear. In other words, the South Korean participants frequently recognized AU-based fear expressions as surprise—a pattern that differed from that observed in Western populations [3].

Together, our pilot study results suggest that South Korean participants have low agreement with previously reported Western annotations, with especially low agreement for surprise and fear expressions. Based on these preliminary findings, we systematically investigate the cultural difference of the recognition and interpretation of surprise emotion using AffectNet dataset. Below sections provide detailed procedure of our main study.

3.2 Participants and Procedure

We recruited participants by advertising our study on a local social community platform In South Korea and on Prolific [64]. Our inclusion criteria were participants who are: (1) born and currently residing in South Korea or the United States, (2) having parents of the same cultural background, and (3) reporting no prior experience of living in another cultural environment for 36 months or longer. One American applicant was excluded due to mismatched parental cultural background.

A total of 36 South Korean participants (K1–K36; 9 male, 27 female) and 34 American participants (A1–A34; 16 male, 15 female, 3 non-binary) completed the study. The age distribution was 18–24 years (8.3%), 25–34 years (38.9%), and 35 years or older (52.8%) for South Korean participants, and 18–24 years (14.7%), 25–34 years (52.9%), and 35 years or older (32.4%) for American participants.

The online survey took approximately 30 minutes to complete. South Korean and American participants received 10,000 KRW and £5 (approximately 7.50 USD each), respectively, as compensation.

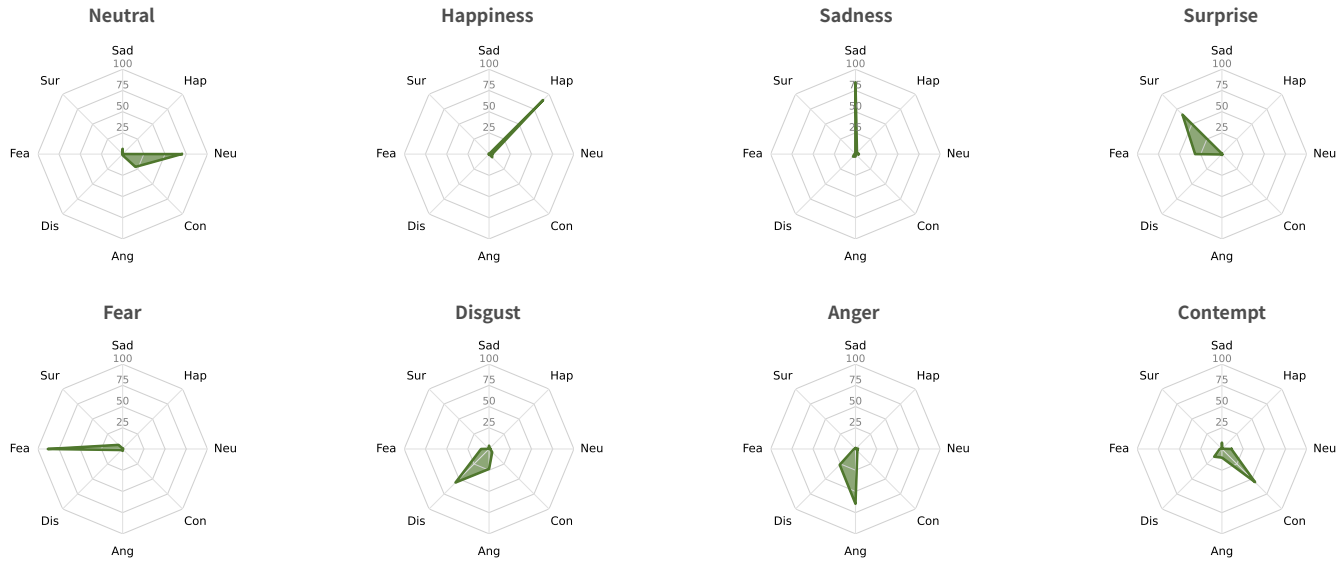

Figure 3: Radar charts illustrating recognition patterns for 64 facial expression images from JACFEE and JACNeuF across eight emotion categories by South Korean participants ($N = 275$) in our pilot study. Each octagonal radar chart represents the set of images that participants labeled with that emotion, showing the distribution of the original JACFEE and JACNeuF labels across the eight emotion axes.

The survey consisted of three sequential tasks designed to address the two research questions. In the first task, participants were asked to describe two of their most recent surprising experiences, specifying what happened and how they felt during the situation. This task aimed to capture the culturally specific valence and context associated with surprise experiences, thereby providing qualitative insight into the **interpretation** (RQ2) dimension of surprise.

In the second and final tasks, participants viewed the same set of 200 facial expression images (100 from *surprise* category and 100 from *fear* category) randomly selected from the AffectNet dataset. The images were presented in a randomized order.

In the second task, participants categorized each image as either "surprise" or "fear." The decision to focus only on these two labels was theoretically grounded. *Surprise* and *fear* often look similar in facial appearance and are easily confused because they share several expressive features. Previous research in FER has consistently reported high confusion rates between these two emotions [11]. Moreover, prior cross-cultural studies suggest that East Asian observers tend to interpret facial expressions as mixed emotions [22, 23], and particularly Koreans have been shown to associate surprise with more negative valence compared to Western norms [39]. This makes *fear* an appropriate comparative category for examining potential cultural differences in how ambiguous *surprise* expressions are recognized. This task therefore allowed us to investigate cross-cultural differences in the **recognition** (RQ1) of ambiguous expressions, specifically whether differences in AU processing lead cultural groups to categorize the same expressions differently.

In the final task, participants rated the perceived valence of each image on a 7-point Likert scale (1: very negative and 7: very positive). Valence judgments were intentionally collected in a separate task to minimize anchoring effects from categorical emotion labeling, which could otherwise bias valence evaluations for ambiguous expressions such as surprise. This provided a direct, quantitative measure of the emotional **interpretation** (RQ2), complementing the categorical recognition data from the second task. A visual summary of the procedure is shown in Figure 4. All participants provided informed consent prior to participation, and the study protocol was approved by our Institutional Review Board.

3.3 Data Analysis and Measurement

3.3.1 Recognition Analysis. We first assessed the aggregated annotations within the two cultural groups (South Korean and American participants) to understand the overall cross-cultural patterns in facial expression recognition. Next, we ran a focused analysis using AU activation [17] on images labeled *surprise*, examining valence-related AU activation to explore how specific facial cues were associated with recognition differences between the two groups.

As a preliminary step, we analyzed the aggregated labels assigned to each image based on the majority voting within each culture to capture overall tendencies in facial expression recognition. Each of the 200 facial expression images (100 from *surprise* category and 100 from *fear* category of AffectNet dataset) was assigned a final label for each cultural group based on the majority response of the participants. A chi-square test compared the overall distribution of the two label types between South Korean and American participants, followed by post-hoc z-tests for proportions to locate specific group differences. In addition, a cross-cultural confusion matrix was constructed to visualize disagreement patterns

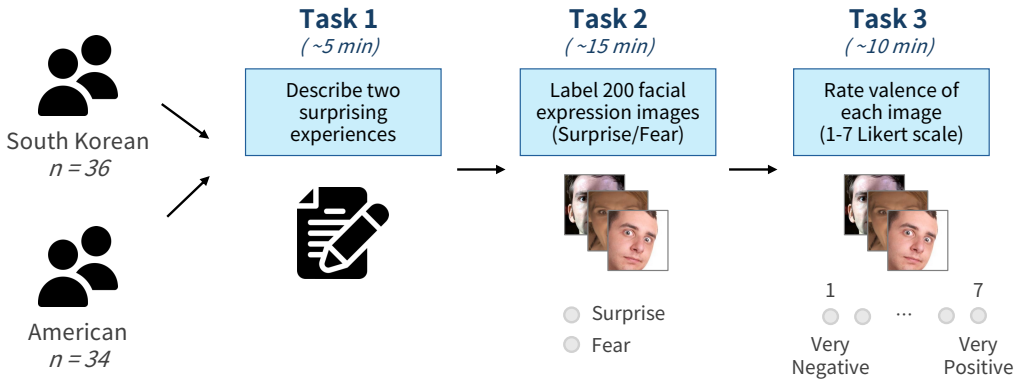

Figure 4: Overview of the online survey procedure.

between the two groups, offering an overview of how differently each culture categorized the same facial expressions. This way, we could establish general recognition tendencies between the two different cultures.

Next, we focused specifically on images labeled as *surprise* to explore which facial cues were most associated with cross-cultural differences. Facial AU features were extracted from all images using OpenFace 2.0 [79], an open-source toolkit for facial behavior analysis. Rather than analyzing the mean intensity of each individual AU, we focused on AUs that have been theoretically linked to affective valence. This approach was chosen because it was considered more appropriate for capturing whether the facial cues underlying *surprise* recognition convey different emotional tones across cultures. Based on prior research connecting particular AUs to positive and negative valence [81], we referred to the AUs that were significantly correlated with valence in that study. Specifically, we utilized the results of their SHAP (SHapley Additive exPlanations) [46] analysis, which computed mean absolute SHAP values for each AU in a random forest regression model to quantify each feature's contribution to subjective valence and arousal rating predictions. From these results, we selected the top five AUs that showed the strongest influence on valence ratings and categorized them into *positive* (AU6: Cheek Raiser, AU9: Nose Wrinkler, AU12: Lip Corner Puller) and *negative* (AU4: Brow Lowerer, AU15: Lip Corner Depressor) groups. The final AU categorization is summarized in Table 1.

Because individual AUs often co-occur and jointly contribute to perceived emotional tone, this aggregation supports cross-cultural comparison at the level of emotional interpretation rather than fine-grained AU-level patterns. To statistically assess differences in valence-oriented facial activation, we compared the mean activation levels of positive and negative AU groups within each cultural sample using paired-samples t-tests. To control for multiple comparisons across cultural groups, p-values were adjusted using the Bonferroni correction. Comparing the patterns observed in each group allowed us to examine whether the two cultural groups differed in the valence orientation of facial cues associated with *surprise* expressions.

3.3.2 Interpretation Analysis. To investigate how recognition patterns relate to cultural variations in emotional interpretation, we

Table 1: Classification of Positive and Negative Action Units based on prior research [81].

Category	Action Units (AUs)
Positive	AU06 (Cheek raiser), AU09 (Nose wrinkler), AU12 (Lip corner puller)
Negative	AU04 (Brow lowerer), AU15 (Lip corner depressor)

conducted two complementary analyses: a quantitative comparison of participants' valence ratings for facial expressions labeled as *surprise*, and a qualitative analysis of their written descriptions of surprising experiences. Together, these analyses aimed to capture both the evaluative and contextual dimensions of how the emotion of surprise is interpreted across the two cultural groups.

For the quantitative analysis, we focused on images labeled as *surprise* by each cultural group and compared their mean valence ratings using an independent-samples t-test to determine whether the two groups differed significantly in their affective evaluations.

For the qualitative data, participants' written descriptions of two surprising experiences each were analyzed using thematic analysis in ATLAS.ti. ATLAS.ti was selected as a widely used qualitative analysis tool that supports systematic code organization and transparent collaboration across multiple coders, facilitating consistent interpretation across cultural groups. In total, 140 accounts were collected from 70 participants (36 South Korean and 34 American). Prior to the thematic analysis, responses were screened to ensure analytic relevance and interpretability. Responses that did not clearly convey a positive or negative affective valence (*e.g.*, neutral descriptions) or that did not align with emergent situational themes were excluded. Based on these criteria, 17 responses were excluded (8 South Korean and 9 American), resulting in a final dataset of 123 responses used for thematic analysis. All responses were first coded by two South Korean researchers, focusing on two key aspects: the valence of the experience (positive or negative) and the situational themes (*e.g.*, natural events such as thunder, personal mistakes, or social interactions). To mitigate potential cultural bias, one American researcher cross-checked the coded data to ensure that the coding decisions were culturally appropriate

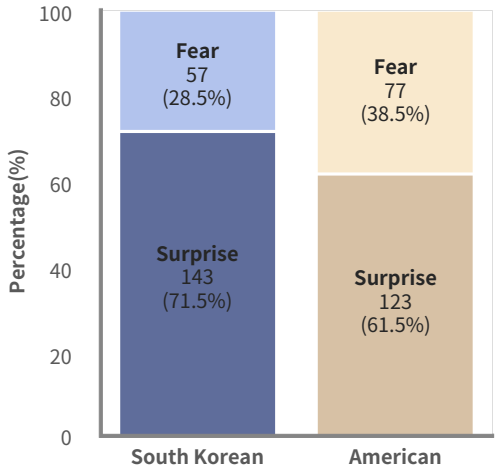

Figure 5: Number of images labeled as *surprise* or *fear*. Each image was labeled based on the aggregated majority voting within each cultural group. South Korean participants labeled 71.5% of the images as *surprise* and 28.5% of the images as *fear* after aggregation, while American participants labeled 61.5% of the images as *surprise* and 38.5% of the images as *fear* after aggregation. This indicates that South Korean participants shown a more conservative attitude in labeling negative emotions to facial expressions.

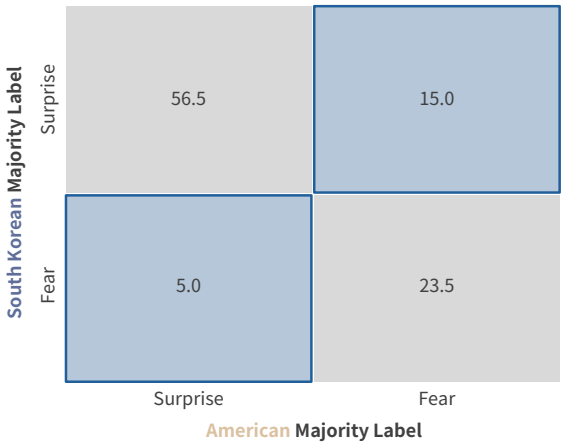

Figure 6: Cross-cultural confusion matrix between aggregated emotion labels of South Korean and American participants. Among images with disagreements (the two anti-diagonal cells), three times more images were labeled as *fear* by American participants, while being labeled as *surprise* by South Korean participants (15%), compared to being labeled as *surprise* by American participants, while being labeled as *fear* by South Korean participants (5%). This asymmetry suggests a relatively more conservative tendency among South Korean participants in assigning negative emotion labels.

and interpretable within context. This qualitative analysis provided complementary insights into how participants from each cultural group emotionally framed and contextualized their experiences.

4 Results

The results are presented in two parts corresponding to the study's main research questions. Section 4.1 demonstrates cross-cultural differences in the recognition of facial expressions through the analyses of AU activation patterns associated with images labeled as *surprise*. Section 4.2 reports cultural differences in the interpretation of these expressions through a mixed-method, where the valence rating of expressions was statistically analyzed and the written descriptions of participants' surprising experiences were qualitatively analyzed.

4.1 RQ1: Cross-Cultural Recognition Differences and AU Pattern Analysis

We analyzed labeling tendencies for both surprise and fear, and automatically computed facial AU patterns for expressions labeled as *surprise*. This section identifies whether cultural groups diverge in *how* they categorize and visually process these expressions.

4.1.1 Overall Aggregated Labels within Cultures and Their Comparison. We first examined the overall distribution of emotion labels to identify general cross-cultural tendencies (Figure 5). A chi-square test revealed a significant difference in the overall distribution between South Korean and American participants, $\chi^2(1, N = 400) = 4.05$, $p = .0441$, indicating that the two cultural groups showed distinct tendencies in emotion recognition. Post-hoc proportion tests further identified which emotions contributed to this difference. For *surprise*, South Korean participants labeled a higher proportion of images as surprise ($z = 2.12$, $p = .0341$). In contrast, for *fear*, South Korean participants were significantly less likely to assign the fear label compared to Americans ($z = -2.12$, $p = .0341$), suggesting a more conservative tendency in labeling negative emotions.

Next, we constructed a cross-cultural confusion matrix based on the labels (Figure 6). Overall, label agreement was higher for *surprise* (56.5%) than for *fear* (23.5%) between the two cultures, indicating that both groups shared relatively similar criteria for identifying surprise expressions. However, asymmetrical patterns emerged in disagreements: facial expressions perceived as *fear* by American participants were often recognized as *surprise* by South Korean participants (15.0%), whereas the reverse (expressions recognized as *fear* by South Koreans but as *surprise* by Americans) occurred less frequently (5.0%). This pattern indicates that the two cultural groups may rely on partly different perceptual cues when recognizing these emotions. Notably, the tendency for expressions labeled as *fear* by American participants to be recognized as *surprise* by South Korean participants aligns with prior research showing that facial expressions perceived as *fear* by Western participants were perceived as *surprise* by East Asian participants [55]. Together, these findings suggest that South Korean participants may recognize *surprise* expressions in a slightly more *fear-like* manner compared to American participants.

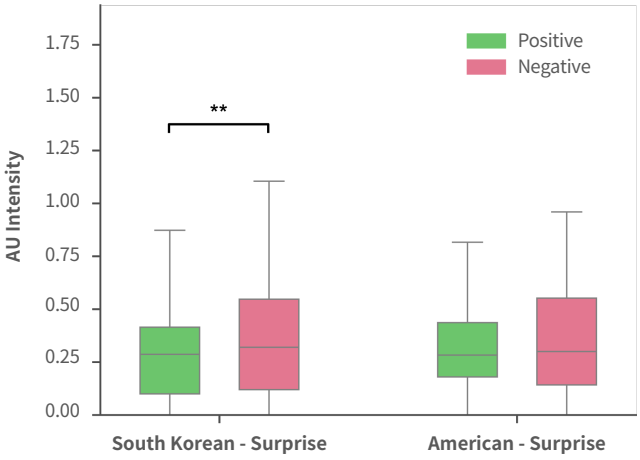

Figure 7: Average intensity of both positive and negative AUs detected in images annotated as *surprise* by the two cultural groups. The intensity of negative AU activations was significantly higher than positive AU activations in images labeled as *surprise* by South Korean participants. There were no significant differences in positive and negative AU activations in images labeled as *surprise* by American participants.

These distributional differences suggest that, even when viewing the same facial expressions, participants from different cultural backgrounds may rely on distinct facial cues to identify emotional categories. To further examine these cue patterns, we next conducted an AU analysis comparing South Korean and American participants, with an emphasis on positive and negative facial muscle activations associated with *surprise* recognition.

4.1.2 Differences in Positive and Negative AU Activation. Figure 7 illustrates the summarized distribution of positive and negative AU activation intensities for the expressions labeled as *surprise* by each cultural group. While the boxplots visualize the overall distribution patterns, paired-samples t-tests were conducted to statistically compare the mean activation levels between positive and negative AUs within each cultural group. To account for multiple comparisons, p-values were adjusted using the Bonferroni correction. Positive and negative AUs were categorized based on prior work on valence-related facial cues, as mentioned in Section 3.3.1. As shown in the figure, facial expression images labeled by the South Korean group exhibited a significant difference between positive and negative AU activation ($t = -2.80$, $p_{\text{adj}} = .0116$), whereas no such difference was found in the American group ($t = -0.46$, $p_{\text{adj}} = 1.00$). Specifically, images labeled as *surprise* by South Korean participants showed lower activation in positive AUs compared to negative AUs, suggesting that their perception of surprise expressions involved relatively more negative facial cues. In contrast, images labeled as *surprise* by American participants showed comparable levels of positive and negative AU activation, implying that the expressions they recognized as *surprise* included more balanced, or slightly more positive, facial signals. These patterns suggest that American participants may interpret *surprise* as an emotion with a more ambivalent or bidirectional valence, whereas South Korean participants tend to perceive it with a more cautious or negative bias.

4.1.3 Summary. Even when viewing the same facial expression images, participants from different cultures categorized the expressions differently. The analysis of *surprise*-labeled faces revealed that each group emphasized facial cues with distinct emotional tones, indicating culturally specific ways of perceiving surprise. South Korean-labeled *surprise* images showed lower activation in positive AUs and higher activation in negative AUs, suggesting more tension- or threat-related cues. In contrast, American-labeled *surprise* images exhibited more balanced activation patterns, reflecting relatively ambivalent signals. Taken together, these results suggest that the two cultural groups recognized ambiguous facial expressions through different affective lenses—South Korean group with a more cautious or negatively biased sensitivity, and American group with a more open or bidirectional interpretation.

4.2 RQ2: Cross-Cultural Differences in Valence Interpretation

To explore the underlying causes of these recognition differences, we next examined *why* each cultural group interpreted the emotional valence of surprise expressions differently. We hypothesized that divergent affective evaluations—how positively or negatively surprise is experienced—may explain the cross-cultural variation observed in recognition patterns. Accordingly, we analyzed both quantitative valence ratings and qualitative narrative descriptions.

4.2.1 Valence Ratings of Surprise Expressions. To examine how each cultural group affectively interprets the facial expressions they recognized as *surprise*, we analyzed the valence ratings assigned to images labeled as *surprise* by South Korean and American participants, respectively (Figure 8). American participants rated their *surprise* images with a significantly higher mean valence ($M = 4.15$, $SD = 1.45$) than South Korean participants ($M = 3.76$, $SD = 1.38$). This difference was statistically significant according to a two-sample t-test ($t(8761) = 13.15$, $p < .001$). These results indicate that the images recognized as *surprise* by American participants were generally associated with higher valence ratings, suggesting that American participants tended to interpret *surprise* as a relatively more positive emotional experience. In contrast, the images labeled as *surprise* by South Korean participants showed lower and more neutral-to-negative valence ratings, reflecting a more cautious or tension-oriented interpretation of the same emotion category.

4.2.2 Qualitative Insights into Cultural Interpretation of Surprise. To gain a deeper insight into the cultural nuances in how surprise is experienced and interpreted, participants were asked to recall two personal situations in which they had felt surprised (see Table 2). South Korean participants predominantly described negative experiences (89.0%), whereas American participants exhibited a more even distribution of affective valence, with positive experiences (52.5%) occurring marginally more often than negative ones (47.5%). This pattern suggests that South Korean participants tended to associate surprise with unpleasant or aversive contexts, while American participants demonstrated comparatively greater affective diversity in their recalled surprise experiences.

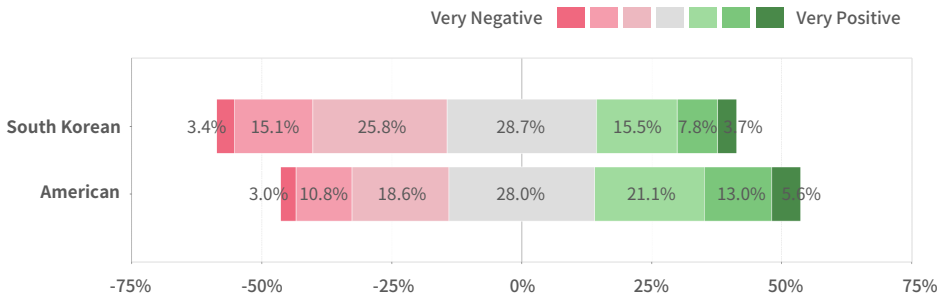

Figure 8: Valence ratings of *surprise* expressions between South Korean participants and American participants. American participants rated their *surprise* images with higher valence than South Korean participants, indicating more positive interpretations of *surprise*.

Table 2: Codebook of positive and negative surprising experiences with frequencies by cultural group. Numbers indicate how many respondents mentioned each code.

Category	Code	Explanation	Example	South Korean	American
Positive	Party/gift	Surprise parties, gifts, or other forms of celebration received from close others.	"*My brother surprised the entire family with a Disney trip!*"(A1)	2	11
	Achievement	Accomplishments or successes of oneself or close others.	"*My colleague won the Employee of the Month award.*" (K25)	2	4
	Health recovery	Recovery, improvement, or positive medical news regarding oneself or close others.	"*My younger sister was brought home from the hospital.*"(A34)	0	1
	Supportive reaction	Supportive behaviors or unexpected kindness from others.	"*My wife reassured me that I could quit my job if I wanted to.*"(K3)	2	3
	Reconnection	Unexpected encounters or renewed contact with old friends.	"*I went to a housework part-time job, and it turned out to be my old elementary school classmate's place.* "(K13)	1	8
	Financial gain	Lottery wins, bonuses, or other unexpected monetary benefits.	"*A man gave me a winning lottery ticket for seemingly no reason.*"(A30)	0	4
Negative	Accident	Traffic collisions or other physical harms that occurred or almost occurred.	"*I hit a taxi while trying to change lanes.*"(K22)	9	5
	Health decline	Illness, injuries, death, or other medical issues experienced by oneself or close others.	"*My father was diagnosed with stage 4 cancer.*"(K5)	13	5
	Interpersonal disruption	Negative and unexpected reactions or behaviors from others in social relationships.	"*My partner was upset with me because I wasn't able to make all of the arrangements he wanted for Father's Day.*"(A8)	5	5
	Financial threat	Phishing, scams, fraud, and other monetary risks experienced by oneself or close others.	"*A phishing scammer called me, pretending there was a delivery for a credit card I never applied for.*"(K17)	5	1
	Violent incident	Physical fights or confrontations, between individuals or involving law enforcement.	"*Two men were yelling and grabbing each other by the collar in a café, and I was sitting at the next table.*"(K14)	2	3
	Animal encounter	Unexpected encounters with animals or insects.	"*When I shook the shorts off the drying rack to put them away, a black spider fell out.*"(K36)	6	5
	Environmental event	Thunder, lightning, heavy rain, or other sudden natural phenomena.	"*I was in the parking lot when it suddenly started thundering and lightning.*"(K10)	7	0
	Work disruption	Layoffs, personnel changes, or other unexpected changes in the workplace.	"*I was informed of my termination by the outsourcing company's manager.*"(K17)	5	0
	Media shock	Shocking news or events encountered through media.	"*I saw a news article while I was browsing news articles on my laptop at home which mentioned that Ozzy Osbourne had passed away.*"(A24)	5	4

For South Korean participants, surprise was often associated with tension, fear, or disappointment triggered by sudden threats or unsettling events. These narratives frequently reflected themes of loss of control and social discomfort, often emerging from interpersonal disruptions or abrupt changes in daily routines. For example, one participant recalled, *"A client harshly vented their emotions over the phone. I was really angry."* (K3) illustrating how unexpected social confrontation evoked frustration and emotional strain. Another described, *"My father was taken to the emergency room."* (K27), reflecting the distress and helplessness evoked by sudden health-related threats. Such accounts highlight that surprise was interpreted as a form of negative emotional arousal linked to frustration, anxiety, and perceived loss of control in social contexts.

In contrast, American participants described a more bidirectional interpretation of surprise, with both positive and negative emotional tones. While some experiences involved stress or fear—such as sudden accidents or unsettling encounters—many others reflected excitement, gratitude, or curiosity arising from pleasant surprises. For instance, one participant shared, *"My three-year-old son suddenly held my hand and said, 'Daddy, you're my best friend.' I felt utterly frozen—it was pure, unfiltered love that caught me off guard."* (A16), describing a mix of excitement and heartfelt appreciation. Another mentioned, *"My manager unexpectedly praised my work in front of everyone, highlighting a project I had been working on quietly,"* (A14) expressing pride and pleasant astonishment at receiving public recognition. These accounts indicate that American participants were more likely to view surprise as a contextually flexible emotion, encompassing both pleasant and unpleasant forms of unexpectedness depending on situational cues.

Overall, these qualitative patterns align with the valence rating results. South Korean participants described surprise primarily within a negative frame, while American participants exhibited a more bi-directional interpretation, encompassing both pleasant and unpleasant aspects of unexpectedness.

4.2.3 Summary. Collectively, RQ2 findings indicate that cultural interpretation of surprise diverged along the valence dimension. American participants tended to perceive surprise as a more positive and open-ended emotional experience. In contrast, South Korean participants often interpreted surprise within negative or tension-related contexts, reflecting a more cautionary response to unexpectedness. Together with the recognition analyses, these findings suggest that cross-cultural differences in emotion perception may arise not only from visual processing of facial cues but also from culturally shaped interpretations of emotional meaning.

5 Discussion

5.1 Considering Cultural Differences in FER Datasets

Our findings highlight the importance of considering cultural variance in FER datasets and AI models, rather than assuming emotion universality. Future FER dataset construction could consider not only balancing demographic factors such as gender and age, but also deliberately diversifying the cultural backgrounds of annotators. Existing datasets can further benefit from systematic re-annotation efforts, for example, by leveraging gamified annotation platforms to scale participation [77], or by selectively targeting culturally ambiguous categories such as fear and surprise to refine label quality. Future research could employ advanced aggregation methods to preserve and leverage annotator diversity, rather than relying on majority voting. Work in crowdsourcing has shown that leveraging diverse human judgments can provide a more reliable estimate of ground truth. For instance, Song *et al.* [69, 70] demonstrated that aggregating diverse responses with different error patterns across tools yields more accurate results. Chung *et al.* [10] showed that for subjective tasks such as emotion labeling, the most effective approach was to retain all plausible labels. Building on existing line of work, advanced aggregation techniques can effectively incorporate subjective annotations into FER datasets. On the modeling side, features such as positive vs. negative AU activations should be used with cultural sensitivity, possibly weighting features differently by predicted user culture. Moreover, rather than relying on hard categorical labels, future models could train with soft/label distributions, incorporate bias-aware regularization, or adopt adversarial debiasing techniques to reduce culture-based bias and better account for annotator variability. Together, these steps can help future FER systems better reflect cultural diversity and reduce annotation bias in cross-cultural deployments.

5.2 Design Implications for Culture-Aware Emotion Recognition Systems

Our findings reveal that cultural background shapes how people interpret ambiguous emotions such as *surprise*. Even when viewing identical faces, South Korean and American participants' interpretations diverged not only in categorical labeling but also in the emotional tone of the cues they emphasized. South Korean participants tended to rely on facial features associated with tension or fear, while American participants interpreted the same expressions through more balanced or ambivalent cues. These results underscore that emotion recognition is not purely a visual process but a culturally grounded interpretation of affective meaning. We outline the following design implications for culture-aware FER systems.

Account for cultural diversity in annotation for ambiguous emotions. While relatively less ambiguous emotions such as happiness or sadness tend to exhibit relatively stable patterns of recognition and interpretation across cultural contexts, our findings indicate that ambiguous emotions like *surprise* are particularly sensitive to cultural variation in both recognition and interpretation. For such emotions, relying on a small or culturally homogeneous annotator pool risks collapsing meaningful interpretive differences into a single dominant label, masking culturally grounded patterns of emotional meaning. To better capture this variability, FER datasets should allocate greater annotation effort and cultural diversity to emotions with high interpretive ambiguity, rather than uniformly sampling annotators across all emotion categories. This may involve recruiting annotators from multiple cultural backgrounds, engaging a larger pool of annotators for culturally ambiguous emotions, or explicitly preserving minority or non-majority labels rather than enforcing strict consensus. Such annotation strategies can help retain culturally meaningful ambiguity at the data level, providing a more faithful foundation for downstream modeling and interpretation of emotions like *surprise*.

Represent culturally ambiguous emotions using probabilistic labels. Rather than assigning a single categorical label, FER systems should support probabilistic representations that capture multiple plausible interpretations. This representation allows systems to communicate uncertainty and capture the ambivalence that often characterizes emotions like *surprise*. For example, a surprised facial expression might be interpreted as 60% surprise and 40% fear for one user, but 70% surprise and 30% happiness for another, depending on cultural framing. By explicitly representing such probabilistic mixtures, systems can preserve cultural ambivalence, improve transparency, and reduce the risk of misinterpretation in downstream applications, while enabling more flexible and empathetic interactions.

Use multimodal and contextual cues to support emotion interpretation. Facial signals alone may not fully convey the intended emotional state of culturally ambiguous emotions such as *surprise*, particularly in cross-cultural contexts where expression norms and display rules vary. To reduce the risk of misinterpretation, future FER systems should integrate complementary cues such as vocal tone, physiological signals, or contextual information to cross-validate emotional inferences. This multimodal integration is especially important in high-stakes domains such as healthcare or education, where confusing tension-related surprise with curiosity-related surprise may lead a system to overlook patient distress or misjudge a student's confusion as engagement.

Ultimately, advancing toward inclusive and culture-aware emotion recognition requires rethinking emotion AI as a contextual and interpretive process rather than a static visual classification problem. Systems that acknowledge and adapt to cultural diversity in emotional expression and perception will not only achieve greater cross-cultural accuracy but also foster more equitable and trustworthy human-AI relationships.

5.3 Limitations and Future Work

Our work has several limitations. First, our study cannot be taken as representative of all East Asian or Western populations. Although many prior studies have drawn this distinction, treating samples from East Asian (*e.g.*, Japanese, Chinese) and Western (*e.g.*, American) populations as representative of their respective cultural groups [33, 35, 55], these categories are far from homogeneous. Recognition tendencies and emotional display rules can differ substantially not only between neighboring countries but also among subcultures within the same society [41]. Future work should therefore expand participant sampling to include a broader range of cultural backgrounds, such as Southeast Asian, European, and Middle Eastern populations, to enable more fine grained cross cultural comparisons. Such efforts would help build a richer understanding of how cultural norms, language, and context jointly shape the perception of ambiguous emotions like surprise, ultimately informing the design of more inclusive and culturally aware emotion recognition systems.

Second, there is an inherent conceptual and linguistic overlap between emotion categories. Emotional vocabulary rarely maps one-to-one across languages, meaning that the semantic boundaries of emotion words themselves may differ across cultures [36, 62]. As a result, people from different linguistic backgrounds might not only interpret an expression differently, but also conceptualize the categorical space of emotions in distinct ways, leading to systematic biases in interpretation. Future research should integrate linguistic and semantic analyses to better clarify how emotional concepts are represented and communicated in different cultural contexts.

6 Conclusion

This study examined how people from different cultural backgrounds recognize and interpret ambiguous emotional expressions, focusing on the emotion *surprise*. By combining AU based analyses with valence ratings and qualitative accounts, we found that the same facial expressions were categorized and interpreted differently between the two cultural groups. Facial expressions labeled as *surprise* by South Korean participants showed lower activation of positive AUs and higher activation of negative AUs, suggesting that the participants tend to recognize *surprise* through tension- or threat-related cues. In contrast, facial expressions labeled as *surprise* by American participants displayed more balanced activation patterns, suggesting relatively more positive or ambivalent interpretations. This pattern was also reflected in their valence ratings and narrative descriptions.

These findings highlight that emotion perception is not universal but jointly shaped by perceptual cues, emotional meaning, and cultural context. By linking recognition (the reading of facial cues) with interpretation (the evaluation of emotional meaning), this study extends traditional FER models to account for cultural variability and ambiguity in *surprise*.

Future emotion recognition systems should incorporate culturally adaptive classifiers and contextual information to improve cross cultural accuracy. Such approaches can support the development of more inclusive and culturally aware affective computing systems that better reflect the diversity of human emotional expression.

Gen AI Usage Disclosure

The authors used generative AI tools (ChatGPT, GPT-5 by OpenAI) for minor language editing and phrasing suggestions. All research content, analyses, and interpretations were created and verified by the authors.

Acknowledgments

We thank the participants of our pilot study and online survey for their time and effort. We also thank the members of the DIAG Lab for their helpful discussions and feedback throughout this work. This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIT) (Nos. RS-2024-00353125 and 2022R1C1C1003123).

References

[1] Naseem Ahmadpour, Danielle Lottridge, Jonas Fritsch, Corina Sas, Marta E Cecchinato, Daniel Harrison, Kristina Höök, Pin Sym Foong, Kiran Ijaz, Phillip Gough, et al. 2025. Affective interaction and affective computing-past, present and future. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–6.

[2] Emad Barsoum, Cha Zhang, Cristian Canton Ferrer, and Zhengyou Zhang. 2016. Training deep networks for facial expression recognition with crowd-sourced label distribution. In *Proceedings of the 18th ACM international conference on multimodal interaction*. 279–283.

[3] Michael Biehl, David Matsumoto, Paul Ekman, Valerie Hearn, Karl Heider, Tsutomu Kudoh, and Veronica Ton. 1997. Matsumoto and Ekman's Japanese and

Caucasian Facial Expressions of Emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal behavior* 21, 1 (1997), 3–21.

[4] Robin Bretin, Mohamed Khamis, Emily Cross, and Mohammad Obaid. 2025. The Role of Drone's Digital Facial Emotions and Gaze in Shaping Individuals' Social Proxemics and Interpretation. *J. Hum.-Robot Interact.* 14, 3, Article 48 (May 2025), 34 pages. doi:10.1145/3714477

[5] Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N Chang, Sungbok Lee, and Shrikanth S Narayanan. 2008. IEMOCAP: Interactive emotional dyadic motion capture database. *Language resources and evaluation* 42, 4 (2008), 335–359.

[6] Tuochao Chen, Yaxuan Li, Songyun Tao, Hyunchul Lim, Mose Sakashita, Ruidong Zhang, Francois Guimbretiere, and Cheng Zhang. 2021. Neckface: Continuously tracking full facial expressions on neck-mounted wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–31.

[7] Yunliang Chen and Jungseock Joo. 2021. Understanding and mitigating annotation bias in facial expression recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 14980–14991.

[8] Seokmin Choi, Yang Gao, Yincheng Jin, Se Jun Kim, Jiyang Li, Wenyao Xu, and Zhanpeng Jin. 2022. PPGface: Like what you are watching? Earphones can" feel" your facial expressions. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 6, 2 (2022), 1–32.

[9] M. Kalpana Chowdary, Tu N. Nguyen, and D. Jude Hemanth. 2023. Deep learning-based facial emotion recognition for human–computer interaction applications. *Neural Computing and Applications* 35, 32 (Nov. 2023), 23311–23328. doi:10.1007/s00521-021-06012-8

[10] John Joon Young Chung, Jean Y Song, Sindhu Kutty, Sungsoo Hong, Juho Kim, and Walter S Lasecki. 2019. Efficient elicitation approaches to estimate collective crowd answers. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–25.

[11] Charles Darwin and Sir Francis Darwin. 1872. *The expression of the emotions in man and animals*. Vol. 3. John Murray London.

[12] Iris Dominguez-Catena, Daniel Paternain, and Mikel Galar. 2022. Gender stereotyping impact in facial expression recognition. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 9–22.

[13] Iris Dominguez-Catena, Daniel Paternain, and Mikel Galar. 2024. Metrics for Dataset Demographic Bias: A Case Study on Facial Expression Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 8 (2024), 5209–5226. doi:10.1109/TPAMI.2024.3361979

[14] Guanglong Du, Shuaiying Long, and Hua Yuan. 2020. Non-contact emotion recognition combining heart rate and facial expression for interactive gaming environments. *IEEE Access* 8 (2020), 11896–11906.

[15] Paul Ekman. 1973. Cross-cultural studies of facial expression. *Darwin and facial expression: A century of research in review* 169222, 1 (1973), 45–60.

[16] Paul Ekman and Wallace V Friesen. 1971. Constants across cultures in the face and emotion. *Journal of personality and social psychology* 17, 2 (1971), 124.

[17] Paul Ekman and Wallace V Friesen. 1978. Facial action coding system. *Environmental Psychology & Nonverbal Behavior* (1978).

[18] Hillary Elfenbein and Nalini Ambady. 2003. When Familiarity Breeds Accuracy: Cultural Exposure and Facial Emotion Recognition. *Journal of Personality and Social Psychology* 85 (08 2003), 276–290. doi:10.1037/0022-3514.85.2.276

[19] Hillary Anger Elfenbein and Nalini Ambady. 2002. On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological bulletin* 128, 2 (2002), 203.

[20] Yingruo Fan, Jacqueline CK Lam, and Victor OK Li. 2021. Demographic effects on facial emotion expression: an interdisciplinary investigation of the facial action units of happiness. *Scientific reports* 11, 1 (2021), 5214.

[21] Le Fang, Bo Chai, Yingqing Xu, and Stephen Jia Wang. 2025. KANFeel: A Novel Kolmogorov-Arnold Network-Based Multimodal Emotion Recognition Framework. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–8.

[22] Xia Fang, Disa A Sauter, and Gerben A Van Kleef. 2018. Seeing mixed emotions: The specificity of emotion perception from static and dynamic facial expressions across cultures. *Journal of Cross-Cultural Psychology* 49, 1 (2018), 130–148.

[23] Xia Fang, Gerben A van Kleef, and Disa A Sauter. 2019. Revisiting cultural differences in emotion perception between easterners and westerners: Chinese perceivers are accurate, but see additional non-intended emotions in negative facial expressions. *Journal of Experimental Social Psychology* 82 (2019), 152–159.

[24] Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin* 76, 5 (1971), 378.

[25] Johnny R. J. Fontaine, Klaus R. Scherer, Etienne B. Roesch, and Phoebe C. Ellsworth. 2007. The World of Emotions is not Two-Dimensional. *Psychological Science* 18, 12 (2007), 1050–1057. doi:10.1111/j.1467-9280.2007.02024.x PMID: 18031411.

[26] Surjya Ghosh, Kaustubh Hiware, Niloy Ganguly, Bivas Mitra, and Pradipta De. 2019. Does emotion influence the use of auto-suggest during smartphone typing?. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 144–149.

[27] Ian J Goodfellow, Dumitru Erhan, Pierre Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, et al. 2013. Challenges in representation learning: A report on three machine learning contests. In *International conference on neural information processing*. Springer, 117–124.

[28] Fernando Gordillo, Lilia Mestas, Miguel Ángel Pérez, José M. Arana, and Eduardo Alejandro Escotto. 2018. Role of Surprise in the Discrimination of the Facial Expression of Fear. *The Spanish Journal of Psychology* 21 (2018), E3. doi:10.1017/sjp.2018.5

[29] Pierre Gosselin and Janik Simard. 1999. Children's knowledge of facial expressions of emotions: Distinguishing fear and surprise. *The Journal of genetic psychology* 160, 2 (1999), 181–193.

[30] Shlomo Hareli, Konstantinos Kafetsios, and Ursula Hess. 2015. A cross-cultural study on emotion expression and the learning of social norms. *Frontiers in psychology* 6 (2015), 1501.

[31] Muhammad Anas Hasnul, Nor Azlina Ab Aziz, Salem Alelyani, Mohamed Mohana, and Azlan Abd Aziz. 2021. Electrocardiogram-based emotion recognition systems and their applications in healthcare—a review. *Sensors* 21, 15 (2021), 5015.

[32] Maryam Imani and Gholam Ali Montazer. 2019. A survey of emotion recognition methods with emphasis on E-Learning environments. *Journal of network and computer applications* 147 (2019), 102423.

[33] Rachael E. Jack, Caroline Blais, Christoph Scheepers, Philippe G. Schyns, and Roberto Caldara. 2009. Cultural Confusions Show that Facial Expressions Are Not Universal. *Current Biology* 19, 18 (Sept. 2009), 1543–1548. doi:10.1016/j.cub.2009.07.051 Publisher: Elsevier.

[34] Rachael E. Jack, Roberto Caldara, and Philippe G. Schyns. 2011. Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *Journal of Experimental Psychology: General* 140, 3 (2011), 575–581. doi:10.1037/a0023463

[35] Rachael E Jack, Oliver GB Garrod, Hui Yu, Roberto Caldara, and Philippe G Schyns. 2012. Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences* 109, 19 (2012), 7241–7244.

[36] Joshua Conrad Jackson, Joseph Watts, Teague R Henry, Johann-Mattis List, Robert Forkel, Peter J Mucha, Simon J Greenhill, Russell D Gray, and Kristen A Lindquist. 2019. Emotion semantics show both cultural variation and universal structure. *Science* 366, 6472 (2019), 1517–1522.

[37] Takeo Kanade, Jeffrey F Cohn, and Yingli Tian. 2000. Comprehensive database for facial expression analysis. In *Proceedings fourth IEEE international conference on automatic face and gesture recognition (cat. No. PR00580)*. IEEE, 46–53.

[38] Hackjin Kim, Leah H Somerville, Tom Johnstone, Sara Polis, Andrew L Alexander, Lisa M Shin, and Paul J Whalen. 2004. Contextual modulation of amygdala responsivity to surprised faces. *Journal of cognitive neuroscience* 16, 10 (2004), 1730–1745.

[39] Sun-Min Kim, Ye-Jin Kwon, Soo-Yun Jung, Min-Ji Kim, Yang Seok Cho, Hyun Taek Kim, Ki-Chun Nam, Hackjin Kim, Kee-Hong Choi, and June-Seek Choi. 2017. Development of the Korean facial emotion stimuli: Korea university facial expression collection 2nd edition. *Frontiers in Psychology* 8 (2017), 769.

[40] Abhiram Kolli, Alireza Fasih, Fadi Al Machot, and Kyandoghere Kyamakya. 2011. Non-intrusive car driver's emotion recognition using thermal camera. In *Proceedings of the Joint INDS'11 & ISTET'11*. IEEE, 1–5.

[41] Ye-jin Lee and Yoshiyuki Matsumoto. 2011. Emotional display rules of Japanese and Koreans. *The Japanese Journal of Psychology* 82, 5 (2011), 415–423. doi:10.4992/jjpsy.82.415

[42] Ke Li, Ruidong Zhang, Siyuan Chen, Boao Chen, Mose Sakashita, François Guimbretière, and Cheng Zhang. 2024. EyeEcho: continuous and low-power facial expression tracking on glasses. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–24.

[43] Evdoxia Eirini Lithoxoidou, George Eleftherakis, Konstantinos Votis, and Tony Prescott. 2025. Advancing affective intelligence in virtual agents using affect control theory. In *Proceedings of the 30th International Conference on Intelligent User Interfaces*. 127–136.

[44] Patrick Lucey, Jeffrey F Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. 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.

[45] Martin Lukac, Gulnaz Zhambulova, Kamila Abdiyeva, and Michael Lewis. 2023. Study on emotion recognition bias in different regional groups. *Scientific Reports* 13, 1 (2023), 8414.

[46] Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[47] David Matsumoto. 1988. Japanese and Caucasian facial expressions of emotion (JACFEE) and neutral faces (JACNeuF). *Intercultual and Emotion Research Laboratory, Department of Psychology* (1988).

[48] David Matsumoto. 1992. American–Japanese Cultural Differences in the Recognition of Universal Facial Expressions. *Journal of Cross-Cultural Psychology* 23, 1 (1992), 72–84. doi:10.1177/0022022192231005

[49] David Matsumoto and Paul Ekman. 1989. American-Japanese cultural differences in intensity ratings of facial expressions of emotion. *Motivation and emotion* 13, 2 (1989), 143–157.

[50] Dhwani Mehta, Mohammad Faridul Haque Siddiqui, and Ahmad Y. Javaid. 2018. Facial Emotion Recognition: A Survey and Real-World User Experiences in Mixed Reality. *Sensors* 18, 2 (2018). doi:10.3390/s18020416

[51] Barbara Mellers, Katrina Fincher, Caitlin Drummond, and Michelle Bigony. 2013. Surprise: A belief or an emotion? *Progress in brain research* 202 (2013), 3–19.

[52] Batja Mesquita and Robert Walker. 2003. Cultural differences in emotions: A context for interpreting emotional experiences. *Behaviour research and therapy* 41, 7 (2003), 777–793.

[53] Sindhu Nair Mohan, Firdaus Mukhtar, and Laura Jobson. 2021. An exploratory study on cross-cultural differences in facial emotion recognition between adults from Malaysia and Australia. *Frontiers in psychiatry* 12 (2021), 622077.

[54] Ali Mollahosseini, Behzad Hasani, and Mohammad H Mahoor. 2017. Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing* 10, 1 (2017), 18–31.

[55] Masahito Nagata and Katsunori Okajima. 2024. Effect of observer's cultural background and masking condition of target face on facial expression recognition for machine-learning dataset. *PLOS ONE* 19, 10 (Oct. 2024), 1–21. doi:10.1371/journal.pone.0313029 Publisher: Public Library of Science.

[56] Sanjeev Nahulanthran, Leimin Tian, Dana Kulić, and Mor Vered. 2025. Explaining Facial Expression Recognition. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems*. 1511–1519.

[57] Subigya Nepal, Arvind Pillai, Weichen Wang, Tess Griffin, Amanda C Collins, Michael Heinz, Damien Lekkas, Shayan Mirjafari, Matthew Nemesure, George Price, et al. 2024. Moodcapture: Depression detection using in-the-wild smartphone images. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–18.

[58] Maital Neta, F Caroline Davis, and Paul J Whalen. 2011. Valence resolution of ambiguous facial expressions using an emotional oddball task. *Emotion* 11, 6 (2011), 1425.

[59] Maital Neta and M Justin Kim. 2023. Surprise as an emotion: A response to Ortony. *Perspectives on Psychological Science* 18, 4 (2023), 854–862.

[60] Marret K. Noordewier and Seger M. Breugelmans. 2013. On the valence of surprise. *Cognition & Emotion* 27 (2013), 1326 – 1334. https://api.semanticscholar.org/CorpusID:27500262

[61] Marret K Noordewier and Seger M Breugelmans. 2013. On the valence of surprise. *Cognition & emotion* 27, 7 (2013), 1326–1334.

[62] Brian Parkinson. 2023. Cross-cultural calibration of words and emotions: Referential, constructionist, and pragmatic perspectives. *Emotion Review* 15, 4 (2023), 348–362.

[63] Catherine Prado, David Mellor, Linda K. Byrne, Christopher Wilson, Xiaoyan Xu, and Hong Liu. 2014. Facial emotion recognition: a cross-cultural comparison of Chinese, Chinese living in Australia, and Anglo-Australians. *Motivation and Emotion* 38, 3 (June 2014), 420–428. doi:10.1007/s11031-013-9383-0

[64] Prolific. 2025. Prolific | Quickly find research participants you can trust. Retrieved Oct 1, 2025 from https://www.prolific.com

[65] I Michael Revina and WR Sam Emmanuel. 2021. A survey on human face expression recognition techniques. *Journal of King Saud University-Computer and Information Sciences* 33, 6 (2021), 619–628.

[66] Annie Roy-Charland, Melanie Perron, Olivia Beaudry, and Kaylee Eady. 2014. Confusion of fear and surprise: A test of the perceptual-attentional limitation hypothesis with eye movement monitoring. *Cognition and Emotion* 28, 7 (2014), 1214–1222. arXiv:https://doi.org/10.1080/02699931.2013.878687 doi:10.1080/02699931.2013.878687 PMID: 24460373.

[67] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

[68] Muhammad Sajjad, Mansoor Nasir, Fath U Min Ullah, Khan Muhammad, Arun Kumar Sangaiah, and Sung Wook Baik. 2019. Raspberry Pi assisted facial expression recognition framework for smart security in law-enforcement services. *Information Sciences* 479 (2019), 416–431.

[69] Jean Y Song, Raymond Fok, Juho Kim, and Walter S Lasecki. 2019. FourEyes: Leveraging Tool Diversity as a Means to Improve Aggregate Accuracy in Crowdsourcing. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 10, 1 (2019), 1–30.

[70] Jean Y Song, Raymond Fok, Alan Lundgard, Fan Yang, Juho Kim, and Walter S Lasecki. 2018. Two tools are better than one: Tool diversity as a means of improving aggregate crowd performance. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces*. 559–570.

[71] Seyma Takir, Elif Toprak, Pinar Uluer, Duygun Erol Barkana, and Hatice Kose. 2023. Exploring the Potential of Multimodal Emotion Recognition for Hearing-Impaired Children Using Physiological Signals and Facial Expressions. In *Companion Publication of the 25th International Conference on Multimodal Interaction* (Paris, France) *(ICMI '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 398–405. doi:10.1145/3610661.3616240

[72] Nicola Webb, Ariel Ruiz-Garcia, Mark Elshaw, and Vasile Palade. 2020. Emotion recognition from face images in an unconstrained environment for usage on social robots. In *2020 international joint conference on neural networks (IJCNN)*. IEEE, 1–8.

[73] Chiyu Wei, Sean Noh, and Ho-Chun Herbert Chang. 2025. Faces Speak Louder Than Words: Emotions Versus Textual Sentiment in the 2024 USA Presidential Election. In *Companion Proceedings of the ACM on Web Conference 2025*. 1390–1393.

[74] Xiaoqian Yan, Timothy J Andrews, and Andrew W Young. 2016. Cultural similarities and differences in perceiving and recognizing facial expressions of basic emotions. *Journal of Experimental Psychology: Human Perception and Performance* 42, 3 (2016), 423.

[75] Zihan Yan, Yufei Wu, Yang Zhang, and Xiang'Anthony' Chen. 2022. EmoGlass: An end-to-end AI-enabled wearable platform for enhancing self-awareness of emotional health. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.

[76] Tongyun Yang, Bishwas Regmi, Lingyu Du, Andreas Bulling, Xucong Zhang, and Guohao Lan. 2025. Through the Eyes of Emotion: A Multi-faceted Eye Tracking Dataset for Emotion Recognition in Virtual Reality. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 3 (2025), 1–41.

[77] Yeonsun Yang, Ahyeon Shin, Nayoung Kim, Huidam Woo, John Joon Young Chung, and Jean Y Song. 2024. Find the Bot!: Gamifying Facial Emotion Recognition for Both Human Training and Machine Learning Data Collection. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–20.

[78] Zhennan Yi, Xianghua Ding, Chenyang You, and Xingchen Zhou. 2025. Smiles, Frowns, and Everything In Between: Understanding User Experiences with Emotion Tracking in Daily Life through Facial Expression Recognition. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference*. 2226–2236.

[79] Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. Openface 2.0: facial behavior analysis toolkit tadas baltrušaitis. In *IEEE International Conference on Automatic Face and Gesture Recognition*.

[80] Jiabei Zeng, Shiguang Shan, and Xilin Chen. 2018. Facial expression recognition with inconsistently annotated datasets. In *Proceedings of the European conference on computer vision (ECCV)*. 222–237.

[81] Junyao Zhang, Wataru Sato, Naoya Kawamura, Koh Shimokawa, Budu Tang, and Yuichi Nakamura. 2024. Sensing emotional valence and arousal dynamics through automated facial action unit analysis. *Scientific Reports* 14, 1 (Aug. 2024), 19563. doi:10.1038/s41598-024-70563-8