Received 22 August 2023, accepted 9 November 2023, date of publication 28 November 2023, date of current version 1 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3334270

RESEARCH ARTICLE

Effects of Facial Expressions and Gestures on the Trustworthiness of a Person

SYAHEERAH LEBAI LUTFI[1], BADR LAHASAN[2], CRISTINA LUNA-JIMÉNEZ[3], ZAHER A. BAMASOOD[1,4,5], AND ZAHID AKHTAR[6]

[1]School of Computer Sciences, Universiti Sains Malaysia, Minden, Penang 11800, Malaysia
[2]Faculty of Computer and Information Technology, University of Shabwah, Shabwah, Yemen
[3]Grupo de Tecnología del Habla y Aprendizaje Automático (THAU Group), Information Processing and Telecommunications Center, E.T.S.I. de Telecomunicación, Universidad Politécnica de Madrid, 28040 Madrid, Spain
[4]Faculty of Computers and Information Technology, Hadhramout University, Hadhramout, Yemen
[5]Hadramout Establishment for Human Development, Mukalla, Hadramout, Yemen
[6]Department of Network and Computer Security, State University of New York Polytechnic Institute, Utica, NY 13502, USA

Corresponding authors: Syaheerah Lebai Lutfi (Syaheerah@usm.my) and Badr Lahasan (albdr2002@gmail.com)

This work was supported in part by University Sains Malaysia under Grant 304/PKOMP/6315137, in part by the Spanish Ministry of Science and Innovation funded by MCIN/AEI/10.13039/501100011033 through the Project GOMINOLA under Grant PID2020-118112RB-C21 and Grant PID2020-118112RB-C22 and the Project AMIC-PoC under Grant PDC2021-120846-C42, in part by the European Union "NextGenerationEU/PRTR," in part by the 2022 Ayudas Program Propio Universidad Politécnica de Madrid (UPM), and in part by the Erasmus+ Program under Grant 2020-1-ES01-KA107-080401.

ABSTRACT Trust is a fundamental element in human relationships, playing a crucial role in decision-making processes. Despite its significance, numerous dimensions of perceived trustworthiness remain unexplored and warrant further investigation. Previous literature has highlighted the influence of emotions and sentiments on how individuals perceive trustworthiness, with visual, vocal, and behavioral cues serving as essential markers. This preliminary study aims to expand the existing knowledge in the field by investigating trustworthiness traits manifested through facial and gesture expressions in emotional videos across diverse cultural contexts. To address this objective, an annotation platform was developed to collect annotation data using the benchmarked One-Minute-Gradual Emotion (OMG) audiovisual corpus, enabling the annotation of actors' perceived trustworthiness levels alongside other inquiries related to emotional state, gesture, activeness, comfort, and speech integrity. The findings of this study demonstrate a positive correlation between higher levels of speaker activity, faster gesturing, and a relaxed demeanor with increased levels of trust gained from audiences. The proposal presented in this paper holds potential for future studies focused on trustworthiness annotation, facilitating the measurement of trust-related features. Moreover, this research serves as a critical step towards understanding the foundations of trustworthiness in the development of synthetic agents that require perceived trustworthiness, particularly in domains involving negotiations or emergency situations where rapid data collection plays a pivotal role in saving lives.

INDEX TERMS Affective computing, man–machine interaction (MMI), face expression, gesture, emotion, trust modeling.

I. INTRODUCTION

Affective computing is a discipline that encompasses various aspects of understanding human psychological states and integrating this knowledge into intelligent devices to enhance human-machine interaction. Emotion interpretation has been a key focus within this field due to its numerous applications and the development of psychological theories. One prominent researcher in the area of emotion recognition is Paul Ekman, who proposed a division of emotions into six universal families: Anger, Disgust, Fear, Joy, Sadness, and Surprise [33]. These families can further be subdivided to capture specific variations in intensity and valence, although most research primarily considers the first level of these

The associate editor coordinating the review of this manuscript and approving it for publication was Paolo Crippa.

six classes for simplicity. Ekman's theory suggests that these basic emotions can be universally recognized across different cultures and countries. While there is some debate regarding the universality of these emotions, Izard [46] argued that individuals can experience more than one emotion simultaneously. Despite the controversy, Ekman's theory has been widely utilized to label many of the existing emotional datasets in the research community such as RAVDESS [57], IEMOCAP [14], AMIGOS [70]. These datasets serve as valuable resources for training and evaluating emotion recognition systems. By applying Ekman's framework, researchers aim to develop algorithms and models that can accurately categorize and understand human emotions. The availability of these labeled datasets has significantly contributed to advancements in the field of emotion recognition and its applications in affective computing.

When considering how to incorporate intelligence into devices, researchers have always looked to nature for inspiration, in this case, human interactions. These interactions require that interlocutors perceive the emotions being displayed, adapt to the detected mood, and respond appropriately and accordingly. For this reason, several studies on affective computing [3], [23], [85], [101] have attempted to pursue this sense-adapt-react mechanism to create more natural and effective agents that are aware of the complex cognitive states of humans.

The first step in the sense-adapt-react strategy is sensing. Machine sensing is often solved by using external sensors that capture environmental information. These sensors generate signals that are subsequently processed [65]. More specifically, these signals are extracted from two main types of body measurements: physiological and behavioral [86]. This study focuses on the behavioral sensing.

The ability to "sense" the intentions of the person with whom we interact is based not only on the content of the message, but also on how it is expressed. The paralinguistics of speech, facial expressions or gestures are modalities that fall within the scope of nonverbal studies [50]. Among nonverbal modalities, facial expression recognition on images is the most commonly used approach for man-machine interaction in emotion recognition tasks [62].

Body movements, which include postures and gestures, also play an excellent role in conveying people's emotional feelings. Body postures are considered static compared to gestures, which involve moving different body parts (e.g., arms, head, fingers, legs, and hands), but both contain relevant emotional information. For example, in [71] when an angry person expresses their anger, they tend to adopt a dominant body posture to infer their emotions. In contrast, the body of an anxious person appears weak and passive and shows avoidance and reclining tendencies. In terms of gestures, researchers have recently focused on investigating which elements have a greater impact on the transmission of emotions. For example, in [82], they found that only features of the arms and upper body can be good indicators of emotion.

In addition to body movements, some models use paralinguistic information about speech to detect emotions. In the study of Pepino and colleagues [80], eGeMaps features (containing spectral and temporal information of the speech signal) were combined with spectrograms of the recordings to feed a wav2vec transform model.

Finally, other studies in emotion recognition have integrated features from different modalities such as facial expressions, body postures, voice features, or text into video sequences to improve recognition rates ([25], [62]).

Although there is a plethora of studies on emotion modeling [3], [4], [11], [15], [63], [64], [66], [68], [85], [96], mostly are concentrated on the full-blown emotions, and there are few reported studies on related, subtle affective features such as trustworthiness recognition, despite their importance. Trust is essential for people and societies. Trust in friendship, love, communities, and organizations is indispensable for developing strong relationships with our colleagues, partners, and neighbors, which fosters a more connected and powerful country.

In addition, trustworthiness plays a key role in economic exchange and politics. For example, market transactions could be harmed if there is a lack of trust between trading partners, as agreements require a certain level of trust on the other side [52]. From a political perspective, decreasing trust in institutions or in a nation's leaders can lead to the collapse of political legitimacy. For this reason, and due to its multiple applications, it is crucial to measure the potential perceived trustworthiness of an individual.

In the context of emotion recognition and trustworthiness, several studies have found that displayed emotions correlate with the decision to trust a person they have just met [38]. For example, facial expressions that express pleasure (such as smiling) are perceived as more trustworthy than the absence of such a signal [55]. Interestingly, a recent perceptual study [60], [61] on cross-cultural trustworthiness between Malaysians and Hungarians found that facial expressions mattered in deciding to trust a person for the former but not for the latter.

However, most of these studies have not analyzed the relationships between emotions and trust [6], [55], [58], [74] using a benchmarked emotionally-inherent video dataset. They also ignored the characteristics of the raters' such as ethnicity and gender. In the attempt to address these gaps, this study uses the benchmarked emotion-annotated OMG dataset [5], to investigate the perceived trustworthiness elicited by gestures and facial expressions of the subjects in the videos among a multicultural audience.

The main goal of the current study is to investigate the contribution of facial expressions and gestures to the level of perceived trustworthiness by human raters.

Two of the main interests of the study are its cross-cultural nature (as it was annotated by raters from different ethnicities, which are compared in the study) and the use of videos as stimuli, as many previous works do not use a trigger (as in

Berg's lab games) or the stimuli are based on pictures. These two features allow us to examine whether trustworthiness is perceived differently across the cultures studied. In addition, videos could allow for more complex scenarios that are closer to the real world compared to laboratory environments and improve understanding of how perceived trustworthiness is conceptualized from first impressions under everyday conditions.

The current analysis could make a significant contribution to the fields of interpersonal relationships and computer-human interactions by providing the first steps toward developing a synthetic agent prototype capable of mimicking trust characteristics and gaining the trust of human users in various domains to improve their user experience. In general, domains that require negotiation would particularly benefit from such agents, such as, economics, politics, and security.

The remainder of the article is organized as follows. Section II summarizes related work in the field of trustworthiness detection from different points of view. Section III describes the participants who contributed to this study, the annotation procedure, and the dataset used for the annotations. Section IV presents the statistics used in detail and explains the main results obtained from their analysis. Finally, in Section VI, we discuss the key conclusions of this study and future research directions.

II. RELATED WORK

Trust is the foundation for the success of any social, economic or political relationship. Social relationships can take various forms, such as partnerships, friendships, and family ties [26], [90]. Economic relationships involve a connection between organizations and their customers [37], [48], [81]. The eBay market is an example of a triumphant connection that is highly dependent on trust [83]. Finally, the relationship between citizens and a nation's political institutions is an example of a political relationship because without trust, the government would collapse [40], [93].

Although the definition and understanding of trust is still an open research topic, it has been studied in several dimensions to understand what characteristics elicit trustworthy responses from a perceptual perspective. Figure 1 summarizes some of the areas of trustworthiness studied in the literature. In the realm of personality traits, we have selectively focused on observable attributes. Among these attributes, friendliness is intertwined with the observer's perception of the speaker's "comfortability" and "activeness" within the presented stimuli. In the context of socio-economic traits, we deliberately refrained from gathering data related to religion and income. This decision is rooted in the fact that these two aspects do not manifest visually as readily as characteristics like ethnicity, gender, and age. Given our primary interest in delving into facial expressions and gestures, we have chosen to exclude vocal information from our analysis. Notably, the study of attractiveness constitutes a separate ongoing project that leverages the OMG framework-an extension derived from the present work.

One of the most important branches of research is the correlation between emotions and perceived trustworthiness ([6], [31], [51], [58], [74]). In [58], emotions with positive valence (especially joy and gratitude) were found to increase trust, while emotions with negative valence (e.g., anger) decreased trust. Subsequently, [74] confirmed that emotions with negative valence appear to decrease trustworthiness under certain circumstances.

More specifically, they discovered that this was only the case for emotions with low certainty ratings, such as fear, whereas anger and guilt with high levels of certainty had no clear effect on trusting behavior. Moreover, [6], [31], [51] observed a robust relationship between the effects of emotions on trust judgments in zero-acquaintance negotiation situations.

The second line of investigation is the relationship between trustworthiness and physical aspects, such as facial expressions or body gestures. For example, smiling is generally translated into trustworthy social attributions [53], [88], especially when perceived as spontaneous [41]. Other studies agree that high variability in temporal emotions and facial movements are understood as more trustworthy and authentic features [91].

Regarding body gestures, the study by Lee et al. [55] analyzed a set of nonverbal cues to discover movements or expressions that correlate with annotated trust values, in a laboratory setting. They concluded that certain gestures such as smiling, leaning forward, or opening arms convey higher levels of trust. In contrast, gestures such as touching our face, crossing our arms, leaning back or touching our hands frequently are perceived as less trustworthy signals.

Attractiveness can also influence the perception of confidence level [19], [92], [98]. In these studies, results obtained using competition-based laboratory games showed that attractive people are more able to gain the trust of others. In addition to emotions, gestures, and attractiveness, other factors such as demographic or socioeconomic aspects may also influence the strength of trustworthiness.

In [30] the authors presented a taxonomy of the current methods and classifications in emotion recognition, along with their limitations. Schneider et al. [29], found that implementing facial expression and gestures performed by a pedagogical agent resulted in successful learning in a separate study. Nápoles et al. [28] investigated the impact of facial expression and gestures on perceptions of choral conductor and ensemble in their study. Recent research by Abi et al. [27] examined the influence of video background, facial expression and gender on trustworthiness and competence in a videoconferencing context, revealing the significant impact of each factor in the presented results.

Researchers [2], [18] have shown that demographic aspects such as people's gender, race, or age have an impact on how much they trust others. In the study by Li et al. for example, it is confirmed that people trust older people more than faces that look younger [56].

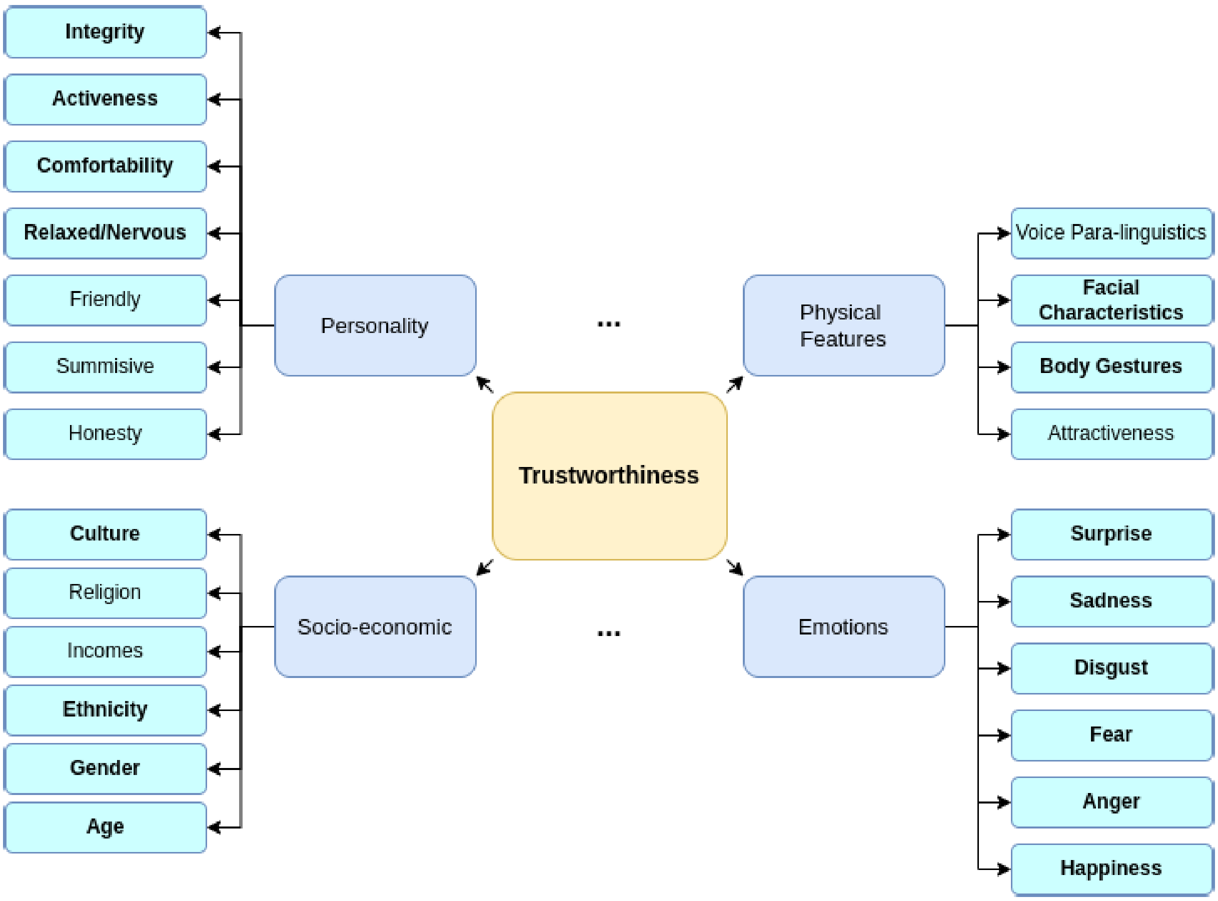


FIGURE 1. **Different dimensions of trustworthiness studies. In bold, the features that were studied in this paper.**

In terms of socioeconomic aspects, some studies have found that countries and regions can also have an impact on trust levels [47], [97], as well as people's reputation to infer the trustworthiness of their current partner [12], [16], [17], [24], [39]. Similarly, expectations also play an important role in trustworthy relationships. The work of [95] showed that different accents were associated with expected behaviors. Thus, when behaviors were congruent, trustworthiness was restored more quickly than when a deceptive action was present. The congruency effect is also noticeable with faces. When pictures show people smiling with happy eyes, raters trust them more than when they smile but the eye muscles assume a position associated with the emotion anger [36].

Perceived personality also plays an important role in trusting a person. Several articles have found a strong relationship between certain universal facial characteristics and social status or traits. For example, adults who have a baby-like face are usually classified as not physically strong, submissive, friendly, naive, warm, and honest [8], [69], [73]). Attractive facial expressions, on their part, are usually associated with high competence and high intelligence level [32]. In addition, angry facial expressions have been observed to signify dominance [45], [49], [72]. Masculine and feminine facial features could also mean that some individuals have dominant personalities [10].

It should be noted that of the previous studies that have examined issues related to facial and body expressions and their possible interpretations, few have used video datasets. For this reason, in this article we address the relationship between trustworthiness and perceived trustworthiness in emotional video datasets at zero acquaintance. Among the several available datasets in the literature with emotional annotations OMG [5], IEMOCAP [14], EmoReact [76], MOSI [100], RAVDESS [57], etc., we chose to use OMG because, this dataset provides a more realistic environment where the actors are of different nationalities, genders, and ages. In addition, the videos were shot in real world conditions where lighting and backgrounds can vary from one video to another. This provides a richer environment to explore what cues influence perceived trustworthiness from new perspectives. Additionally, this dataset is especially interesting for our study due to the different nationalities of the actors, as well as their rich expressive content.

OMG is a multimodal dataset of emotions and contains 173 relatively long videos of monologues, lasting on average about one minute. The dataset contains audiovisual content with their transcriptions. In addition, each video in the OMG dataset was split into individual clips based on utterances and annotated in terms of arousal-valence dimensions following Ekman and Friesen [33].

III. METHODOLOGY: EFFECT OF FACE AND GESTURE EXPRESSIONS ON TRUST

A. PARTICIPANTS

In this research, there were two groups involved: the Yemeni group and the mixed ethnic group. The Yemeni group consisted of 127 students, with 18 females (14,17%) and 109 (85,83%) males, from the University of Science and Technology, Hadhramaut, Yemen (UST). These students shared similar physical characteristics with people from the Middle Eastern or North African ethnic backgrounds. Each student in this group evaluated between 5 and 10 videos, resulting in a total of 1,373 annotations. Figure 2 provides information about the age distribution within the Yemeni ethnic group. Among the Yemeni participants, 15 (11,81%) were under 20 years old, 92 (72,44%) were between 20 to 25 years old, and 7 (5,51%) were between 36 and 45 years old.

As for the mixed ethnic group, there were 17 participants who completed the surveys, including 8 males (47,06%) and 9 females (52,94%). The participants in this group belonged to different ethnicities: 11 Asians, 4 Blacks/African Americans, 1 white, and 1 Hispanic. Each participant in the mixed ethnic group analyzed at least 5 videos, resulting in a total of 238 annotations. Figure 2 also displays the age distribution within the mixed ethnic group. None of the participants in this group were under 20 years old. Among them, 7 (64,70%) were between 20 and 25 years old, 5 (29,41%) were between 26-35 years old, and 3 (17,65%) were between 36 and 45 years old. Additionally, there were 2 participants (11,76%) in the age range of 46-60.

The inclusion criteria encompass three aspects: English language proficiency, as all OMG videos were presented in English, and age within the specified range of 20 to 45 years, indicating a focus on youth participants. Notably, no explicit exclusion criteria were employed, as the fulfillment of the inclusion criteria was incorporated within the Google form (in the instructions, specifically), resulting in the exclusion of individuals who did not meet the specified criteria from participating in the study. For example, participants who had a previous experience rating the OMG videos are not encouraged to participate. Our data shows that none of these participants fulfilled the exclusion criteria.

The participants were instructed to complete a Google form consisting of video clips sourced from OMG, accompanied by a set of itemized questions for response. The questionnaire demonstrated clarity, and the instructions were straightforward, requiring participants to view the video content and provide corresponding answers to the given questions.

Each participant used their own computer or mobile phone for the purpose of accessing and engaging with the videos.

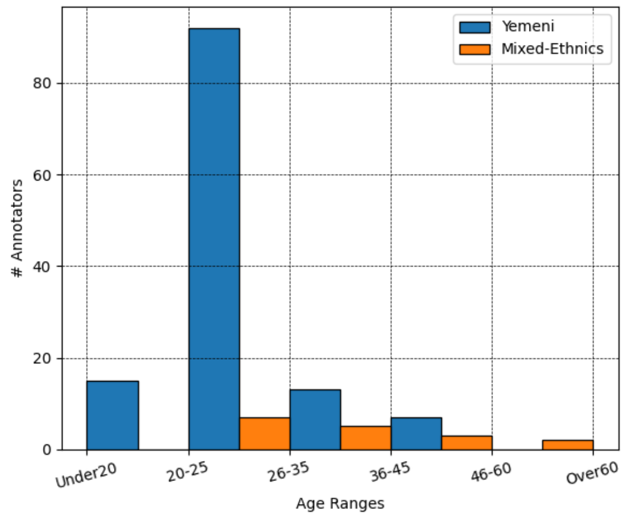


FIGURE 2. Age distribution of annotators per ethnic. In blue, Yemeni annotators (127 in total); in orange, mixed-ethnics annotators (17 in total).

This computer/mobile setup was seamlessly integrated into the Google forms platform, enabling participants to watch and listen to the videos within the context of the study.

The allocation of videos to participant groups was performed through a random assignment process. This way, we control that each group does not repeat the same videos. Each group of participants was exposed to a randomized selection of videos, ensuring that each video was viewed by a minimum of three participants within the group. Likewise, each participant from a given group provided ratings for at least three videos. It is important to note that the distribution of videos among participants was completely randomized, devoid of any predetermined patterns or biases.

B. ANNOTATION PROCEDURE

To annotate the data, we developed two annotation set-ups, one local and one online, in which the participants had to fill in a questionnaire about certain perceived features of the subject in the video.

The 'local set-up' was used with the Yemeni group. All the students watched between 5 to 10 videos locally in different sessions: some using local computers and others watching the videos on a big HD screen. In both cases, participants filled in their answers to a printed questionnaire.

Unlike the local set-up, for the online version, we developed a system using YouTube and Google Forms. First, all the selected videos were uploaded to YouTube, and later, integrated into Google forms on sets of 5 videos. This set-up was employed with the mix-group to obtain answers from different nationalities and compare them to the Yemini group.

Thus, all the questionnaires contained from 5 to 10 videos with their six associated questions about the perceived features of the subjects that appeared on the video. The participants in this study were asked to rate their perceived trustworthiness of the speakers based on facial and gesture expressions by answering six questions relating to the speakers' behavior: (emotional) state, gesture, activeness, comfort, and integrity. The specific queries can be consulted in Table 1.

The purpose of using the six questions individually is described afterward. Firstly, for investigating whether the state of the speaker affects her/his trustworthiness. In other words, to discover whether a relaxed person could be perceived as more trustworthy than a stressed one. Secondly, with the question that refers to the gesture of the speaker we expect to discover if people generally trust more on speakers whose gestures are faster than others showing slow movements. Next, another interesting feature to consider was the relationship between activeness and trust. After that, the purpose of the comfort question is to understand if people would feel more comfortable with whom they trust or not. The final question's target is to show if people who display integrity during their speech, gain more trust from the audience or not.

Regarding the responses, the linguistic answers can represent numerical ranges, although they are displayed as sentences to help annotators to understand the meaning of the ranges. While the linguistic responses to the first three questions vary, the last three queries share the same linguistic answers, although all of them represent similar values. Apart from the questions associated with each video, the forms also collected demographic information of the respondents: their age (20-25, 26-35, 36-45, 46-60, and above 60 years); gender (male or female); and ethnicity/race (White; Latino, Hispanic, Spanish, Middle Eastern or North African; Black or African American; Asian; American Indian or Alaska Native; and Native Hawaiian or other Pacific Islanders as classified in [84]).

Overall, the surveys aimed to uncover the effects of body expressions, represented by gestures, and facial expressions on trustworthiness.

C. STIMULI

The main corpus employed for this analysis was 'OMG-Emotion' (One-Minute Gradual-Emotion Recognition). This dataset contains 173 videos collected from YouTube with people displaying different emotions evolving over time.

Each video displays a single person on the scene under different light conditions, scenarios, and costumes simulating 'in the wild' environments. Characters appear in distinct camera shots and situations, but in most cases, they show the whole body. These features allow annotators annotate aspect such as activeness and gestures and compare them with trustworthiness annotations, which is one of the aims of this study. Additionally, this corpus is an open-source corpus that was developed and employed in the IJCNN challenge [5] and used in previous studies for emotion recognition [102], [103]. From the 173 videos, we selected 64 videos that displayed contrasts in gestures and facial expressions to study their effects on perceived trustworthiness.

TABLE 1. Questions asked in each survey and their available answers.

Labels	Questions	Responses
State	In which state do you think the speaker is in most of the time?	Very Distress Distress In between Relaxed Very relaxed
Gesture	What do you think of the gesture of the speaker?	Not clear Very Slow Slow Normal Fast Very Fast
Activeness	Do you think the speaker is an active person?	Not at all Not really Yes, somewhat Yes
Comfort	Do you think the speaker is someone you would feel comfortable with?	Not at all Not really I don't know Yes, sometimes Yes, most of the time
Trustworthiness	Do you think the speaker is a trustworthy person?	Not at all Not really I don't know Yes, sometimes Yes, most of the time
Integrity	Do you think the speaker displays integrity in her/his speech?	Not at all Not really I don't know Yes, sometimes Yes, most of the time

The nature of the speakers affected the grouping of the OMG video clips. These 64 speakers were selected randomly from the OMG dataset. As sometimes there are several videos of the same person, the speakers' videos of each actor were extracted and joined into one clip with a recorded variety of emotions and were usually not longer than 46 seconds. However, when the same emotion was expressed by a unique speaker in more than one clip, only one of the clips was randomly chosen. In other words, for one speaker there was just one video in which several emotions (ranging from happiness, sadness, fear, disgust, anger, surprise, or neutral) are represented. The respondents were then requested to evaluate the subjects in each video by providing answers to the six posed questions.

Overall, the number of questions for all 64 videos was 384, with six questions per recording. Table 1 shows the questions and the available participants' responses.

IV. RESULTS AND DISCUSSION

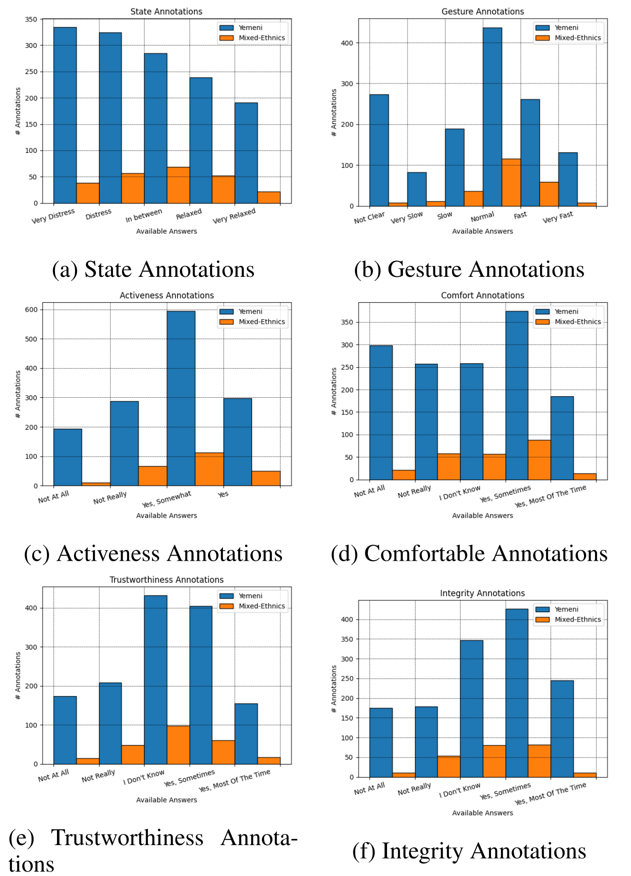
In this research, version 26.0 of the IBM SPSS statistics for windows was used to analyze the responses retrieved from the questionnaires. The major purpose of this research is to analyze the assessment of people's perceived trust based on the facial expressions and gestures performed by actors in videos during their speeches. This study was applied to 2 groups, namely Yemeni and mixed ethnic, to interpret possible variations across cultures. Since we have multiple rates who independently evaluate the same set of responses, we calculated Cohen's kappa coefficient to determine the degree of agreement beyond what would be

expected by chance. This coefficient ranges from -1 to 1, with higher values indicating stronger agreement. The ranges of agreements are as follows:
Less than 0: Indicates no agreement beyond chance.
0.01 to 0.20: Slight agreement.
0.21 to 0.40: Fair agreement.
0.41 to 0.60: Moderate agreement.
0.61 to 0.80: Substantial agreement.
0.81 to 1: Almost perfect agreement.
We included data that achieved 0.05 and above.

A. NORMALITY TEST

To understand the distribution of collected annotations by ethnicity and decide whether to use parametric or nonparametric statistics, we performed a Shapiro-Wilk test for normality by trait (state, gesture, activity, comfort, trustworthiness, and integrity) and ethnicity (Yemeni and Mix). The Shapiro-Wilk tests yielded significance values of less than 0.05 ($p < 0.05$), indicating that the data distribution was significantly different from a normal distribution in all cases. Figure 3 shows the histograms generated by the collected annotations and the results of the Shapiro-Wilk Test.

These results encourage the use of non-parametric statistics for comparing features and annotations across cultures. For this reason, we have used metrics of this type in the following subsections to examine the significance of respondent-reported results and to examine correlations between the features and constructs in the questionnaire.



(a) State Annotations
(b) Gesture Annotations
(c) Activeness Annotations
(d) Comfortable Annotations
(e) Trustworthiness Annotations
(f) Integrity Annotations

FIGURE 3. Normality study of annotations collected in the surveys. In blue, Yemeni annotations; in orange, mixed-ethnics annotations.

B. ETHNIC INFLUENCE ON ANSWERS

Statistics were used to test whether ethnicity introduced divergences across cultures in the perceived trustworthiness. For these experiments, we compared the distribution of labeled features (state, gesture, activeness, comfort, trust, and integrity) between the Yemeni and mixed-ethnic groups by employing Mann-Whitney tests.

The Mann-Whitney test is a nonparametric measurement equivalent to the two-sample t-test, but without the assumptions of normality of distributions and equal variance, which are not met for the sampled responses. The only requirement for using the Mann-Whitney method is independence of the samples, which is met because the selected commentators had different nationalities or ethnicities. We used the Mann-Whitney metric to test the null hypothesis (H_0): Each annotated feature (state, gesture, activity, comfort, trustworthiness, and integrity) had the same distribution for the two ethnicities (Yemeni and Mix).

Table 2, shows the U-Mann-Whitney test performed for each feature and the significance result obtained using the SPSS software. In all cases, the significance indicated was the 'exact significance (bilateral)'.

Analyzing the first feature, the distribution of the 'State' characteristic given by Yemeni ethnic respondents was not significantly different from that given by mixed-ethnics (U-Mann-Whitney = 153566.5, p = 0.13). So experiments reveal not finding a significant difference between ethnics annotations for the 'state' feature, so the H_0 hypothesis is maintained: 'The two groups have the same distribution'. Similar conclusions can be extracted for 'Activeness', 'Comfort', 'Trustworthiness' and 'Integrity'.

The unique divergence in the results can be observed in the 'Gestures' query, in which the significance value is lower than 0.05 ($p = 0.001 < 0.05$). Therefore, in this case, this statistic suggests that when evaluating the number of movements of a person in a video, different ethnicities could not have the same perception. Another hypothesis could be that the 'Not clear or Null' option is introducing certain biases in annotations, not being understood in the same way by different cultures or creating confusion about its meaning.

The reader can find more details about the range comparisons performed by the Mann-Whitney test per culture in Table 3.

C. FEATURES CORRELATION

Nonparametric Spearman correlations were extracted to examine the relationship between each queried characteristic, focusing on the impact that activity level might have on perceptions of trustworthiness. For this analysis, we combined responses from all the annotators ($N = 1660$) as there were

TABLE 2. Mann-Whitney comparative results of ethnicity per question. Significance level of 0.05.

Feature	U Mann-Whitney	Significance (Bilateral)
State	153566,5	0.13
Gesture	141993,5	0.001*
Activeness	155193	0,192
Comfort	153949	0,144
Trustworthiness	157165,5	0,331
Integrity	155193	0,192

TABLE 3. Details of Mann-Whitney ranges comparison by ethnic per question.

Feature	Ethnicity	N	Average Range	Sum of Ranges
State	Yemeni	1373	813,15	1116458,5
	Mix	238	764,74	182007,5
Gesture	Yemeni	1373	790,42	1085244,5
	Mix	238	895,89	213221,5
Activeness	Yemeni	1373	800,03	1098444
	Mix	238	840,43	200022
Comfort	Yemeni	1373	799,13	1097200
	Mix	238	845,66	201266
Trustworthiness	Yemeni	1373	810,53	1112859,5
	Mix	238	779,86	185606,5
Integrity	Yemeni	1373	800,03	1098444
	Mix	238	840,43	200022

no significant differences between the cultures studied for most features.

Focusing on the first row of the correlation matrix displayed in Figure 4, it shows the significant correlation that there exist between 'Trustworthiness' and certain characteristics, which are: 'State' ($\rho = -0.25$), 'Activeness' ($\rho = 0.26$), 'Integrity'($\rho = 0.49$) and 'Comfortableness' ($\rho = 0.56$). These results suggest that active people seem to be more trustworthy than non-active people when their movement is not associated with a negative state such as stress or anxiety. Furthermore, these metrics conjecture that trustworthiness increases among people perceived as honest and friendly. In the same way, subjects annotated as trustworthy receive also high scores for the 'Comfortableness' inquiry, so it seems that trustworthiness is necessary to feel comfortable with a person's company.

Another correlation of note appears between the speech integrity of the person displayed in the video and the perception of being a person that the annotators could feel comfortable with ($\rho = 0.37$). This high correlation may indicate that people with high integrity levels in their speech are seen with other good moral or communication skills (e.g. honesty, friendliness ...) that make the annotator believe they would feel comfortable in their company. As in the previous analysis of trustworthiness, the 'State' also shows a negative correlation with 'Comfortableness', implying that more nervous or distressed people are perceived as less pleasant to be with, enjoying more the company of a more relaxed or calmed person at zero acquaintance.

Particularly relevant is the correlation ($\rho = -0.43$) that exists between the 'state' of the communicator and the expected level of 'comfortability' perceived by the receiver (in our case, the annotators). The correlation results

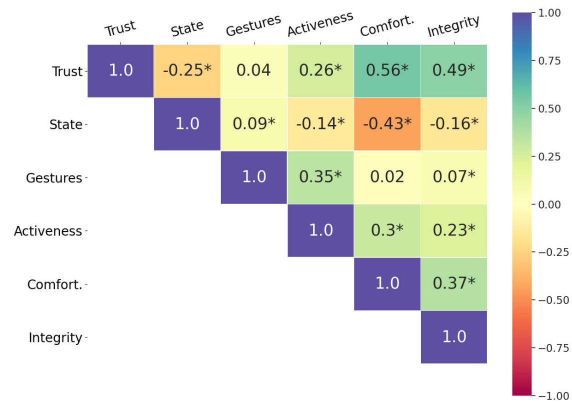


FIGURE 4. Spearman's correlation between features. (*) indicates statistical significance ($p < 0.01$).

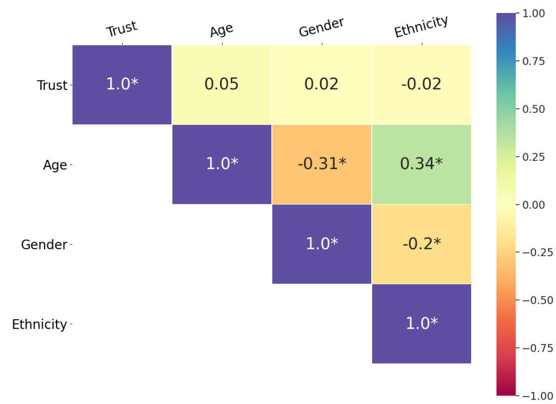


FIGURE 5. Spearman's correlation between trustworthiness and annotators demographics. (*) Indicates statistical significance ($p < 0.01$).

suggest that individuals are uncomfortable in the company of worrying individuals, while they enjoy the company of a more relaxed or calm person at zero acquaintance. This last finding should be examined more closely to understand the basis of this rejection, as it could be related to perceptions of psychological problems. In this case, this issue could be particularly relevant to help people with mental illness in their daily communication.

Finally, 'Activeness' and 'Gesture' show a positive correlation ($\rho = 0.35$). This correlation confirms that people making more movements with hands or faces are seen as more active too, which demonstrates that the annotators and surveys are able to capture correlations between related concepts.

D. TRUSTWORTHINESS DEMOGRAPHICS CORRELATION

From the demographic correlation study (see Figure 5), demographic characteristics such as age, ethnicity, or gender do not appear to influence trustworthiness ratings. To confirm these results, future experiments will collect more samples of the underrepresented population, e.g., from the mixed-ethnic group or from women, to confirm that this behavior can be observed in a larger population sample and can be replicated in other scenarios.

V. LIMITATIONS

One of the main limitations of this study is the small number of annotated videos and respondents. Since this is a study conducted among the students of UST, the population sample is mainly concentrated in the 20-25 age group, with a higher proportion of males than females. For this reason, we will expand the study to other cultures and populations in future versions to determine if the findings found in this population sample can be applied across ethnics.

VI. CONCLUSION

Understanding and studying trustworthiness is crucial in various aspects of society, negotiations, and relationships. It is essential to explore its dimensions in different scenarios to gain valuable insights. This preliminary article takes a multicultural perspective and investigates trustworthiness through Yemeni annotations and a group of mixed ethnicities. The study focuses on videos with high emotional content and examines the impact of gestures on perceived trustworthiness when people have no prior acquaintance.

The results, obtained through the Mann-Whitney test and Spearman rank order correlations, reveal significant findings regarding gestures and facial expressions in relation to social traits. The key discoveries from the study are as follows:

- More active people convey higher levels of trustworthiness across cultures ($\rho = 0.26$).
- People who look honest and display integrity in their speeches also received higher trustworthiness scores ($\rho = 0.49$).
- If all the previously mentioned factors increase the trustworthiness of the speakers, this in turn increases the comfort of the listeners.
- Conversely, individuals who are perceived as stressed instead of relaxed are considered less trustworthy ($\rho = -0.25$), and make them feel less comfortable ($\rho = -0.43$).
- There are no discernible differences between the comments collected for Yemeni ethnic groups compared to mixed ethnic groups. Neither demographic characteristics such as age nor gender reached a level of significance suggesting a correlation with perceived trustworthiness.

In summary, people generally trust speakers who display positive emotions, such as activity, less stressed, and integrity in their movements and speech. The more pronounced these features are, the more trustworthy they appear and the more comfortable the audience feels. Moving forward, our research endeavors will be extended to include a broader scope, encompassing a larger corpus of videos and an increased number of annotators hailing from diverse cultural backgrounds. This expansion aims to enhance the reliability and generalizability of our findings. To ensure consistency, we will also normalize the potential responses to the queries posed during the annotation process.

Furthermore, our future investigations will delve deeper into the understanding of the rejection of distress traits. This aspect will be explored with meticulous attention, shedding light on the dynamics of trust perception in relation to emotional states.

As our annotated dataset continues to grow, we will leverage advanced techniques such as automated deep learning systems. These systems will be designed to evaluate trustworthiness, taking into account a wide array of factors including facial expressions, gestures, ethnicity, familiarity, and potentially more. By employing such comprehensive approaches, we aim to develop robust models that can effectively assess trustworthiness across multiple dimensions.

Through these future advancements, our research aims to contribute to a more nuanced understanding of trust perception and its intricate interplay with various factors. The integration of advanced technologies will enable us to uncover valuable insights and pave the way for practical applications in domains where trust is of paramount importance.

REFERENCES

[1] S. Agrawal, S. Dutta, and B. K. Patra, "Sentiment analysis of short informal text by tuning BERT-Bi-LSTM model," in *Proc. IEEE 19th Int. Conf. Smart Technol. (EUROCON)*, Jul. 2021, pp. 98–102, doi: 10.1109/EUROCON52738.2021.9535535.

[2] A. Alesina and E. La Ferrara, "Who trusts others?" *J. Public Econ.*, vol. 85, no. 2, pp. 207–234, 2002, doi: 10.1016/S0047-2727(01)00084-6.

[3] A. Azazi, S. L. Lutfi, and I. Venkat, "Identifying universal facial emotion markers for automatic 3D facial expression recognition," in *Proc. Int. Conf. Comput. Inf. Sci. (ICCOINS)*, Jun. 2014, pp. 1–6.

[4] A. Azazi, S. L. Lutfi, I. Venkat, and F. Fernández-Martínez, "Towards a robust affect recognition: Automatic facial expression recognition in 3D faces," *Expert Syst. Appl.*, vol. 42, no. 6, pp. 3056–3066, Apr. 2015.

[5] P. Barros, N. Churamani, E. Lakomkin, H. Siqueira, A. Sutherland, and S. Wermter, "The OMG-emotion behavior dataset," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–7, doi: 10.1109/IJCNN.2018.8489099.

[6] L. Y. Belkin and N. B. Rothman, "Do I trust you? Depends on what you feel: Interpersonal effects of emotions on initial trust at zero-acquaintance," *Negotiation Conflict Manage. Res.*, vol. 10, no. 1, pp. 3–27, Feb. 2017, doi: 10.1111/ncmr.12088.

[7] J. Berg, J. Dickhaut, and K. McCabe, "Trust, reciprocity, and social history," in *Games and Economic Behaviour*. Amsterdam, The Netherlands: Elsevier, 1995, pp. 122–142.

[8] D. S. Berry and L. Z. McArthur, "Perceiving character in faces: The impact of age-related craniofacial changes on social perception," *Psychol. Bull.*, vol. 100, no. 1, pp. 3–18, 1986.

[9] A. Biglan, "Distressed behavior and its context," *Behav. Analyst*, vol. 14, no. 2, pp. 157–169, Oct. 1991, doi: 10.1007/bf03392566.

[10] L. G. Boothroyd, B. C. Jones, D. M. Burt, and D. I. Perrett, "Partner characteristics associated with masculinity, health and maturity in male faces," *Personality Individual Differences*, vol. 43, no. 5, pp. 1161–1173, Oct. 2007.

[11] C. Bourke, K. Douglas, and R. Porter, "Processing of facial emotion expression in major depression: A review," *Austral. New Zealand J. Psychiatry*, vol. 44, no. 8, pp. 681–696, Aug. 2010.

[12] J. Bracht and N. Feltovich, "Whatever you say, your reputation precedes you: Observation and cheap talk in the trust game," *J. Public Econ.*, vol. 93, nos. 9–10, pp. 1036–1044, Oct. 2009, doi: 10.1016/j.jpubeco.2009.06.004.

[13] M. Brülhart and J.-C. Usunier, "Does the trust game measure trust?" *Econ. Lett.*, vol. 115, no. 1, pp. 20–23, Apr. 2012.

[14] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan, "IEMOCAP: Interactive emotional dyadic motion capture database," *Lang. Resour. Eval.*, vol. 42, no. 4, pp. 335–359, Dec. 2008, doi: 10.1007/s10579-008-9076-6.

[15] L. D. Cameron and N. C. Overall, "Suppression and expression as distinct emotion-regulation processes in daily interactions: Longitudinal and meta-analyses," *Emotion*, vol. 18, no. 4, p. 465, 2018.

[16] L. J. Chang, B. B. Doll, M. van 't Wout, M. J. Frank, and A. G. Sanfey, "Seeing is believing: Trustworthiness as a dynamic belief," *Cognit. Psychol.*, vol. 61, no. 2, pp. 87–105, Sep. 2010, doi: 10.1016/j.cogpsych.2010.03.001.

[17] G. Charness, N. Du, and C.-L. Yang, "Trust and trustworthiness reputations in an investment game," *Games Econ. Behav.*, vol. 72, no. 2, pp. 361–375, Jun. 2011, doi: 10.1016/j.geb.2010.09.002.

[18] A. Chaudhuri and L. Gangadharan, "An experimental analysis of trust and trustworthiness," *Southern Econ. J.*, vol. 73, no. 4, pp. 959–985, Apr. 2007, doi: 10.1002/j.2325-8012.2007.tb00813.x.

[19] J. Chen, J. Zhong, Y. Zhang, P. Li, A. Zhang, Q. Tan, and H. Li, "Electrophysiological correlates of processing facial attractiveness and its influence on cooperative behavior," *Neurosci. Lett.*, vol. 517, no. 2, pp. 65–70, May 2012, doi: 10.1016/j.neulet.2012.02.082.

[20] Z. Chen, D. Huang, Y. Wang, and L. Chen, "Fast and light manifold CNN based 3D facial expression recognition across pose variations," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018, pp. 229–238.

[21] O. Collignon, S. Girard, F. Gosselin, S. Roy, D. Saint-Amour, M. Lassonde, and F. Lepore, "Audio-visual integration of emotion expression," *Brain Res.*, vol. 1242, pp. 126–135, Nov. 2008.

[22] M. Coulson, "Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence," *J. Nonverbal Behav.*, vol. 28, no. 2, pp. 117–139, 2004, doi: 10.1023/b:jonb.0000023655.25550.be.

[23] N. Dael, M. Mortillaro, and K. R. Scherer, "Emotion expression in body action and posture," *Emotion*, vol. 12, no. 5, p. 1085, 2012.

[24] M. R. Delgado, R. H. Frank, and E. A. Phelps, "Perceptions of moral character modulate the neural systems of reward during the trust game," *Nature Neurosci.*, vol. 8, no. 11, pp. 1611–1618, Nov. 2005, doi: 10.1038/nn1575.

[25] J. J. Deng, C. H. C. Leung, and Y. Li, "Multimodal emotion recognition using transfer learning on audio and text data," in *Computational Science and Its Applications* (Lecture Notes in Computer Science), vol. 12951, O. Gervasi et al., Eds. Cham, Switzerland: Springer, 2021, doi: 10.1007/978-3-030-86970-0_39.

[26] M. Deutsch, "Trust and suspicion," *J. Conflict Resolution*, vol. 2, no. 4, pp. 265–279, Dec. 1958, doi: 10.1177/002200275800200401.

[27] A. Cook, M. Thompson, and P. Ross, "Virtual first impressions: Zoom backgrounds affect judgements of trust and competence," *PLoS ONE*, vol. 18, no. 9, 2023, Art. no. e0291444, doi: 10.1371/journal.pone.0291444.

[28] J. Nápoles, B. A. Silvey, and M. Montemayor, "The influences of facial expression and conducting gesture on college musicians' perceptions of choral conductor and ensemble expressivity," *Int. J. Music Educ.*, vol. 39, no. 2, pp. 260–271, May 2021, doi: 10.1177/0255761420926665.

[29] S. Schneider, F. Krieglstein, M. Beege, and G. D. Rey, "The impact of video lecturers' nonverbal communication on learning—An experiment on gestures and facial expressions of pedagogical agents," *Comput. Educ.*, vol. 176, Jan. 2022, Art. no. 104350, doi: 10.1016/j.compedu.2021.104350.

[30] N. Ahmed, Z. Al Aghbari, and S. Girija, "A systematic survey on multimodal emotion recognition using learning algorithms," *Intell. Syst. With Appl.*, vol. 17, Feb. 2023, Art. no. 200171, doi: 10.1016/j.iswa.2022.200171.

[31] D. Dunning, J. E. Anderson, T. Schlsser, D. Ehlebracht, and D. Fetchenhauer, "Trust at zero acquaintance: More a matter of respect than expectation of reward," *J. Personality Social Psychol.*, vol. 107, no. 1, pp. 122–141, 2014.

[32] A. H. Eagly, R. D. Ashmore, M. G. Makhijani, and L. C. Longo, "What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype," *Psychol. Bull.*, vol. 110, no. 1, pp. 109–128, Jul. 1991.

[33] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *J. Personality Social Psychol.*, vol. 17, no. 2, pp. 124–129, 1971, doi: 10.1037/h0030377.

[34] P. Ekman, "An argument for basic emotions," *Cognit. Emotion*, vol. 6, nos. 3–4, pp. 169–200, 1992, doi: 10.1080/02699939208411068.

[35] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," *Pattern Recognit.*, vol. 36, no. 1, pp. 259–275, Jan. 2016, doi: 10.1016/S0031-3203(02)00052-3.

[36] A. F. Martín, Á. P. Plaza, C. L. Concepción, and M. G. Calvo, "Trustworthiness of a smile as a function of changes in the eye expression,' *Psicothema*, vol. 29, no. 4, pp. 462–468, 2017, doi: 10.7334/psicothema2016.306.

[37] D. Fetchenhauer and G. van der Vegt, "Honesty, trust and economic growth," *Zeitschrift für Sozialpsychologie*, vol. 32, no. 3, pp. 189–200, Sep. 2001.

[38] J. P. Forgas, "Affective influences on attitudes and judgments," in *Handbook of Affective Sciences* (Series in Affective Science), R. J. Davidson, K. R. Scherer, and H. H. Goldsmith, Eds. London, U.K.: Oxford Univ. Press, 2003, pp. 596–618. [Online]. Available: https://psycnet.apa.org/record/2009-07773-030

[39] E. Fouragnan, G. Chierchia, S. Greiner, R. Neveu, P. Avesani, and G. Coricelli, "Reputational priors magnify striatal responses to violations of trust," *J. Neurosci.*, vol. 33, no. 8, pp. 3602–3611, Feb. 2013, doi: 10.1523/jneurosci.3086-12.2013.

[40] F. Fukuyama, *Trust: The Social Virtues and the Creation of Prosperity*, vol. 99. New York, NY, USA: Free Press, 1995.

[41] M. Gagnon, L. Chérif, and A. Roy-Charland, "Contextual cues about reciprocity impact ratings of smile sincerity," *Cognition Emotion*, vol. 36, no. 6, pp. 1181–1195, Sep. 2022, doi: 10.1080/02699931.2022.2090903.

[42] J. J. Gross and O. P. John, "Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being," *J. Personality Social Psychol.*, vol. 85, no. 2, pp. 348–362, 2003.

[43] H. Gunes and M. Piccardi, "Bi-modal emotion recognition from expressive face and body gestures," *J. Netw. Comput. Appl.*, vol. 30, no. 4, pp. 1334–1345, Nov. 2007, doi: 10.1016/j.jnca.2006.09.007.

[44] H. Gunes and M. Piccardi, "From mono-modal to multi-modal: Recognition using visual modalities," in *Intelligent Environments* (Advanced Information and Knowledge Processing), D. Monekosso et al., Eds. Springer-Verlag, 2009, doi: 10.1007/978-1-84800-346-0_10.

[45] U. Hess, S. Blairy, and R. E. Kleck, "The influence of facial emotion displays, gender, and ethnicity on judgments of dominance and affiliation," *J. Nonverbal Behav.*, vol. 24, no. 4, pp. 265–283, 2000.

[46] C. E. Izard, "Basic emotions, relations among emotions, and emotion-cognition relations," *Psychol. Rev.*, vol. 99, no. 3, pp. 561–565, 1992.

[47] N. D. Johnson and A. A. Mislin, "Trust games: A meta-analysis," *J. Econ. Psychol.*, vol. 32, no. 5, pp. 865–889, Oct. 2011, doi: 10.1016/j.joep.2011.05.007.

[48] S. Knack and P. Keefer, "Does social capital have an economic payoff? A cross-country investigation," *Quart. J. Econ.*, vol. 112, no. 4, pp. 1251–1288, Nov. 1997, doi: 10.1162/003355300555475.

[49] B. Knutson, "Facial expressions of emotion influence interpersonal trait inferences," *J. Nonverbal Behav.*, vol. 20, no. 3, pp. 165–182, Sep. 1996.

[50] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis ;Using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012, doi: 10.1109/T-AFFC.2011.15.

[51] D. T. Kong, K. T. Dirks, and D. L. Ferrin, "Interpersonal trust within negotiations: meta-analytic evidence, critical contingencies, and directions for future research," *Acad. Manage. J.*, vol. 57, no. 5, pp. 1235–1255, Oct. 2014, doi: 10.5465/amj.2012.0461.

[52] M. Kosfeld, M. Heinrichs, P. J. Zak, U. Fischbacher, and E. Fehr, "Oxytocin increases trust in humans," *Nature*, vol. 435, no. 7042, pp. 673–676, Jun. 2005.

[53] E. Krumhuber, A. S. R. Manstead, D. Cosker, D. Marshall, P. L. Rosin, and A. Kappas, "Facial dynamics as indicators of trustworthiness and cooperative behavior," *Emotion*, vol. 7, no. 4, pp. 730–735, 2007.

[54] D. N. Kurien, "Body language: Silent communicator at the workplace," *IUP J. Soft Skills*, vol. 4, nos. 1–2, pp. 29–36, 2010. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1610020

[55] J. J. Lee, W. B. Knox, J. B. Wormwood, C. Breazeal, and D. DeSteno, "Computationally modeling interpersonal trust," *Frontiers Psychol.*, vol. 4, Dec. 2013, Art. no. 56004, doi: 10.3389/fpsyg.2013.00893.

[56] Y. Li, Z. Chen, X. Liu, and Y. Qi, "Perceiving the facial trustworthiness: Facial age, emotional expression, and attractiveness," *Quart. J. Exp. Psychol.*, vol. 75, no. 5, pp. 818–829, May 2022, doi: 10.1177/17470218211047176.

[57] S. R. Livingstone and F. A. Russo, "The Ryerson audio-visual database of emotional speech and song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English," *PLOS ONE*, vol. 13, no. 5, 2018, Art. no. e0196391, doi: 10.1371/journal.pone.0196391.

[58] J. R. Dunn and M. E. Schweitzer, ''Feeling and believing: The influence of emotion on trust,'' *J. Personality Social Psychol.*, vol. 88, no. 5, p. 736, 2005.

[59] G. Lucas, G. Stratou, S. Lieblich, and J. Gratch, ''Trust me: Multimodal signals of trustworthiness,'' in *Proc. 18th ACM Int. Conf. Multimodal Interact.*, Oct. 2016, pp. 5–12, doi: 10.1145/2993148.2993178.

[60] C. Luna-Jiménez, S. L. Lutfi, F. Fernández-Martínez, and A. Tick, ''Measuring trust at zero-acquaintance: A cross-cultural study between Malaysians and Hungarians,'' in *Proc. IEEE 26th Int. Conf. Intell. Eng. Syst. (INES)*, Malaysia, Aug. 2022, pp. 267–272.

[61] C. Luna-Jimenez, S. L. Lutfi, M. Gil-Martin, R. Kleinlein, J. M. Montero, and F. Fernandez-Martinez, ''Measuring trust at zero-acquaintance using acted-emotional videos,'' in *Proc. IberSPEECH*, 2022, pp. 206–210, doi: 10.21437/IberSPEECH.2022-42.

[62] C. Luna-Jiménez, D. Griol, Z. Callejas, R. Kleinlein, J. M. Montero, and F. Fernández-Martínez, ''Multimodal emotion recognition on RAVDESS dataset using transfer learning,'' *Sensors*, vol. 21, no. 22, p. 7665, Nov. 2021, doi: 10.3390/s21227665.

[63] S. L. Lutfi, J. M. Montero, R. Barra-Chicote, J. M. Lucas-Cuesta, and A. Gallardo-Antolín, ''Expressive speech identifications based on hidden Markov model,'' in *Proc. HEALTHINF*, 2009, pp. 488–494.

[64] S. L. Lutfi, F. Fernández-Martínez, J. M. Lucas-Cuesta, L. López-Lebón, and J. M. Montero, ''A satisfaction-based model for affect recognition from conversational features in spoken dialog systems,'' *Speech Commun.*, vol. 55, nos. 7–8, pp. 825–840, Sep. 2013.

[65] C. Maaoui and A. Pruski, ''Emotion recognition through physiological signals for human-machine communication,'' in *Cutting Edge Robotics*, 2010, pp. 317–332, doi: 10.5772/10312.

[66] J. Manfredonia, A. Bangerter, N. V. Manyakov, S. Ness, D. Lewin, A. Skalkin, M. Boice, M. S. Goodwin, G. Dawson, R. Hendren, B. Leventhal, F. Shic, and G. Pandina, ''Automatic recognition of posed facial expression of emotion in individuals with autism spectrum disorder,'' *J. Autism Develop. Disorders*, vol. 49, no. 1, pp. 279–293, Jan. 2019.

[67] V. Maslej-Krešňáková, M. Sarnovský, P. Butka, and K. Machová, ''Comparison of deep learning models and various text pre-processing techniques for the toxic comments classification,'' *Appl. Sci.*, vol. 10, no. 23, p. 8631, Dec. 2020, doi: 10.3390/app10238631.

[68] D. Matsumoto and H. S. C. Hwang, ''Culture, emotion, and expression,'' in *Cross-Cultural Psychology: Contemporary Themes and Perspectives*. Hoboken, NJ, USA: Wiley, 2019, pp. 501–515, doi: 10.1002/9781119519348.ch24.

[69] L. Z. McArthur and K. Apatow, ''Impressions of baby-faced adults,'' *Social Cognition*, vol. 2, no. 4, pp. 315–342, 1984.

[70] J. A. M. Correa, M. K. Abadi, and N. I. S. Patras, ''AMIGOS: A dataset for affect, personality and mood research on individuals and groups,'' *IEEE Trans. Affect. Comput.*, vol. 12, no. 2, pp. 479–493, Nov. 2018, doi: 10.1109/taffc.2018.2884461.

[71] C. J. Mondloch, N. L. Nelson, and M. Horner, ''Asymmetries of influence: Differential effects of body postures on perceptions of emotional facial expressions,'' *PLoS ONE*, vol. 8, no. 9, Sep. 2013, Art. no. e73605, doi: 10.1371/journal.pone.0073605.

[72] J. M. Montepare and H. Dobish, ''The contribution of emotion perceptions and their overgeneralizations to trait impressions,'' *J. Nonverbal Behav.*, vol. 27, no. 4, pp. 237–254, 2003.

[73] J. M. Montepare and L. A. Zebrowitz, ''Person perception comes of age: The salience and significance of age in social judgments,'' in *Advances in Experimental Social Psychology*, vol. 30, P. Mark, Ed. New York, NY, USA: Academic, 1998, pp. 93–161.

[74] C. D. Mayers and D. Tingley, ''The influence of emotion on trust,'' in *Political Analysis*, vol. 24, no. 4. Cambridge, U.K.: Cambridge Univ. Press, 2016, pp. 492–500, doi: 10.1093/pan/mpw026.

[75] O. Niebuhr, J. Voße, and A. Brem, ''What makes a charismatic speaker? A computer-based acoustic-prosodic analysis of steve jobs tone of voice,'' *Comput. Hum. Behav.*, vol. 64, pp. 366–382, Nov. 2016, doi: 10.1016/j.chb.2016.06.059.

[76] B. Nojavanasghari, T. Baltrušaitis, C. E. Hughes, and L.-P. Morency, ''EmoReact: A multimodal approach and dataset for recognizing emotional responses in children,'' in *Proc. 18th ACM Int. Conf. Multimodal Interact. (ICMI)*, Oct. 2016, pp. 137–144, doi: 10.1145/2993148.2993168.

[77] N. N. Oosterhof and A. Todorov, ''The functional basis of face evaluation,'' *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 32, pp. 11087–11092, Aug. 2008.

[78] M. Pantic and M. S. Bartlett, ''Machine analysis of facial expressions,'' in *Face Recognition*. I-Tech Education and Publishing, 2007, pp. 377–416. [Online]. Available: https://research.utwente.nl/en/publications/machine-analysis-of-facial-expressions

[79] M. Pantie and L. J. M. Rothkrantz, ''Automatic analysis of facial expressions: The state of the art,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1424–1445, Dec. 2000, doi: 10.1109/34.895976.

[80] L. Pepino, P. Riera, and L. Ferrer, ''Emotion recognition from speech using wav2vec 2.0 embeddings,'' in *Proc. INTERSPEECH*, 2021, pp. 3400–3404, doi: 10.21437/interspeech.2021-703.

[81] R. D. Putnam, R. Leonardi, and R. Y. Nonetti, *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton Univ. Press, 1993, doi: 10.2307/j.ctt7s8r7.

[82] J. Ren, R. Ding, S. Li, M. Zhang, D. Wei, C. Feng, P. Xu, and W. Luo, ''Features and extra-striate body area representations of diagnostic body parts in anger and fear perception,'' *Brain Sci.*, vol. 12, no. 4, p. 466, Mar. 2022, doi: 10.3390/brainsci12040466.

[83] P. Resnick and R. Zeckhauser, ''Trust among strangers in internet transactions: Empirical analysis of eBay's reputation system,'' in *The Economics of the Internet and e-Commerce* (Advances in Applied Microeconomics), vol. 11, M. R. Baye, Ed. Leeds, U.K.: Emerald Group Publishing Limited, 2002, pp. 127–157, doi: 10.1016/S0278-0984(02)11030-3.

[84] U. C. Bureau, ''Memorandum 2018.02: Using two separate questions for race and ethnicity in 2018 end-to-end census test and 2020 census,'' United States Dept. Commerce, Econ. Statist. Admin., U.S. Census Beureau, Washington, DC, USA, Tech. Rep. 727, 2018.

[85] D. L. Roter, R. M. Frankel, J. A. Hall, and D. Sluyter, ''The expression of emotion through nonverbal behavior in medical visits,'' *J. Gen. Internal Med.*, vol. 21, no. 1, pp. 28–34, 2006.

[86] N. Ruiz, ''Cognitive load measurement in multimodal interfaces,'' M.S. thesis, Univ. New South Wales, Sydney, NSW, Australia, doi: 10.26190/unsworks/23501.

[87] E. Sariyanidi, H. Gunes, and A. Cavallaro, ''Automatic analysis of facial affect: A survey of registration, representation, and recognition,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 6, pp. 1113–1133, Jun. 2015, doi: 10.1109/TPAMI.2014.2366127.

[88] J. P. W. Scharlemann, C. C. Eckel, A. Kacelnik, and R. K. Wilson, ''The value of a smile: Game theory with a human face,'' *J. Econ. Psychol.*, vol. 22, no. 5, pp. 617–640, 2001, doi: 10.1016/S0167-4870(01)00059-9.

[89] P. F. Secord, W. F. Dukes, and W. Bevan, ''Personalities in faces: I. An experiment in social perceiving,'' *Genetic Psychol. Monographs*, vol. 49, p. 231279, 1954.

[90] J. A. Simpson, ''Foundations of interpersonal trust,'' in *Social Psychology: Handbook of Basic Principles*, 2nd ed. The Guilford Press, 2007, pp. 587–607. [Online]. Available: https://www.researchgate.net/publication/232444661_Foundations_of_interpersonal_trust

[91] M. L. Slepian and E. W. Carr, ''Facial expressions of authenticity: Emotion variability increases judgments of trustworthiness and leadership,'' *Cognition*, vol. 183, pp. 82–98, Feb. 2019, doi: 10.1016/j.cognition.2018.10.009.

[92] C. Sofer, R. Dotsch, D. H. J. Wigboldus, and A. Todorov, ''What is typical is good: The influence of face typicality on perceived trustworthiness,'' *Psychol. Sci.*, vol. 26, no. 1, pp. 39–47, Jan. 2015, doi: 10.1177/0956797614554955.

[93] J. L. Sullivan and J. E. Transue, ''The psychological underpinnings of democracy: A selective review of research on political tolerance, interpersonal trust, and social capital,'' *Annu. Rev. Psychol.*, vol. 50, no. 1, pp. 625–650, Feb. 1999.

[94] A. Todorov, C. Y. Olivola, R. Dotsch, and P. Mende-Siedlecki, ''Social attributions from faces: Determinants, consequences, accuracy, and functional significance,'' *Annu. Rev. Psychol.*, vol. 66, no. 1, pp. 519–545, Jan. 2015.

[95] I. Torre, J. Goslin, L. White, and D. Zanatto, ''Trust in artificial voices,'' in *Proc. Technol., Mind, Soc. TechMindSociety, Technol., Mind, Soc.*, 2018, pp. 1–6, doi: 10.1145/3183654.3183691.

[96] S. W. White, L. Abbott, A. T. Wieckowski, N. N. Capriola-Hall, S. Aly, and A. Youssef, "Feasibility of automated training for facial emotion expression and recognition in autism," *Behav. Therapy*, vol. 49, no. 6, pp. 881–888, Nov. 2018.
[97] M. Willinger, C. Keser, C. Lohmann, and J. C. Usunier, "A comparison of trust and reciprocity between France and Germany: Experimental investigation based on the investment game," *J. Econ. Psychol.*, vol. 24, no. 4, pp. 447–466, 2003, doi: 10.1016/S0167-4870(02)00165-4.
[98] R. K. Wilson and C. C. Eckel, "Judging a book by its cover: Beauty and expectations in the trust game," *Political Res. Quart.*, vol. 59, no. 2, pp. 189–202, Jun. 2006.
[99] Y. Wu and T. S. Huang, "Human hand modeling, analysis and animation in the context of HCI," in *Proc. Int. Conf. Image Process.*, vol. 3, 1999, pp. 6–10, doi: 10.1109/ICIP.1999.817058.
[100] A. Zadeh, R. Zellers, E. Pincus, and L.-P. Morency, "MOSI: Multimodal corpus of sentiment intensity and subjectivity analysis in online opinion videos," 2016, *arXiv:1606.06259*.
[101] Y. Zhao, N. D. Georganas, E. M. Petriu, and Y. Zhao, "Human emotion recognition from body language of the head using soft computing techniques," *J. Ambient Intell. Humanized Comput.*, vol. 4, pp. 121–140, Feb. 2012.
[102] Z. Du, S. Wu, D. Huang, W. Li, and Y. Wang, "Spatio-temporal encoder-decoder fully convolutional network for video-based dimensional emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 12, no. 3, pp. 565–578, Jul. 2021, doi: 10.1109/TAFFC.2019.2940224.
[103] J. Han, Z. Zhang, Z. Ren, and B. Schuller, "EmoBed: Strengthening monomodal emotion recognition via training with crossmodal emotion embeddings," *IEEE Trans. Affect. Comput.*, vol. 12, no. 3, pp. 553–564, Jul. 2021, doi: 10.1109/TAFFC.2019.2928297.

SYAHEERAH LEBAI LUTFI received the D.E.A. and Ph.D. (cum laude) degrees from the Grupo de Technologia del Habla (GTH)-Speech Technology Group, University Politecnica de Madrid, Spain, in 2013. She is currently the Deputy Director of global engagement and strategies with the USM International Office and the Project Manager of the Tropical Data Science School (TDSS) and the NUSA Internship Program, which places Indonesian interns in industry in Penang. Her current research interests include human–computer interaction, especially affective computing (behavioral analysis of personality, culture, mood and emotion analysis, and modeling).

BADR LAHASAN received the B.S. degree in computer science from Irbid University, Jordan, in 2004, the M.S. degree in computer science from Universiti Sains Malaysia, in 2012, and the Ph.D. degree in computer science (computer vision and image analysis) from Universiti Sains Malaysia, in 2017. He is currently an Associate Professor with the Faculty of Computer and Information Technology, University of Shabwah, Shabwah, Yemen. His current research interests include face recognition, biometrics, pattern recognition, and artificial intelligence.

CRISTINA LUNA-JIMÉNEZ received the bachelor's, master's, and Ph.D. degrees in telecommunication engineering from Universidad Politécnica de Madrid (UPM).

During her thesis, she did a three-month research stay with Universiti Sains Malaysia (USM). She received the one-year Erasmus grant for studying with Katholieke Universiteit Leuven (KUL) during the master's degree. She is currently a Postdoctoral Researcher with Universidad de Granada, attached to the International Project MENHIR and the National Project GOMINOLA. She was a Researcher and an Assistant Professor with Universidad Politécnica de Madrid, for more than 4 years, associated with the Department of Electronic Engineering, UPM. As a result, she has collaborated on more than six Spanish national projects, one seed project with USM, and one international project (ASTOUND). As a result of her investigation efforts, she has authored and coauthored more than 18 papers in international journals and conferences. Her current research interests include affective computing (emotion and trustworthiness recognition), perception, and multimedia signal processing.

ZAHER A. BAMASOOD received the M.Sc. degree from the King Fahd of Petroleum and Minerals (KFUPM), in 2014. He was a Network Engineer. He is currently a Researcher of data science, with an emphasis on topics related to big data and data mining. He is also a Teacher of computer science with Hadhramout University and a Supervisor of the IT Department, Nesto Tourist Hotel, Mukalla, Yemen.

ZAHID AKHTAR received the Ph.D. degree in electronic and computer engineering from the University of Cagliari, Italy. He is currently an Assistant Professor with the Department of Network and Computer Security, State University of New York (SUNY) Polytechnic Institute, USA. Prior to that, he was a Research Assistant Professor with the University of Memphis, USA, and a Postdoctoral Fellow with the INRS-EMT, University of Quebec, Canada; the University of Udine, Italy; Bahcesehir University, Turkey; and the University of Cagliari. His current research interests include computer vision and machine learning with applications to cybersecurity, biometrics, affect recognition, image and video processing, and audiovisual multimedia quality assessment.