Computers, Materials & Continua
DOI:10.32604/cmc.2022.023738
Article

Tech Science Press

Empathic Responses of Behavioral-Synchronization in Human-Agent Interaction

Sung Park[1,*], Seongeon Park[2] and Mincheol Whang[2]

[1]Savannah College of Art and Design, Savannah, GA, 31401, USA
[2]Sangmyung University, Seoul, 03016, Korea
*Corresponding Author: Sung Park. Email: spica7601@gmail.com
Received: 19 September 2021; Accepted: 20 October 2021

Abstract: Artificial entities, such as virtual agents, have become more pervasive. Their long-term presence among humans requires the virtual agent's ability to express appropriate emotions to elicit the necessary empathy from the users. Affective empathy involves behavioral mimicry, a synchronized co-movement between dyadic pairs. However, the characteristics of such synchrony between humans and virtual agents remain unclear in empathic interactions. Our study evaluates the participant's behavioral synchronization when a virtual agent exhibits an emotional expression congruent with the emotional context through facial expressions, behavioral gestures, and voice. Participants viewed an emotion-eliciting video stimulus (negative or positive) with a virtual agent. The participants then conversed with the virtual agent about the video, such as how the participant felt about the content. The virtual agent expressed emotions congruent with the video or neutral emotion during the dialog. The participants' facial expressions, such as the facial expressive intensity and facial muscle movement, were measured during the dialog using a camera. The results showed the participants' significant behavioral synchronization (i.e., cosine similarity $\geq .05$) in both the negative and positive emotion conditions, evident in the participant's facial mimicry with the virtual agent. Additionally, the participants' facial expressions, both movement and intensity, were significantly stronger in the emotional virtual agent than in the neutral virtual agent. In particular, we found that the facial muscle intensity of AU45 (Blink) is an effective index to assess the participant's synchronization that differs by the individual's empathic capability (low, mid, high). Based on the results, we suggest an appraisal criterion to provide empirical conditions to validate empathic interaction based on the facial expression measures.

Keywords: Facial emotion recognition; facial expression; virtual agent; virtual human; embodied conversational agent; empathy; human-computer interaction

This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1 Introduction

The prevalence of AI technology, including deep fake or advanced 3D modeling, has introduced unprecedented virtual humans closely resembling human appearance and behavior. For example, virtual human Roji created by Sidus Studio X has gained immense popularity in Korea as an advertisement model through its life-like motion and behavior. People did not realize that Roji was a virtual human until the company revealed it four months after its public debut. The human-like character was achieved using an AI model, learning actual human facial expressions and behavior, and applying the learned facial and bodily patterns to the 3D model.

However, virtual humans are not new. They have been used in many domains, including advertisements as well as medical practice [1,2], healthcare [3,4], education [5,6], entertainment [7,8], and the military [9,10], interacting with the user and acting on the environment to provide a positive influence, such as behavioral change of the human counterpart [11]. The emphasis on interactivity with the user brought the term virtual agent, which utilizes verbal (e.g., conversation) and non-verbal communication (e.g., facial expressions and behavioral gestures) channels to learn, adapt, and assist the human. Similar to human-to-human communication, understanding the human dyad's emotion [12], expressed by speech, body motion, and facial expression, is paramount for an effective conversation. For the virtual agent to help humans with their mental or health-related problems or assist their daily activities, the virtual agent should be capable of empathizing with the human user. That is, recognizing humans' emotional states, thoughts, and situations and behave accordingly.

1.1 Literature Review on Empathy Research

We feel similar emotions to other people, which is sometimes a result of understanding others' thoughts and feelings. Empathy involves "an affective response more appropriate to someone else's situation than to one's own" [13]. The crux considers the other's affective state and situation for cooperation, prosocial behavior, and positive relationships [13–16].

Empathy research has been conducted in social development, clinical psychology, and neuroscience. Since discovering mirror neurons in monkeys [17], neuroscientists have identified underlying neurological evidence for empathy [18]. Overlapping brain patterns were observed when an observer perceived the same emotions from a target, suggesting shared affective neural networks [19–21].

However, there is no consensus on the definition of empathy. The number of definitions is proportional to the number of researchers [22]. Researchers agree that empathy consists of multiple subcomponents [13,23,24], and some critical elements of empathy (recognition, process, outcome, response) are commonly identified (for an extensive review of empathy as a concept, see [25]). A typical empathic episode initiates when the observer perceives empathic cues (expression or situation) from the target through verbal (e.g., "I don't feel well") or non-verbal channels. The observer then engages in an internal affective or cognitive process, which may result in a congruent emotional state (i.e., feeling through the target), and if willing, an empathic response (e.g., "I understand that you are in a bad mood"). Based on the most prominent empathy theories [13,23,26–28], the affective or cognitive processes are the underlying mechanisms that produce empathic outcomes. Because the latter, cognitive empathy, is an extensive research field, including perspective taking, we limited our research to affective empathy.

The crux of affective empathy is motor mimicry, an observer's automatic and unconscious imitation of the target. Mimicry was first described by Lipps and organized by Hoffman [29] into a two-step process: 1) the observer imitates the target's empathic expressions (e.g., facial expression, voice, and posture); 2) this imitation results in afferent feedback that produces a parallel affect congruent

with the target's feedback. For example, a virtual agent may imitate a human's facial expression, who looks cheerful and changes their emotional state accordingly. This mechanism is also referred to as primitive emotional contagion [30] or the chameleon effect [31]. Mimicry is essential in building rapport [32] and makes the observer more persuasive [33]; however, in certain situations, it may have a diminishing effect.

1.2 Limitations of Empathy Research with Virtual Agents

Because empathy is a directional construct involving an observer empathizing with a target, empathy research related to a virtual agent is twofold: research on 1) the virtual agent (observer) empathizing with the human user (target), or 2) the human user (observer) empathizing with the virtual agent (target).

Empathic virtual agents have been studied in the context of playing games [34–36], health-care interventions [37,38], job-interviews [39], email assistance [40], social dialog [41], or even a story narrative [42]. A typical empirical study evaluated the participants' perceptions when interacting with or observing an empathic virtual agent compared to a non-empathic one. Overall, empathic agents were perceived positively in liking [35,37,41,42], trust [35,37] and felt more human-like [34], caring [34,35], attractive [34], respected [37], and enjoyable [38]. While most studies were based on one-time interaction, a few studies identified the participants' intention to use empathic virtual agents longer [37,38]. The research community certainty has established a grounding that an empathic virtual agent, when implemented to provide an appropriate response congruent with the situation, elicits a positive perception of the users with a perspective for long-term interaction.

However, some studies have investigated the participants' empathic and emotional responses to virtual agents. The participants' affective states were estimated through a dialog (e.g., multiple choices) [39–41,43], or more direct measures such as physiological measures (e.g., skin conductance, EMG) [34,36,39] and facial expressions [38,44]. Lisetti et al. [38] developed a virtual counselor architecture with a facial expression recognizer model to recognize the participant's facial expression as part of its multimodal empathy model. The participants' facial photos were obtained from the JPEG-Cam library and sent for analysis to the processing module. The analysis is yet to return limited output such as smiling status and five emotion categories (happy, sad, angry, surprised, and neutral) and not the kind of advanced analysis such as facial mimicry. That is, the study did not directly investigate the facial synchronization. Prendinger et al. [39] evaluated psychophysiological measures (e.g., skin conductance and heart rate) of the participants interacting with an empathic companion. The study suggested that the companion's comments, displayed through text, positively affected the participant's stress level. The companion was designed to express empathy based on a decision network. However, the research did not provide direct evidence suggesting behavioral mimicry, such as the convergence of heart rate [45], evidence of empathy. Ochs et al. [40] suggested an emotion model for an empathic dialog agent that infer a user's emotions considering the user's beliefs, intentions, and uncertainties. The model captured the dynamic change of the agent's mental state due to an emotion eliciting event. However, the inference is yet a logical approximation, absent of direct empirical evidence suggesting empathic synchronization.

1.3 Contribution of Current Study

While some studies have suggested a promising empathy recognition model, few studies empirically validated its effectiveness with human participants. Empirical data are limited to indirect methods, such as multiple choices during dialog [39–41]. The empathic synchronization is approximated

with the participant's subjective responses to an empathic agent [40]. That is, no study directly compares the emotional expression between a participant and a virtual agent on an equal scale to assess empathy.

Our research is interested in understanding the participants' behavioral mimicry when they empathize with the virtual agent by analyzing the participants' facial expression changes, including the intensity and movement of the facial appearance and muscles. To the best of our knowledge, this is the first study to directly compare a battery of facial expression measures between a participant and an empathic virtual agent. Furthermore, this is the first research to analyze how such measures differed as a function of participants' empathy capability (low, mid, high). Based on the analysis, we provide an empirically validated criterion for detecting a participant's empathic state when interacting with a virtual agent. Such a criterion is paramount to inform the virtual agent or any kind of AI system to adapt its emotion and behavior to the user in real time.

2 Methods

By definition, the observer's (the participant) empathy occurs when the observer recognizes and relates to the target's (the virtual agent) emotional expression (facial expression, behavioral gesture, voice) congruent with the situation. We designed our experiment to achieve an emotion-embedded shared experience (viewing either positive or negative video clips) between the participant and the virtual agent. Through dialog, the virtual agent expresses an emotion congruent with the valence of the video so that the participant can empathize with the virtual agent. The emotion expressed is varied; emotional or neutral, and the valence of the video stimuli is either positive or negative.

2.1 Research Hypothesis

Our research is directed towards verifying the following three hypotheses:

- H_1: There is a difference in the facial synchronization between the participant and virtual agent when interacting with an emotional virtual agent and a neutral virtual agent.
- H_2: The participants' facial expressions differ when interacting with an emotional virtual agent and a neutral virtual agent.
- H_3: The participants' facial expressions differ depending on the level of the empathic capability of the participant.

2.2 Manipulation Check

The current research used a video stimulus to evoke emotions and a dialog to evoke empathy. We conducted a manipulation check to ensure that the video stimuli and dialog interaction were effective before the main experiment, in which participants' facial data were acquired. Thirty university students were recruited as the participants. The participants' ages ranged from 21 to 40 years (mean = 28, SD = 4), with 16 males and 14 females. We defined four interaction cases between the *emotion valence* factor (negative or positive) and *virtual agent expression* factor (emotional or neutral) where the participant viewed the video stimulus, conversed with the virtual agent, and responded to a questionnaire. We adopted Russell's valence dimension in his circumplex model [46], where emotional states can be defined at any or neutral level of valence. The materials (video clip and virtual agent) were identical to those used in the main experiment. We used video stimuli known to elicit emotions, which were organized and empirically validated by Stanford University (n = 411, [47]).

The experimenter explained the experiment procedure and clarified the terms in the questionnaire. After viewing each stimulus and conversing with the virtual agent, the participants responded to the survey (see Fig. 1). The interaction with the virtual agent was identical to the main experiment where the virtual agent led the conversation asking a series of questions related to the shared viewing experience of the content (details of the virtual agent and dialog will be explained in Section *2.6 Materials and Data Acquisition*). The order of the stimulus and the virtual agent were randomized.

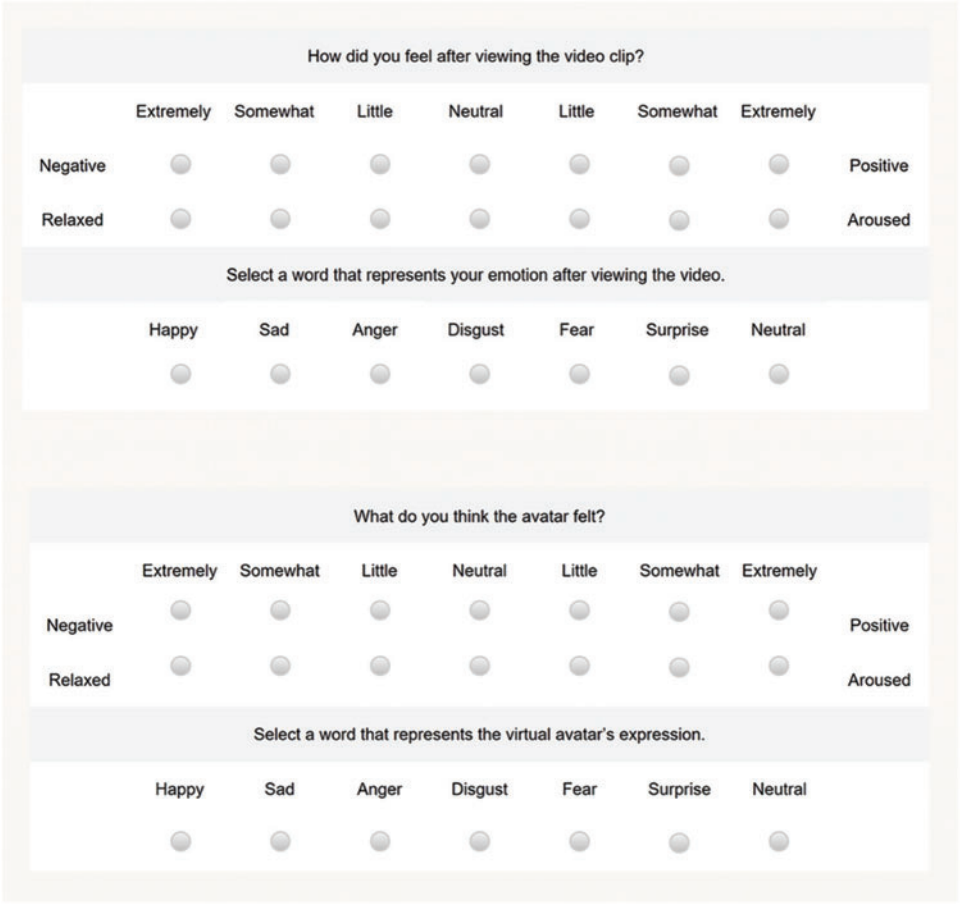

Figure 1: Questionnaire for manipulation check

We analyzed the participants' perceptions (i.e., questionnaire responses) to determine whether the video stimuli could evoke emotion and if the virtual agent could elicit empathy from the participant. We concluded that participants reported congruent emotions with the target emotion of the stimulus (▲ in Fig. 2). That is, the participants reported negative valence in the negative emotion condition and positive valence in the positive emotion condition.

To test whether the virtual agent would elicit empathy, we conducted a paired sample t-test on the questionnaire data after validating the normality through the Shapiro-Wilk test (see Fig. 3). Whether the participant emphasized the virtual agent was analyzed based on the vector distance between the participant's emotional state and perceived emotional state of the virtual agent. In the negative emotion condition, the distance between the participant's emotional state and perceived emotional state of the virtual agent was significantly smaller in the emotional virtual agent condition than in the

neutral condition ($t = -5.14$, $p < .001$) (Figs. 2a and 3a). Similarly, in the positive emotion condition, the vector distance between the participant's emotional state and perceived emotional state of the virtual agent was significantly lower in the emotional virtual agent condition than in the neutral condition ($t = -6.41$, $p < .001$) (Figs. 2b and 3b).

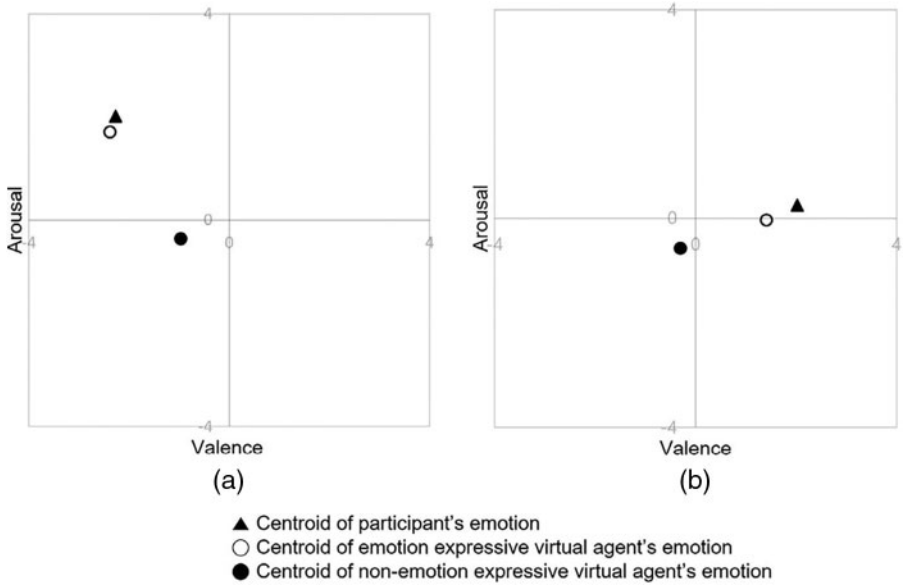

Figure 2: Distance between emotional states. (a) Negative emotion (b) Positive emotion

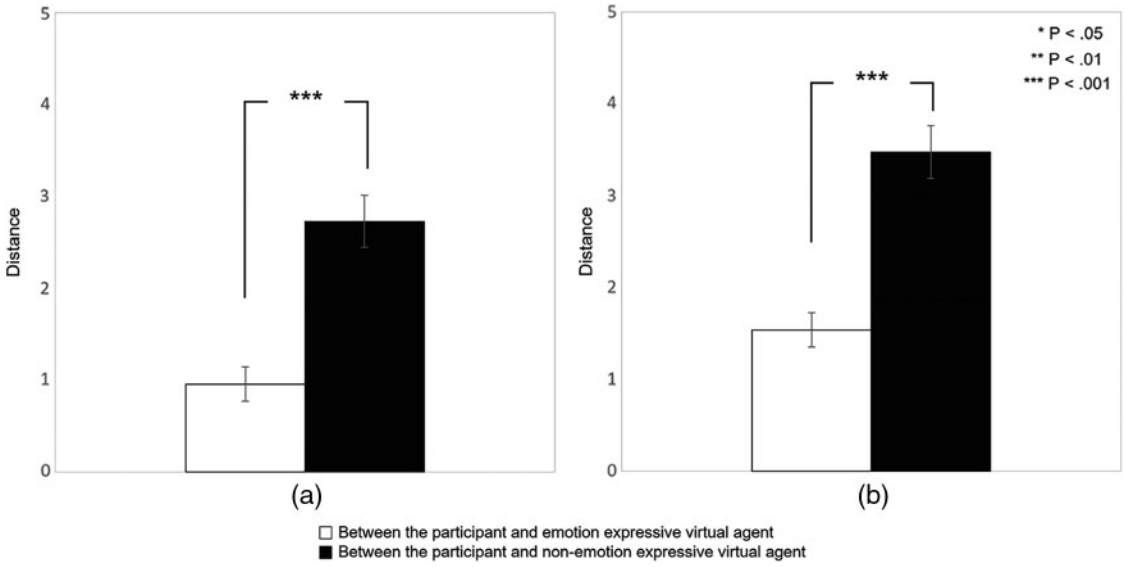

Figure 3: T-tests results on mean distance between emotional states. (a) Negative emotion (b) Positive emotion

2.3 Experiment Design

The main experiment was a mixed factorial design of 2 (virtual agent expression: emotional and neutral) × 2 (emotion valence: negative and positive). The virtual agent expression was a between-subject factor, whereas the emotion valence factor was a within-subject factor design. That is, a participant interacted with only one type of virtual agent's expression, but the virtual agent would express with all two emotions. The participants were randomly distributed with equal numbers between a group that would interact with an emotional virtual agent and a group that would interact with a neutral virtual agent.

2.4 Participants

We conducted an a priori power analysis with the program G*Power with power set at 0.8 and $\alpha = 0.05$, $d = 0.6$ (independent t-test) and 0.4 (one-way repeated ANOVA), two-tailed. The results suggest an n of approximately 45 would be needed to achieve appropriate statistical power. Therefore, forty-five university students were recruited as participants. The participants' ages ranged from 20 to 30 years (mean = 28, SD = 2.9), with 20 (44%) males and 25 (56%) females. We selected participants with a corrective vision of .8 or above without any vision deficiency, to ensure the participant's reliable recognition of visual stimuli. We recommended that participants have sufficient sleep and prohibit alcohol, caffeine, and smoke the day before the experiment. Because the experiment requires a valid recognition of the participant's facial expression, we limited the use of glasses and cosmetic makeup. All participants were briefed on the purpose and procedure of the experiment and signed a consent form. They were then compensated with participation fees.

2.5 Experiment Procedure

The experiment consisted of two sessions with a week interval between the participant's two visits to the lab (see Fig. 4). The main experiment used identical experimental materials as those used in the manipulation check. In the first session, the experimenter explained the experiment procedure and clarified the terms in the questionnaire to be administered. The participants then responded to a Korean adaptation version of the Empathy Quotient (EQ) survey [48] to assess their empathic capability. The survey is explained in detail in Section *2.6 Materials and Data Acquisition*.

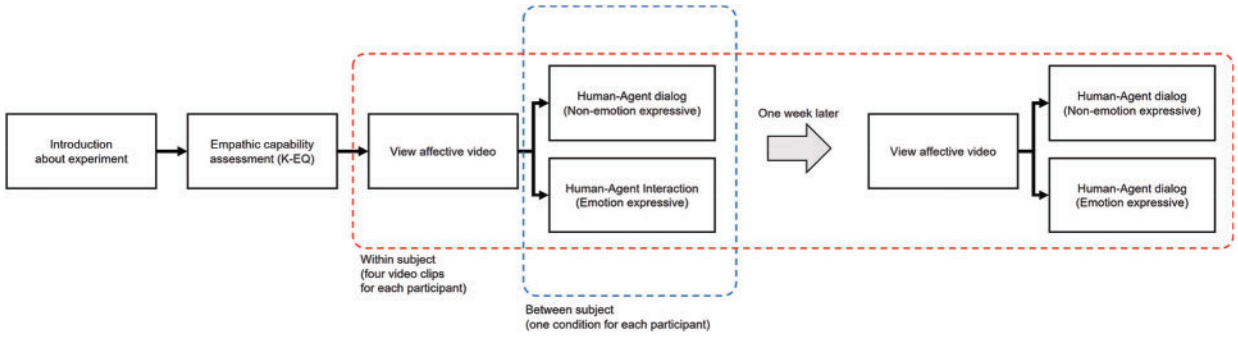

Figure 4: Experiment procedure

The participants were then placed in an experiment room with two monitors. The video stimulus was displayed on the left monitor, whereas the virtual agent was displayed on the right. The camera was placed on top of the right monitor to capture the participant's facial data during the dialog.

The participant then conversed with the virtual agent to build a rapport. After the conversation, the participant viewed the affective video that was designed to evoke emotions; negative emotions in the first session and positive emotions in the second. The virtual agent was programmed to face toward the left monitor as if the virtual agent was viewing the video content played on the left monitor. That is, the participants had a sense of watching the stimulus together. Each video clip lasted for 30 s.

The participant then engaged in a series of interactive dialogs with the virtual agent. The dialog is explained in detail in Section 2.6 *Materials and Data Acquisition*. The dialog lasted approximately four to seven minutes. After the dialog, the participant relaxed for 180 s. The participants then viewed the second video clip with the same emotion, followed by an identical interaction with the virtual agent. The camera captured the participant's facial expressions during the entire dialog. The order of the video clips was randomized. The participants then retired and revisited the lab after a week. Because the evoked emotion tends to permeate throughout the day, we had the interval between the two sessions so that the results could be attributed to a single evoked emotion. In the second session, the participants followed the same steps but with a positive emotion stimulus.

2.6 Materials and Data Acquisition

2.6.1 Empathy Quotient (EQ)

To validate the relationship between the participants' empathic capability and facial expression when interacting with the virtual agent (H_3), we administered a battery of empathy quotient surveys. While there is a long academic history involving the development of a survey to measure empathy in adults, scales have evolved to fully capture empathy without being confounded with other constructs. For example, one of the early scales, Hogan's Empathy (EM) [49], in a strict sense, only had one factor, sensitivity, related to empathy as a construct. The Questionnaire Measure of Emotional Empathy (QMEE) [50] is also considered to measure empathy, but the authors themselves suggest a confounding variable such as being emotionally aroused to the environment that is not relevant to empathy in an interpersonal interaction [51]. Davis' Interpersonal Reactivity Index (IRI) [52] started to include higher-level cognitive empathy, such as perspective taking**,** but is considered to be broader than construct empathy [48].

The most recent Baron-Cohen's EQ (Emotion Quotient) [48] scale was designed to capture both the cognitive and affective components of empathy and validated by a panel of six psychologists to determine whether the battery agrees with the scholarly definition of empathy. The original Baron-Cohen scale is a 4-point scale that includes 40 items and 20 control items, which conveys either a positive or negative emotional valence.

Because any kind of scale designed to measure a psychological construct is affected by cultural differences, we used an adapted and translated version for Koreans, the K-EQ [53]. The Baron-Cohen scale has a four-point Likert scale with bidirectional ends (positive and negative). To put the responses on a unidirectional scale, the K-EQ converts the answer items (1–4) to a three-point Likert scale (0–2). That is, participants answered 0, 1, or 2, so the total score ranged from 0 to 80. The validity and reliability of the K-EQ battery have been verified [53,54].

Using the questionnaire data, we divided the participants into three groups (high, mid, and low) according to the cumulative percentage of the K-EQ score so that the three groups had a distribution of 3:4:3 (see Fig. 5). We report and discuss the participants' differential facial expressions as a function of empathic capability in the Results section.

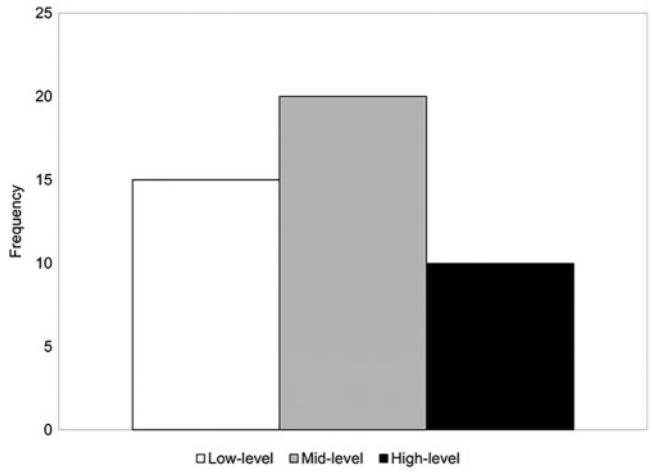

Figure 5: Distribution of empathic capability

2.6.2 Virtual Agent and Dialog

Research indicates that emotional expression elicits differential social responses from different cultural members [55–58]. Ethnic stereotyping also applies to virtual agents [59,60]. Specifically, participants responded to a virtual human of the same ethnicity as more intelligent, trustworthy, attractive, and persuasive [61]. In our research, because the participant would emphasize the virtual agent, we designed the agent to match the participant's ethnicity and race.

The virtual agent was a three-dimensional female character that was refined for the experiment (see Fig. 6). We used the animation software Maya 2018 (Autodesk) to modify an open-source, FBX (Filmbox) formatted virtual agent model. Specifically, we adjusted the number and location of the cheekbone and chin to express the facial expressions according to the experiment design (i.e., negative and positive emotion expressions). We used the Unity 3D engine to animate the modified model and developed an experiment program using C# 4.5, an objective-oriented programming language.

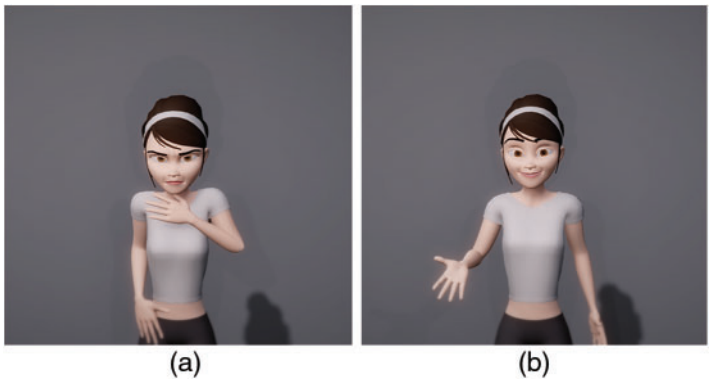

Figure 6: (a) Negative and (b) positive expressions of virtual agent

The virtual agent expressed emotion through three means: facial expression, behavioral gestures, and voice. The agent expressed two emotions, positive and negative, on the valence plane based on the dimensional affect by Russell [46]. The corresponding facial expression was designed based on the

Facial Action Code System (FACS) [62]. The behavioral gestures were designed based on previous studies on the perceived intentions and emotions of gestures [63–65]. For example, in the negative emotion condition, the palms were facing inward, and the arms were bent, concealing the chest (see (a) in Fig. 6). In the positive emotion condition, the virtual agent had the palms facing upward with the arm and chest opened (see (b) in Fig. 6). We used a voice recording of a female in her 20s, congruent with the appearance of the virtual agent. To make the expression as natural and believable as possible, we guided the voice actor to speak as similar as possible to the visual appearance. The tone and manner were congruent with the dialog script. We designed the virtual agent's lips to synchronize with the vocal recordings.

To elicit empathy from the participant, the virtual agent was designed to converse with the participant on the shared experience that just occurred. We referred to the dialog script from previous virtual agent systems [37,42,66,67]. The following is a representative dialog script, which includes rapport building. In this case, the video viewed by the participant and virtual agent was about a playful episode between a baby and her father. Her father teased her by saying, "make an evil look," when obviously she is too young to understand the meaning. She made a frown as if she was responding to her father. The video was empirically validated to elicit a positive response [47]. This video stimulus was one of the stimuli used in the experiment. The bolded script affects the virtual agent's facial expression, behavioral gesture, and voice depending on the expression condition (i.e., emotional *vs.* neutral).

Agent: Hi, my name is Mary. How did you come to the lab today?

Participant: Hi, I took a bus.

Agent: I see. It is nice to see you. Can I ask you about your hobbies?

Participant: I like to listen to music and watch a movie.

Agent: I am really into watching a YouTube video. I want to see something with you. Could you please watch this with me?

Participant: Sure.

(The virtual agent turns her posture on the left monitor).

(After watching the stimulus video).

Agent: **(gestures) Have you seen it well?**

Participant: Yes.

Agent: **(gestures) (in positive condition) The video was really funny (in negative condition: unpleasant).** How about you?

Participant: I thought it was funny too.

Agent: **(gestures) Could you tell me what part of the video made you feel that way?**

Participant: The baby seemed to understand what her father was saying, and her response was so funny.

Agent: **(gestures) How would you feel if the same thing happened to you?**

Participant: I would keep smiling because the baby is so cute!

Agent: **(gestures) How would you behave in that situation?**

Participant: I would tease the baby just like the video.

Agent: **(gestures) Have you experienced something similar in real life?**

Participant: I used to tease my young cousin, similar to the video.

Agent: **(gestures) Could you share an experience that has a similar emotion to the video?**

Participant: I felt similar emotions when viewing a video of a puppy.

Agent: **(gestures) I see. What do you think my expressions were like when viewing the video?**

Participant: I would imagine that you were smiling and had your corner of the lip pulled, just like me.

Agent: **(gestures) I see. What do you think my feelings are now?**

Participant: You probably feel better. You know babies; they make us happy.

The dialog script was presented using the Wizard of Oz method [68] as if the virtual agent was conversing naturally with the participant. We recorded and analyzed the participants' facial expressions during the conversation, which will be explained in detail in the following section.

2.6.3 Facial Data

The facial data were captured by a camera (Logitech, C920), which was fixed on the monitor with the virtual agent to record the participant's frontal facial expressions. The video was captured at a 1920×1080 pixel resolution at a frame rate of 30 fps in an MP4 format. The facial movement was measured through Open face [69], which is an open-source tool developed based on a machine learning model. The library can analyze and track a participant's face in real time. We extracted the facial landmarks, blend shape, action units (AU) and their strength, and head pose data at a rate of 30 fps. The dependent variables elicited from the facial data are as follows.

Facial appearance movement. The Open Face provides feature characteristics such as facial landmarks that include the participant's facial appearance. The two-dimensional landmark coordination (x, y) is a result of the AI decoder, which is trained based on a machine learning model, which consists of 68 features. Because the data are dependent on the size (height, breadth) and location of the face, we normalized the data through the min-max normalization, using Eqs. (1) and (2), resulting in a normalized value [0,1], referred to as the facial appearance movement.

$$X_{normalized}(i = 0, 1, \ldots 67) = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

$$Y_{normalized}(i = 0, 1, \ldots 67) = \frac{y_i - y_{\min}}{y_{\max} - y_{\min}} \tag{2}$$

Facial muscle movement. AU are modular components (i.e., facial muscles) that can be broken down from facial expressions [62]. AU is considered the basic analytical element of a facial cognitive system, the Facial Action Code System (FACS). We used 27 AU, which is the centroid value between the facial landmarks (see Tab. 1). The centroid value is the average of the respective two-dimensional coordinates (x, y) of the facial landmarks that are related to the respective AU.

Facial muscle intensity. Through the dynamic model based on machine learning, Open Face provides the degree of muscle contraction (0–1) through a variable that captures the strength of the facial muscles from AU1 to AU45.

Table 1: Action Units (AU) definitions

Action Units	Full name	Landmarks (Left)	Landmarks (Mid)	Landmarks (Right)
AU1	Inner brow raiser	18, 19, 20, 21		22, 23, 24, 25
AU2	Outer brow raiser	17, 18, 19		24, 25, 26
AU4	Brow lower		17, 18, 25, 26	
AU5	Upper lid raiser	36, 37, 38, 39		42, 43, 44, 45
AU6	Cheek raiser	2, 31, 41		14, 35, 46
AU7	Lid tightener	36, 38, 39, 41		42, 43, 45, 46
AU9	Nose wrinkler	30, 31, 33		30, 33, 35
AU10	Upper lid raiser		31, 35, 49, 51, 53	
AU11	Nasolabial deepener		29, 30, 34, 35	
AU12	Lip corner puller	3, 31, 48		13, 35, 54
AU14	Dimpler	3, 4, 31, 48		12, 13, 35, 54
AU15	Lip corner depressor	4, 6, 48		10, 12, 54
AU16	Lower lip depressor		6, 7, 8, 9, 10, 55, 56, 57, 58, 59	
AU18	Lip puckerer		60, 61, 62, 63, 64, 65, 66, 67	
AU20	Lip stretcher	3, 4, 5, 48		11, 12 , 13 , 54
AU21	Neck tightener		5, 7, 8, 9, 10, 57	
AU22	Lip funneler		50, 52, 56, 58	
AU25	Lips part		61, 62, 63, 65, 66 , 67	
AU26	Jaw drop		6, 8, 10, 57	
AU27	Mouth stretch		6, 10, 48, 54	
AU28	Lip suck		48, 51, 54, 57	
AU29	Jaw thrust		7, 9, 56, 58	
AU30	Jaw sideways		3, 7, 9, 13	
AU35	Cheek suck	2, 4, 31		12, 14, 35
AU38	Nostril dilate	30, 31, 32		30, 34, 35
AU40	Eye closure	36, 37, 38, 39, 40, 41		42, 43, 44, 45, 46, 47
AU45	Blink	36, 37, 38, 39, 41		42, 43, 44, 45, 46

Facial expressive intensity. Using the animation software Maya (Autodesk), we produced AU-based blend shapes for the virtual agent in the experiment. For a more natural look, the blend shapes morphed the base shape to the target shape of the face, using the linear interpolation method. The

facial expressive intensity variable represents the strength of the blend shapes, between 0 and 100, which are based on the facial regions (e.g., brows, eyes, nose, cheek, mouth) involving facial expressions.

Head pose movement. The head pose movement involves the orientation of the participant's head. The research analyzed the head pose to understand a person's point of attention and interest [70–72]. The participant's head pose variable consists of the yaw (X), pitch (Y), and roll (Z) on three-dimensional planes, represented by *Euler angles (see* Fig. 7*)*.

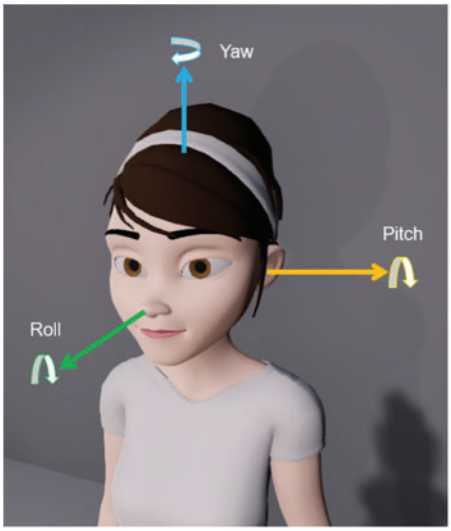

Figure 7: Rotation of the virtual agent's head pose

2.7 Analysis Plan

We eliminated any data outside of the mean ± 3SD as outliers [73]. To validate H_1, we conducted an independent t-test on the degree of facial synchronization of the participant to the virtual agent between the two conditions: emotional and neutral virtual agents. To measure the degree of synchronization, we used cosine similarity, which captures the similarity between the *facial expressive intensity* of the participant and virtual agent. Cosine similarity was used as a distance metric for facial verification [74]. Prior to the t-test, we conducted the Shapiro-Wilt and Levene's test to confirm data normality and homoscedasticity.

To validate H_2, we conducted an independent t-test on dependent measures of the facial expression (*facial expressive intensity, facial appearance movement, facial muscle movement, and head pose*) between the two conditions: emotional and neutral virtual agents. We used the same methods to test data normality and homoscedasticity.

To validate H_3, we conducted a one-way ANOVA test on the *facial muscle intensity*, followed by a post-hoc Scheffe's test. If the data did not meet normality, we conducted the Kruskal-Wallis test instead. If the data did not meet the equal variance criterion, we conducted Welch's ANOVA, followed by a post-hoc Games-Howell test.

3 Results

3.1 Synchronization of Facial Expression

Fig. 8 depicts the differences in the average cosine similarity in the negative condition between the emotional and neutral virtual agents. The brows frown on both sides ($p < .001$), brows down on both sides ($p < .001$), mouth sad on both sides ($p < .001$), narrowing the mouth on both sides ($p < .001$) had a significantly higher cosine similarity in the emotional condition than in the neutral condition. Remarkably, brow frowning and narrowing the mouth had a cosine similarity of more than 0.5. Cosine similarity above 0.5, is considered as a strong synchronization [75,76].

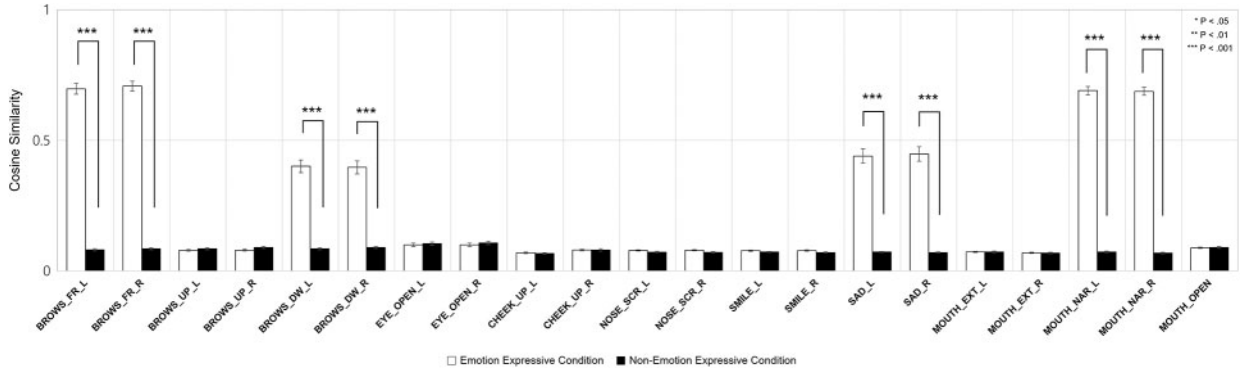

Figure 8: Averaged cosine similarity in the negative emotion condition

Fig. 9 depicts the differences in the average cosine similarity in the positive condition between the emotional and neutral virtual agents. The brows raised on both sides ($p < .001$), cheek raise on both sides ($p < .001$), nose scrunching on both sides ($p < .001$), mouth smiling on both sides ($p < .001$), extending the mouth on both sides ($p < .001$), and mouth opening ($p < .001$) had significantly higher cosine similarities in the emotional condition than the neutral condition. Remarkably, the brows raising, mouth smiling, and mouth opening had a cosine similarity of more than .05.

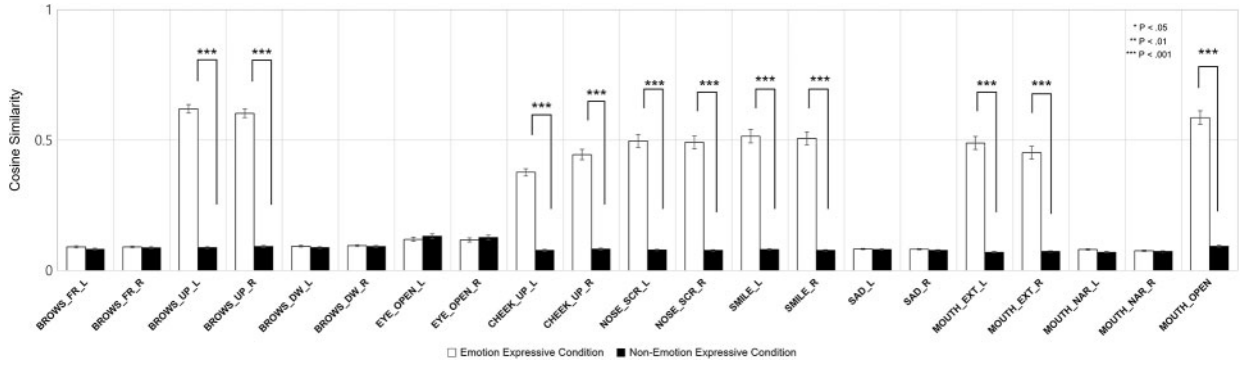

Figure 9: Average cosine similarities in the positive emotion condition

3.2 Difference in Facial Appearance

Fig. 10 depicts the average *facial expressive intensity* (i.e., blend shape strength) in the negative emotion condition. The *facial expressive intensity* of the brows frown on both sides ($p < .001$), brows down on both sides ($p < .05$), and mouth narrow on the left side ($p < .001$) in the emotional condition

was significantly higher than in the neutral condition. Conversely, the *facial expressive intensity* of the brows raised on both sides ($p < .05$) and mouth extension on the left side ($p < .001$) was significantly higher in the neutral condition than in the emotional condition.

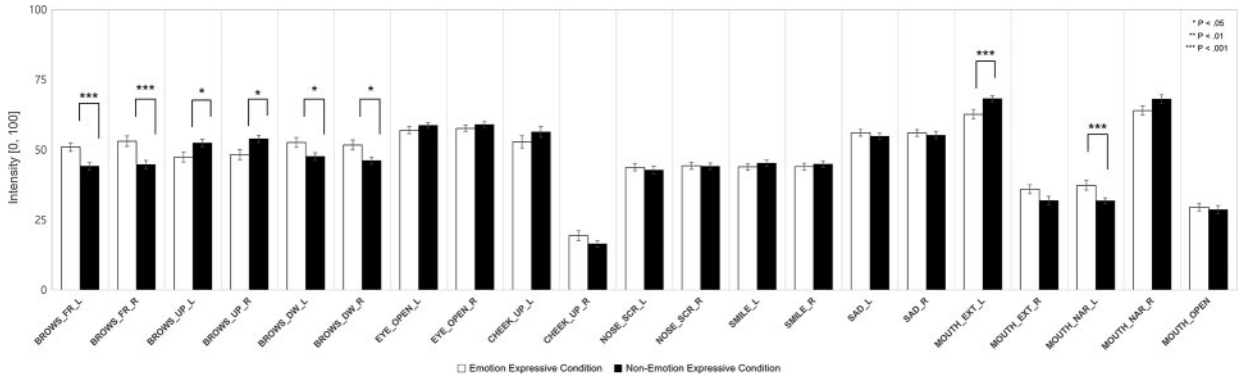

Figure 10: Average facial expressive intensity in the negative emotion condition

Fig. 11 depicts the average *facial expressive intensity* (i.e., blend shape strength) in the positive emotion condition. The *facial expressive intensity* of the mouth extension on the right side ($p < .01$) was higher in the emotional condition than in the neutral condition. The *facial expressive intensity* of the eyes open on both sides ($p < .05$) and the mouth narrow on the right ($p < .05$) was higher in the neutral condition than in the emotional condition.

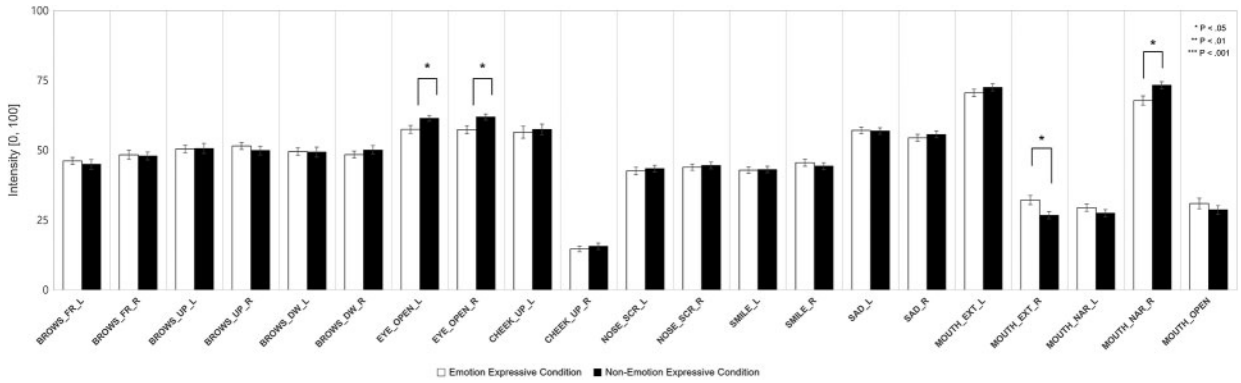

Figure 11: Average facial expressive intensity in the positive emotion condition

Fig. 12a depicts the *facial appearance movement* of the participants in the negative emotion condition. The appearance movement of the eyebrows, eyes, nose, and lips was significantly higher in the emotional condition than in the neutral condition. Fig. 12b shows the positive emotion condition. The movement of the lower jaw and left cheekbone was significantly higher in the emotional condition than in the neutral condition.

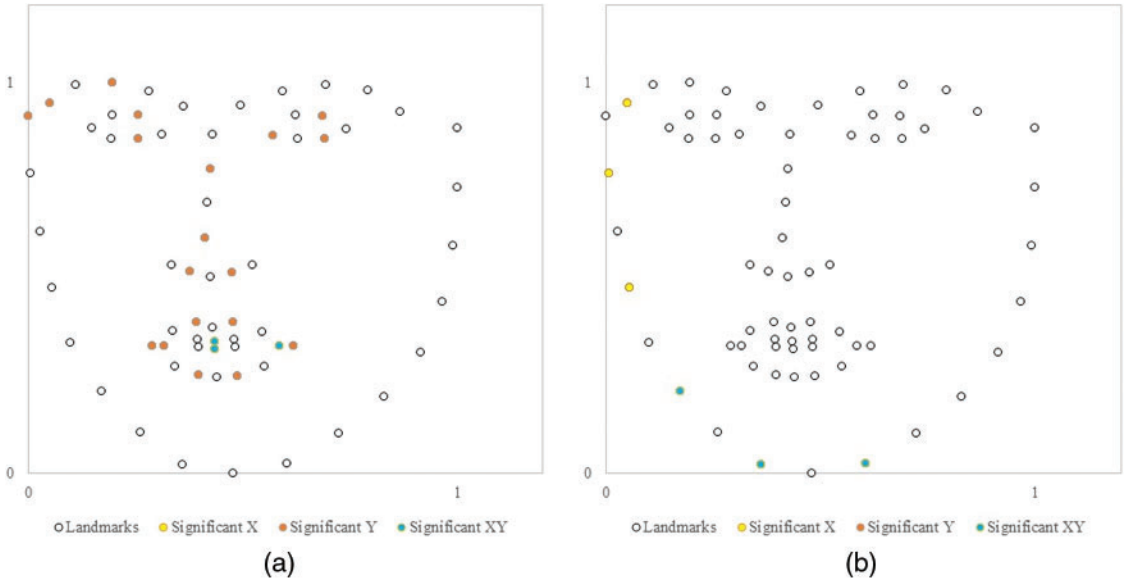

Figure 12: Difference of facial appearance movement. (a) Negative condition (b) Positive condition

Fig. 13a depicts the *facial muscle movement* of the participants in the negative emotion condition. The 68 landmarks and 27 AU are plotted on an X-Y axis. Based on the t-test results, AU with a significant difference ($p < .05$) on the X-axis are colored with yellow, on the Y-axis are colored with orange, and on both axes are colored with green. The movement of the eyes, nose, mouth, chin, cheekbone, and philtrum was significantly higher in the emotional condition than in the neutral condition. Fig. 13b shows the positive emotion condition. The muscle movements of the eyes, nose, mouth, and chin were significantly higher in the emotional condition than in the neutral condition.

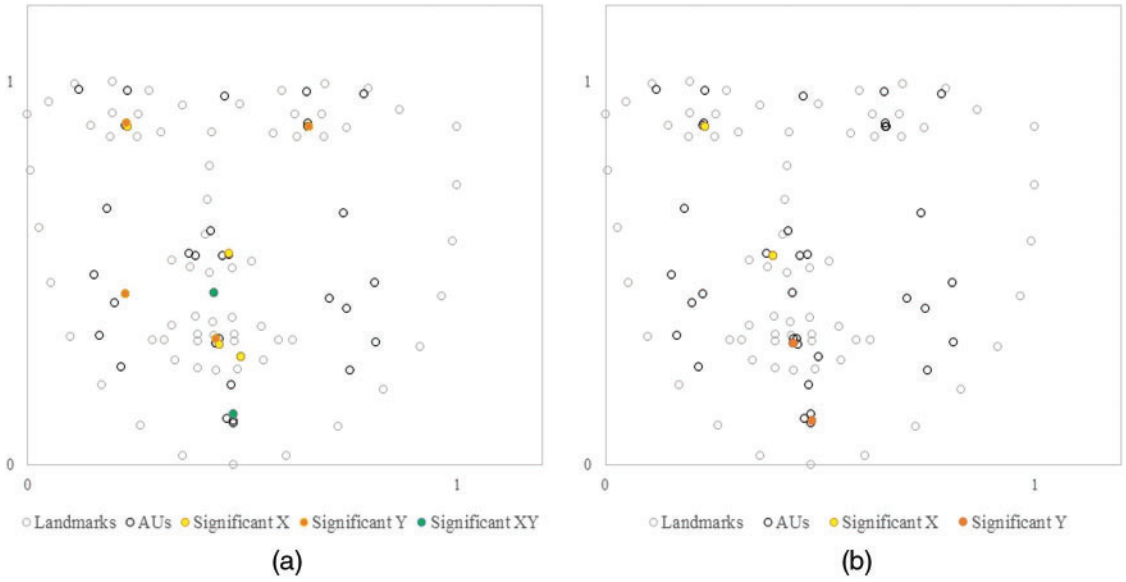

Figure 13: Difference of facial muscle movement. (a) Negative condition (b) Positive condition

Finally, we conducted an independent t-test on the *head pose movement* between the emotional and neutral conditions. The movement indicates the rate of change of the Euler angles. We found no significant difference in both conditions (see Figs. 14 and 15), involving all three-dimensional planes (yaw (X), pitch (Y), and roll (Z)).

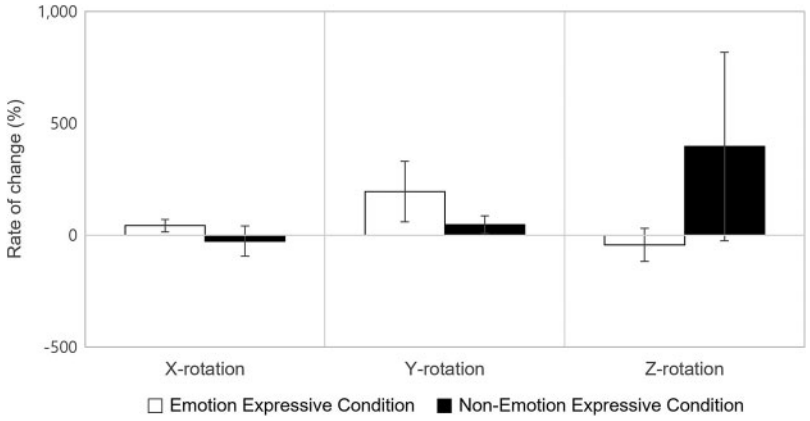

Figure 14: Rate of change in head pose movement in the negative condition

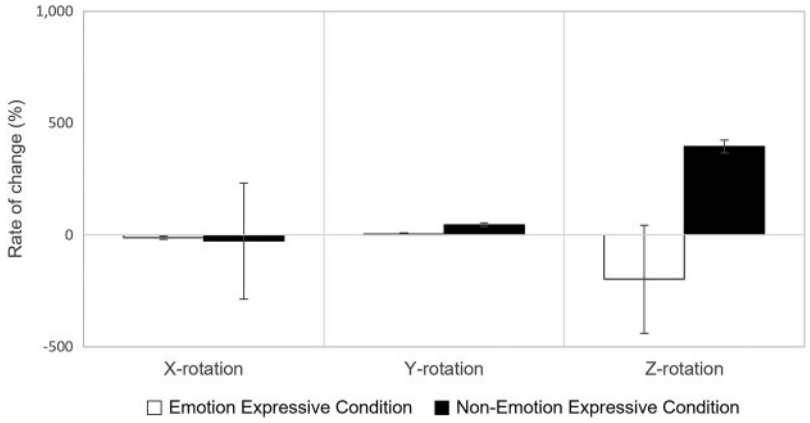

Figure 15: Rate of change in head pose movement in the positive condition

3.3 Facial Muscle Intensity as a Function of the Empathic Capability

We conducted a one-way ANOVA on the *facial muscle intensity* but found no significant difference between the different empathic capabilities in the negative emotion condition. However, in the positive emotion condition, we found a significant difference in the *facial muscle intensity* in AU45 (Blink) ($p < .5$, $F = 3.737$) (see Fig. 16). We conducted a post hoc Games-Howell test and found a significant difference between the low and high empathy ($p < .01$), the mid *vs.* high empathy group ($p < .05$), and the low *vs.* mid empathy group ($p < .05$).

Involving the AU that did not meet normality, we conducted the Kruskal-Wallis H test instead. The results showed a significant difference in AU1 (Inner brow raiser, $p < .01$, $\chi^2 = 9.252$), AU6 (Cheek raiser, $p < .05$, $\chi^2 = 7.686$), AU12 (Lip corner puller, $p < .001$, $\chi^2 = 24.025$), and AU25 (Lips, $p < .05$, $\chi^2 = 6.079$). The post hoc pairwise comparison results showed a significant difference in AU1 between

the low and high empathy group ($p < .01$), in AU6 between the low and high empathy group ($p < .05$), in AU12 between the low and mid empathy group ($p < .05$), and between the low and high empathy group ($p < .001$), and in AU25 between the low and high empathy group ($p < .05$).

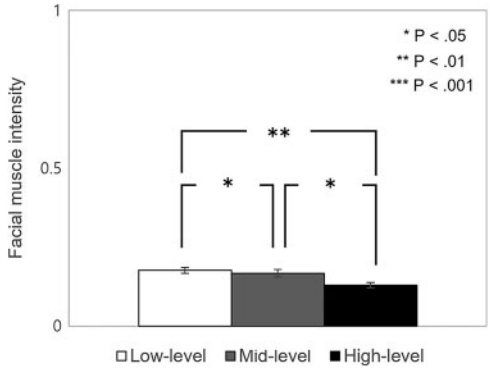

Figure 16: Difference in facial muscle intensity of AU45 between three empathy groups

4 Conclusion and Discussion

This work evaluated whether a participant (observer) can empathize with a virtual agent (target) and exhibit behavioral mimicry accordingly. We designed an interaction context in which the participants viewed the stimuli in the presence of a virtual agent to give a sense of shared experience. We developed a dialog script to build rapport, and had the virtual agent express an emotion congruent (or neutral) with the context via facial expressions, behavioral gestures, and voice. We analyzed the participants' facial expressions during the conversation to validate facial synchronization, which is evidence of facial mimicry.

In summary, when the two dyads (the participant and virtual agent) shared the same emotion (negative or positive) in a shared task (i.e., viewing a video together), we found that the participant mimicked the virtual agent's facial expression if the virtual agent projected an emotion congruent with the emotion of the stimulus. To the best of our knowledge, this is the first study to provide extensive evidence, albeit limited to facial expressions, that humans can exhibit behavioral mimicry to virtual agents in a shared empathic context.

In the negative emotion condition, the level of synchronization of the brows frown, brows down, mouth sad, and mouth narrow in both the left and right was higher in the emotional condition than in the neutral condition. Such expressive elements (i.e., blend shape), with the exception of the nose scrunch, were the elements that the virtual agent utilized to express negative emotion. This implies that the participants were synchronized to nearly all facial regions corresponding to the virtual agent's negative facial expressions. The exception was because the nose scrunch was not prominent in the design, such that the participants were not able to perceive it clearly. The prerequisite of empathizing is a clear recognition by the observer of the emotional cue from the target [77]. Furthermore, the cosine similarity of the brows frown and mouth narrow was more than 0.5, which is regarded as a strong synchronization [74]. Such expressive elements can be used as appraisal criteria for detecting empathy. Future studies may directly test this.

In the positive emotion condition, the level of synchronization of the brows raised, cheek raise, nose scrunch, mouth smile, mouth extension, and mouth opening on both the left and right was higher in the emotional condition than in the neutral condition. Again, such expressive elements were the

elements that the virtual agent utilized to express positive emotions. Furthermore, the cosine similarity of the brows raised, mouth smile, and mouth opening was greater than 0.5.

Additionally, we found that in both emotion conditions (positive and negative), the facial expressions were stronger in the emotional condition than in the neutral condition. That is, all three variables that constitute the facial expression, *facial expressive intensity, facial appearance movement, and facial muscle movement*, were higher in the emotional condition than in the neutral condition.

The results on the *facial expressive intensity* and the expression of the virtual agent had a greater effect on the participants more so in the negative emotion condition than in the positive emotion condition. Specifically, in the negative condition, there was a difference in the expressive intensity between the emotional and neutral virtual agents in all brow-related measures, mouth extension, and narrowing. However, in the positive condition, there was no difference in the expressive intensity in any brow-related measures, mouth extension, and mouth narrowing.

The results of the *facial appearance movement* showed that in the negative condition, the movement related to the brow, eyes, nose, and lips was higher in the emotional condition than in the neutral condition. In the positive emotion condition, the movement of the lower jaw and left cheekbone was higher in the emotional condition than in the neutral condition.

The results of the *facial muscle movement* showed that in the negative condition, the movement of the eyes, nose, mouth, chin, cheekbone, and philtrum was higher in the emotional condition than in the neutral condition. In the positive emotion condition, the movement of the eyes, nose, mouth, and chin was higher in the emotional condition than in the neutral condition.

Finally, we confirmed that there was a difference in the *facial muscle intensity* between the participants in the different empathic capability groups. The intensity of AU45 (Blink) was lower in the higher empathy group and vice versa. AU45 consists of the relaxation of the levator paplebrae and the contraction of the orbicularis oculi. The low intensity of these muscles indicates a low number of eye blinks. This implies that the higher the empathic capability, the more likely it is to engage in an empathic process, resulting in a low number of eye blinks. Future studies may analyze the differential weight of AU45's components. Additionally, there was a significant difference in AU1 (Inner brow raise), AU6 (Cheek raiser), and AU25 (Lips part) between the low and high-level empathy groups. Specifically, in the high-level empathy group, the muscle intensity was significantly higher. The results imply that the high-level empathy group utilized facial muscles such as frontalis, orbicularis oculi, and depressor labbi more so than the low-level empathy group. The low empathy group had a lower use of the zygomatic major than the other two empathy groups.

Empathic responses initiate perceptual information involving the observer's environment to be sent to the superior temporal sulcus (STS). This information is used to determine whether the observer is in danger. This is an unconscious response known as neuroception [78]. If the observer's external environment is perceived as safe, the nucleus ambiguus (NA) becomes activated, suppressing the observer's defense mechanism. The observer is then in a state of social engagement, controlling the facial muscles responsible for pro-social behavior. That is, the observer controls the muscles to hear the target's voice better and orients the observer's gaze toward the target. A more fluid facial expression is now possible [79]. Because the high-empathy group would be more likely to transit to a pro-social state when interacting with the virtual agent, with a more fixed eye gaze, the number of eye blinks found in our experiment would be lower.

In summary, we demonstrated that humans could synchronize their facial expressions with a virtual agent in a shared emotion context with an emotional virtual agent, evident in a significant increment in nearly all dependent measures (movement and intensity) in facial expressions. We also found that such measures differed as a function of the empathy capability. Based on these findings, we suggested two evaluation criteria to assess whether a human user empathizes with the virtual agent:

1. Criterion for facial mimicry: the synchronization level (cosine similarity) is on and above 0.5.
2. Criterion for empathic capability: the *facial muscle intensity* of AU45 is within the following range:
   (1) Low empathy group: $0.18 \pm .05$
   (2) Mid empathy group: $0.16 \pm .08$
   (3) High empathy group: $0.13 \pm .04$

This criterion can be used as a modular assessment tool that can be implemented in any interactive system involving human users to validate whether the user empathizes with the system, including the virtual agent. The application of such modules is significant. The gaming and content media industry is heading toward interactive storytelling based on the viewer's input or response. So far, it has mostly been the user's explicit feedback that drives the story, but with the implementation of an empathic recognition system, it can be a more fluid and seamless interactive storytelling experience.

Since the introduction of voice recognition systems such as Amazon Echo, we have been experiencing AI devices with a mounted camera. For example, Amazon Show has a camera to detect, recognize, and follow the user faces during a dialog. The system can now tap into the user's response in real time, whether the content is congruent with the user's emotional state, and changes the response or service accordingly.

The implications of this study extend to social robots. They typically have a camera mounted on their head for eye contact and indicate their intention through their gaze. The social robot's emotional expression can now be amended according to the facial feed from the camera.

We acknowledge the limitations of this study. For ecological validity, we manipulated the virtual agent's emotional expression at three different levels (facial expression, voice, and behavioral gestures). Although we found an effect, we cannot attribute the results to a single manipulation from the three because of the limitations of the experimental design. Further studies may dissect each manipulation and determine which modality of the virtual agent has a differential impact on the participant's behavioral mimicry.

The study is also limited to statistical test results. We conducted a targeted t-test and ANOVA to validate the hypothesis established from the integrated literature review. However, future studies may train a model with modern methods (machine learning, deep learning, fuzzy logic) [80,81] using the dependent measures (movement and intensity) in facial expressions and output (emotional and neutral). Such an approach may provide weights and parameters that accurately predict the participant's behavioral synchronization. In such a model, the context identification module for an empathic virtual agent is critical because empathy is affected by interaction context and task.

Acknowledgement: Authors thank those who contributed to write this article and give some valuable comments.

Funding Statement: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2020R1A2B5B02002770, Recipient: Whang, M.). URL: https://english.msit.go.kr/eng/index.do.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] K. Glanz, A. S. Rizzo and K. Graap, "Virtual reality for psychotherapy: Current reality and future possibilities," *Psychotherapy: Theory, Research, Practice, Training*, vol. 40, no. 1–2, pp. 55, 2003.

[2] M. Javaid and A. Haleem, "Virtual reality applications toward medical field," *Clinical Epidemiology Global Health*, vol. 8, no. 2, pp. 600–605, 2020.

[3] D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast *et al.,* "Simsensei kiosk: a virtual human interviewer for healthcare decision support," in *Proc. Int. Conf. on Autonomous Agents and Multi-Agent Systems*, Paris, France, pp. 1061–1068, 2014.

[4] P. Kenny, T. Parsons, J. Gratch and A. Rizzo, "Virtual humans for assisted health care," in *Proc. Int. Conf. on Pervasive Technologies Related to Assistive Environments*, Athens, Greece, pp. 1–4, 2008.

[5] W. Ward, R. Cole, D. Bolaños, C. Buchenroth-Martin, E. Svirsky *et al.,* "My science tutor: A conversational multimedia virtual tutor," *Journal Educational Psychology*, vol. 105, no. 4, pp. 1115, 2013.

[6] G. Castellano, A. Paiva, A. Kappas, R. Aylett, H. Hastie *et al.,* "Towards empathic virtual and robotic tutors," in *Proc. Int. Conf. on Artificial Intelligence in Education*, Memphis, TN, USA, pp. 733–736, 2013.

[7] H. Prendinger, J. Mori and M. Ishizuka, "Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game," *International Journal of Human Computer Studies*, vol. 62, no. 2, pp. 231–245, 2005.

[8] K. Zibrek, E. Kokkinara and R. McDonnell, "The effect of realistic appearance of virtual characters in immersive environments-does the character's personality play a role?" *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 4, pp. 1681–1690, 2018.

[9] J. J. Roessingh, A. Toubman, J. van Oijen, G. Poppinga, M. Hou *et al.,* "Machine learning techniques for autonomous agents in military simulations—Multum in parvo," in *Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC)*, Banff, Canada, pp. 3445–3450, 2017.

[10] R. W. Hill Jr, J. Gratch, S. Marsella, J. Rickel, W. R. Swartout *et al.,* "Virtual humans in the mission rehearsal exercise system," *Künstliche Intelligenz*, vol. 17, no. 4, pp. 5, 2003.

[11] T. W. Bickmore, E. Kimani, H. Trinh, A. Pusateri, M. K. Paasche-Orlow *et al.,* "Managing chronic conditions with a smartphone-based conversational virtual agent," in *Proc. Int. Conf. on Intelligent Virtual Agents*, Sydney, Australia, pp. 119–124, 2018.

[12] A. Metallinou, Z. Yang, C. Lee, C. Busso, S. Carnicke *et al.,* "The USC CreativeIT database of multimodal dyadic interactions: From speech and full body motion capture to continuous emotional annotations," *Language Resources and Evaluation*, vol. 50, no. 3, pp. 497–521, 2016.

[13] M. L. Hoffman, *In Empathy and Moral Development: Implications for Caring and Justice*, Cambridge, UK: Cambridge University Press, 2001.

[14] C. D. Batson and L. L. Shaw, "Evidence for altruism: Toward a pluralism of prosocial motives," *Psychological Inquiry*, vol. 2, no. 2, pp. 107–122, 1991.

[15] N. Eisenberg and A. S. Morris, "The origins and social significance of empathy-related responding. a review of empathy and moral development: Implications for caring and justice by ML hoffman," *Social Justice Research*, vol. 14, no. 1, pp. 95–120, 2001.

[16] D. Hume, "A treatise of human nature," *British Moralists*, pp. 1650–1800, 1978.

[17] G. Rizzolatti, L. Fadiga, V. Gallese and L. Fogassi, "Premotor cortex and the recognition of motor actions," *Cognitive Brain Research*, vol. 3, no. 2, pp. 131–141, 1996.

[18] F. De Vignemont and T. Singer, "The empathic brain: How, when and why?" *Trends in Cognitive Sciences*, vol. 10, no. 10, pp. 435–441, 2006.

[19] B. Wicker, C. Keysers, J. Plailly, J. -P. Royet, V. Gallese *et al.,* "Both of us disgusted in My insula: The common neural basis of seeing and feeling disgust," *Neuron*, vol. 40, no. 3, pp. 655–664, 2003.

[20] C. Keysers, B. Wicker, V. Gazzola, J. -L. Anton, L. Fogassi *et al.,* "A touching sight: SII/PV activation during the observation and experience of touch," *Neuron*, vol. 42, no. 2, pp. 335–346, 2004.
[21] T. Singer, B. Seymour, J. O'doherty, H. Kaube, R. J. Dolan *et al.,* "Empathy for pain involves the affective but not sensory components of pain," *Science*, vol. 303, no. 5661, pp. 1157–1162, 2004.
[22] J. Decety and P. L. Jackson, "The functional architecture of human empathy," *Behavioral and Cognitive Neuroscience Reviews*, vol. 3, no. 2, pp. 71–100, 2004.
[23] S. D. Preston and F. B. M. De Waal, "Empathy: Its ultimate and proximate bases," *Behavioral and Brain Sciences*, vol. 25, no. 1, pp. 1–20, 2002.
[24] M. H. Davis, "Measuring individual differences in empathy: Evidence for a multidimensional approach," *Journal of Personality and Social Psychology*, vol. 44, no. 1, pp. 113, 1983.
[25] B. M. P. Cuff, S. J. Brown, L. Taylor and D. J. Howat, "Empathy: A review of the concept," *Emotion Review*, vol. 8, no. 2, pp. 144–153, 2016.
[26] M. L. Hoffman, "Interaction of affect and cognition in empathy," *Emotions*, *Cognition, and Behavior*, pp. 103–131, 1984.
[27] M. H. Davis, *In Empathy: A Social Psychological Approach*, London, UK: Routledge, 2018.
[28] N. Eisenberg, C. L. Shea, G. Carlo and G. P. Knight, "Empathy-related responding and cognition: A 'chicken and the egg' dilemma," *Handbook of Moral Behavior and Development*, vol. 2, pp. 63–88, 2014.
[29] M. L. Hoffman, "Toward a theory of empathic arousal and development," *The Development of Affect*, pp. 227–256, 1978.
[30] E. Hatfield, J. T. Cacioppo and R. L. Rapson, *Emotional contagion: Studies in emotion and social interaction*, Cambridge, UK: Cambridge University Press, 1994.
[31] T. L. Chartrand and J. A. Bargh, "The chameleon effect: The perception–behavior link and social interaction," *Journal of Personality and Social Psychology*, vol. 76, no. 6, pp. 893, 1999.
[32] J. B. Bavelas, A. Black, C. R. Lemery and J. Mullett, "I show how you feel': Motor mimicry as a communicative act," *Journal of Personality and Social Psychology*, vol. 50, no. 2, pp. 322, 1986.
[33] J. N. Bailenson and N. Yee, "Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments," *Psychological Science*, vol. 16, no. 10, pp. 814–819, 2005.
[34] C. Becker, H. Prendinger, M. Ishizuka and I. Wachsmuth, "Evaluating affective feedback of the 3D agent max in a competitive cards game," in *Proc. Int. Conf. on Affective Computing and Intelligent Interaction*, Beijing, China, pp. 466–473, 2005.
[35] S. Brave, C. Nass and K. Hutchinson, "Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent," *International Journal of Human Computer Studies*, vol. 62, no. 2, pp. 161–178, 2005.
[36] S. W. McQuiggan, J. L. Robison, R. Phillips and J. C. Lester, "Modeling parallel and reactive empathy in virtual agents: an inductive approach," in *Proc. Int. Joint Conf. on Autonomous Agents and Multiagent Systems-Volume 1*, Estoril, Portugal, pp. 167–174, 2008.
[37] T. W. Bickmore and R. W. Picard, "Establishing and maintaining long-term human-computer relationships," *ACM Transactions on Computer-Human Interaction*, vol. 12, no. 2, pp. 293–327, 2005.
[38] C. Lisetti, R. Amini, U. Yasavur and N. Rishe, "I can help you change! an empathic virtual agent delivers behavior change health interventions," *ACM Transactions on Management Information Systems*, vol. 4, no. 4, pp. 1–28, 2013.
[39] H. Prendinger and M. Ishizuka, "The empathic companion: A character-based interface that addresses users'affective states," *Applied Artificial Intelligence*, vol. 19, no. 3–4, pp. 267–285, 2005.
[40] M. Ochs, D. Sadek and C. Pelachaud, "A formal model of emotions for an empathic rational dialog agent," *Autonomous Agents and Multi-Agent Systems*, vol. 24, no. 3, pp. 410–440, 2012.
[41] H. Boukricha, I. Wachsmuth, M. N. Carminati and P. Knoeferle, "A computational model of empathy: empirical evaluation," in *Proc. Humaine Association Conf. on Affective Computing and Intelligent Interaction*, Geneva, Switzerland, pp. 1–6, 2013.

[42] S. H. Rodrigues, S. F. Mascarenhas, J. Dias and A. Paiva, "I can feel it too!': emergent empathic reactions between synthetic characters," in *Proc. Affective Computing and Intelligent Interaction and Workshops*, Cambridge, UK, pp. 1–7, 2009.
[43] T. W. Bickmore and R. W. Picard, "Establishing and maintaining long-term human-computer relationships," *ACM Transactions on Computer-Human Interaction*, vol. 12, no. 2, pp. 293–327, 2005.
[44] S. H. Rodrigues, S. Mascarenhas, J. Dias and A. Paiva, "A process model of empathy for virtual agents," *Interacting with Computers*, vol. 27, no. 4, pp. 371–391, 2015.
[45] K. Kodama, S. Tanaka, D. Shimizu, K. Hori, H. Matsui *et al.,* "Heart rate synchrony in psychological counseling: A case study," *Psychology*, vol. 9, no. 7, pp. 1858, 2018.
[46] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161, 1980.
[47] A. C. Samson, S. D. Kreibig, B. Soderstrom, A. A. Wade and J. J. Gross, "Eliciting positive, negative and mixed emotional states: A film library for affective scientists," *Cognition and Emotion*, vol. 30, no. 5, pp. 827–856, 2016.
[48] S. Baron-Cohen and S. Wheelwright, "The empathy quotient: An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences," *Journal of Autism and Developmental Disorder*, vol. 34, no. 2, pp. 163–175, 2004.
[49] R. Hogan, "Development of an empathy scale," *Journal of Consulting and Clinical Psychology*, vol. 33, no. 3, pp. 307, 1969.
[50] A. Mehrabian and N. Epstein, "A measure of emotional empathy," *Journal of Personality*, vol. 40, no. 4, pp. 525–543, 1972.
[51] A. Mehrabian, A. L. Young and S. Sato, "Emotional empathy and associated individual differences," *Current Psychology*, vol. 7, no. 3, pp. 221–240, 1988.
[52] J. A. Simpson, "Psychological foundations of trust," *Current Directions in Psychological Science*, vol. 16, no. 5, pp. 264–268, 2007.
[53] J. -H. Heo and C. -J. Lee, "Psychometric analysis of the empathy quotient (EQ) scale," *Humanitity Science*, vol. 24, pp. 183–200, 2010.
[54] J. -H. Heo and C. -J. Lee, "The effects of empathy on self-esteem and subjective well-being," *The Journal of the Acoustical Society of Korea*, vol. 29, no. 5, pp. 332–338, 2010.
[55] H. Adam, A. Shirako and W. W. Maddux, "Cultural variance in the interpersonal effects of anger in negotiations," *Psychological Science*, vol. 21, no. 6, pp. 882–889, 2010.
[56] H. Adam and A. Shirako, "Not all anger is created equal: The impact of the expresser's culture on the social effects of anger in negotiations," *Journal of Applied Psychology*, vol. 98, no. 5, pp. 785, 2013.
[57] D. Matsumoto and S. H. Yoo, "Toward a new generation of cross-cultural research," *Perspectives on Psychological Science*, vol. 1, no. 3, pp. 234–250, 2006.
[58] D. Cohen and A. Gunz, "As seen by the other: Perspectives on the self in the memories and emotional perceptions of easterners and westerners," *Psychological Science*, vol. 13, no. 1, pp. 55–59, 2002.
[59] A. L. Baylor and Y. Kim, "Pedagogical agent design: the impact of agent realism, gender, ethnicity, and instructional role," in *Proc. Int. Conf. on Intelligent Tutoring Systems*, pp. 592–603, 2004.
[60] J. A. Pratt, K. Hauser, Z. Ugray and O. Patterson, "Looking at human–computer interface design: Effects of ethnicity in computer agents," *Interacting with Computers*, vol. 19, no. 4, pp. 512–523, 2007.
[61] C. Nass, K. Isbister and E. -J. Lee, "Truth is beauty: Researching embodied conversational agents," *Embodied Conversational Agents*, pp. 374–402, 2000.
[62] P. Ekman and W. V. Friesen, "*M anual of the Facial Action Coding System (FACS)*," Consulting Psychologists Press: Palo Alto, CA, USA, 1978.
[63] P. Collett, *In the Book of Tells: how to Read People's Minds from Their Actions*, New York, NY, USA: Random House, 2004.
[64] A. Nelson and S. K. Golant, *In You Don't say: Navigating Nonverbal Communication Between the Sexes*, Hoboken, NJ, USA: Prentice Hall, 2004.

[65] B. Pease and A. Pease, *In the Definitive Book of Body Language: the Hidden Meaning Behind People's Gestures and Expressions*, New York, NY, USA: Bantam, 2008.
[66] S. W. McQuiggan and J. C. Lester, "Modeling and evaluating empathy in embodied companion agents," *International Journal of Human-Computer Studies*, vol. 65, no. 4, pp. 348–360, 2007.
[67] A. Paiva, I. Leite, H. Boukricha and I. Wachsmuth, "Empathy in virtual agents and robots: A survey," *ACM Transactions on Interactive Intelligent Systems*, vol. 7, no. 3, pp. 1–40, 2017.
[68] N. Dahlbäck, A. Jönsson and L. Ahrenberg, "Wizard of Oz studies—why and how," *Knowledge-based Systems*, vol. 6, no. 4, pp. 258–266, 1993.
[69] T. Baltrušaitis, P. Robinson and L. -P. Morency, "Openface: an open source facial behavior analysis toolkit," in *Proc. IEEE Winter Conf. on Applications of Computer Vision*, Lake Placid, NY, USA, pp. 1–10, 2016.
[70] J. L. Tracy and D. Matsumoto, "The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays," *Proc. National Academy of Sciences*, vol. 105, no. 33, pp. 11655–11660, 2008.
[71] Z. Hammal, J. F. Cohn, C. Heike and M. L. Speltz, "What can head and facial movements convey about positive and negative affect?" in *Proc. Int. Conf. on Affective Computing and Intelligent Interaction*, Xian, China, pp. 281–287, 2015.
[72] A. Adams, M. Mahmoud, T. Baltrušaitis and P. Robinson, "Decoupling facial expressions and head motions in complex emotions," in *Proc. Int. Conf. on Affective Computing and Intelligent Interaction*, Xian, China, pp. 274–280, 2015.
[73] D. C. Howell, M. Rogier, V. Yzerbyt and Y. Bestgen, *Statistical Methods in Human Sciences*, New York, NY: Wadsworth, 1998.
[74] H. V. Nguyen and L. Bai, "Cosine similarity metric learning for face verification," in *Proc. Asian Conf. on Computer Vision*, Queenstown, New Zealand, pp. 709–720, 2010.
[75] P. G. Ipeirotis, L. Gravano and M. Sahami, "Probe, count, and classify: categorizing hidden web databases," in *Proc. ACM SIGMOD Int. Conf. on Management of Data*, Santa Barbara, California, USA, pp. 67–78, 2001.
[76] N. Koudas, A. Marathe and D. Srivastava, "Flexible string matching against large databases in practice," in *Proc. Int. Conf. on Very Large Data Bases*, Toronto, Canada, pp. 1078–1086, 2004.
[77] J. Zaki, N. Bolger and K. Ochsner, "It takes two: The interpersonal nature of empathic accuracy," *Psychological Science*, vol. 19, no. 4, pp. 399–404, 2008.
[78] S. W. Porges, "Neuroception: A subconscious system for detecting threats and safety," *Zero to Three*, vol. 24, no. 5, pp. 19–24, 2004.
[79] S. W. Porges, "The polyvagal theory: Phylogenetic contributions to social behavior," *Physiology & Behavior*, vol. 79, no. 3, pp. 503–513, 2003.
[80] Mustaqeem and S. Kwon, "Optimal feature selection based speech emotion recognition using twostream deep convolutional neural network," *International Journal of Intelligent Systems*, vol. 36, pp. 5116–5135, 2021.
[81] S. Kwon, "Optimal feature selection based speech emotion recognition using two-stream deep convolutional neural network," *International Journal of Intelligent Systems*, 2021.