2026 IEEE/SICE International Symposium on System Integration (SII)
January 11-14, 2026. Cancún, México

Emotion Recognition of Avatar Facial Expressions Generated Using Action Units Based on Plutchik's Emotion Model

Nanami Ishizu[1], Rei Taguchi[1], Wen Liang Yeoh[1], and Osamu Fukuda[1]

Abstract— **With conversational artificial intelligence (AI) becoming increasingly present in our daily lives, some people have begun to view it as a friend, rather than merely a tool. Among young people in particular, the use of conversational AIs as confidants is becoming more widespread, signaling the emergence of a new form of interaction between humans and AI. However, current conversational AIs are limited to text-based communication and lack the nonverbal cues found in human communication, such as facial expressions, intonation, and gestures. This lack of nonverbal cues raises concerns that the user might miss the nuanced richness of meaning that is intended to be conveyed. To expand the expressive capabilities of future conversational AIs, we propose an avatar that conveys facial expressions, which play an important role in conveying emotions. Based on Plutchik's emotion model, the avatar can display eight primary facial expressions and composite expressions by combining them. We experimentally investigated how participants perceive the expressions of the proposed avatar. This experiment revealed that anger, fear, disgust, and trust are difficult to recognize. Additionally, the results suggest that, even when male and female avatars express the same emotions, they may be perceived slightly differently.**

I. INTRODUCTION

Conversational artificial intelligence (AI) based on large language models, such as ChatGPT [1] and Gemini [2], has been increasingly integrated into our daily lives. According to a survey conducted by Nomura Research Institute Ltd. in September 2024, the awareness rate of ChatGPT in Japan was 72.2%, while the usage rate of it was 20.4% [3]. The survey also reported that the primary uses of ChatGPT were writing, gathering information, and summarizing sentences. These indicate conversational AI's widespread use as a tool. On the other hand, a June 2025 survey by DENTSU Inc. of Japanese people aged 12 to 69 found that 64.9% shared their feelings with conversational AI [4], compared to that of close friends (64.6%) and mothers (62.7%). This rate was even higher among younger people, with over 70% reporting that they could share their feelings with a conversational AI. This survey suggests that some people are beginning to treat AI like a friend. The spread of conversational AI may create a new interaction between humans and conventional AI, which has up to now been seen as a tool.

*This research was funded by JSPS KAKENHI (grant number 23H03440).

[1]Nanami Ishizu is with Faculty of Sciences and Engineering, Saga University, 1 Honjo, Saga, Japan 24742101@edu.cc.saga-u.ac.jp

[1]Rei Taguchi is with Faculty of Science and Engineering, Saga University, 1 Honjo, Saga, Japan 22238245@cc.saga-u.ac.jp

[1]Wen Liang Yeoh is with Faculty of Science and Engineering, Saga University, 1 Honjo, Saga, Japan wlyeoh@cc.saga-u.ac.jp

[1]Osamu Fukuda is with Faculty of Science and Engineering, Saga University, 1 Honjo, Saga, Japan fukudao@cc.saga-u.ac.jp

Due to the increasing number of people sharing their worries with conversational AI, there is increasing interest in using conversational AI for psychological support and counseling [5], [6]. However, there is no definitive evidence to show that conversational AI is effective as a counselor [7]. Despite this, interacting with conversational AI could have detrimental effects. In February 2024, a child who had been talking about his worries to a conversational AI has ended up taking his own life [8]. A possible reason for this is that current conversational AI is limited to text-based communication. In human conversation, nonverbal cues such as facial expressions, intonation, and gestures are crucial. However, text-based communication lacks such nonverbal cues, and this difference has been found to affect the transmission of information [9]. With conversational AI increasingly becoming a confidant for humans, it may be beneficial for current conversational AI to be able to express nonverbal cues.

Human facial expressions play an important role in conveying emotions [10]. We are focusing on facial expressions and working to create avatars capable of expressing natural human-like expressions as part of developing a new form of interactive AI. A common approach to representing emotions in avatars is the expression control method, which is based on combinations of Action Units (AUs) in the Facial Action Coding System (FACS) [11]. FACS is a system that expresses facial expressions using combinations of units called AUs. AUs are defined based on the movements of a facial muscle or muscle groups that produce the expression of an emotion. For example, the movement of the muscles that raise the corners of the lips is defined as AU12. Ekman et al. (1969, 1971) identified six emotions that people display universal facial expressions across cultures: anger, disgust, fear, joy, sadness, and surprise [12], [13], and these same emotions could be expressed by a combination of AUs [14]. Since then, many previous studies have controlled avatar expressions based on these six basic emotions [15].

On the other hand, Plutchik proposes eight primary human emotions: joy, trust, fear, surprise, sadness, disgust, anger, and anticipation [16]. This emotion model is classified based on the psychological structure of humans and has been frequently used in research to estimate the emotions contained in text [17], [18]; however, there has been very little research extending this model to human facial expressions. Park (2025) analyzed the relationship between emotions and facial muscle movements based on Plutchik's eight basic emotion model, but did not conduct empirical user evaluations on the appropriateness of the proposed expressions [19].

978-1-6654-5784-2/26/$31.00 ©2026 IEEE

In this paper, we control the face of the avatar based on Plutchik's eight primary emotions using AUs and experimentally investigate how humans perceive the expressions of the proposed avatars.

II. Control of the Face of the Avatar Using AU

MetaHuman [20], an avatar developed by Epic Games whose facial expression could be controlled using AUs, was used in this paper. As illustrated in Fig. 1, the Facial Control Rig, which is the facial expression controller of MetaHuman, performs analogous functions to AUs defined by Ekman. By specifying the Facial Control Rig value as a floating-point number between 0.0 and 1.0, or -1.0 and 1.0, you can change the facial expressions of MetaHuman along the AU.

We controlled the face of the avatar along Plutchik's model of eight primary emotions, expanding on Ekman's six basic emotions by adding two additional emotional states: anticipation and trust. Figure 2 shows the eight facial expressions displayed by the avatar and the AUs used to control them. AUs, which were used for facial expressions controlled based on Ekman's six basic emotions, were referenced from previous research by Vicente-Querol et al [21]. Since anticipation and trust are not defined in Ekman's basic six emotions, we referred to Park's previous research [19]. Anticipation and trust were expressed using AUs corresponding to the muscles discussed by him. Table I shows the correspondence between facial muscles used to display anticipation and trust described in Park's study and the AUs used in our research.

In general, our face expresses not only eight primary emotions but also many other emotions. To make it possible for an avatar to express complex facial expressions like humans, we propose composite facial expressions that are expressed by combining the eight emotions. Note that the composite facial expressions in this paper indicate the facial expression by linearly adding the AUs used in the eight emotions. Figure 3 shows examples of the composite facial expressions displayed by the avatar.

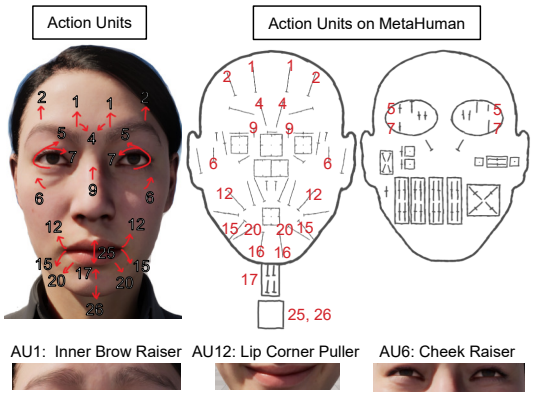


Fig. 1. Action Units and MetaHuman Facial Control Rig

III. Methods

A. Participants

Ten students from Saga University participated in the experiment (nine males, one female, mean age 23.0 ± 0.4).

TABLE I
Facial Muscles and AU to Trust and Anticipation

Trust

Facial Muscles	Action Unit
Frontails	AU1, AU2
Procerus	-
Orbicularis Oculi	AU6, AU7
Zygomaticus Major	AU12
Levator Anguli Oris	AU13

Anticipation

Facial Muscles	Action Unit
Procerus	-
Orbicularis Oculi	AU6, AU7
Zygomaticus Major	AU12
Levator Anguli Oris	AU13
Risorius	AU20

This experiment was conducted with the approval of the Ethics Committee of Saga University School of Medicine (approval number R7-4).

B. Task

Participants were asked to answer questions about the avatar displayed on the monitor, which emotion the avatar expressed, and how strong the expression was, using the user interface shown in Fig. 4. The question number appeared at the top of the user interface, and the target avatar face image was displayed below it. Participants selected facial expressions and their intensity by moving the Answer Slider, and submitted their answers by pressing the Confirm Button. Once a participant had submitted their answers, automatically proceeded to the next question. Participants had the option to press the Reset Button to change the facial expression and intensity they had selected for a question.

C. Conditions

To investigate how well participants recognized the avatar's facial expressions, two conditions were arranged: one with a single expression and another with a composite expression. Under the condition of a single expression, the avatar displayed one of the nine emotions: Joy, Sadness, Anticipation, Surprise, Anger, Fear, Disgust, Trust, or Neutral. Each emotion, except for Neutral, had three levels of intensities, which were low, middle, and high. A total of 25 avatar facial expressions were shown to the participants in the condition of a single expression. Under the condition of a composite expression, the avatar displayed a combination of two expressions from the eight expressions: Joy, Sadness, Anticipation, Surprise, Anger, Fear, Disgust, and Trust. There were three kinds of intensity for each combination: low, medium, and high. In total, 84 avatar facial expressions were presented to the participants as a composite expression.

We prepared male and female avatars for both single and composite expressions to explore whether the gender of the avatar influences participants' recognition of the avatar's facial expressions. The same AUs controlled the facial expressions of both male and female avatars.

Fig. 2. Basic eight emotions and Action Unit

Fig. 3. Examples of composite expressions

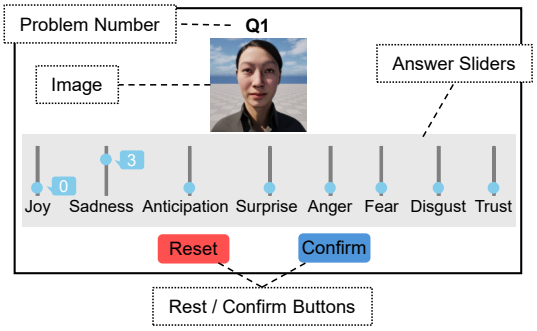


Fig. 4. User interface used in the experiment

D. Protocol

The experiment was conducted on an online web page that we built on Amazon Web Services servers. First, the participants answered practice questions to understand how to use the user interface. Next, the participants were asked to answer questions under the condition that the male and female avatars displayed single expressions. After allowing at least 12 hours after the first session, they were asked to answer questions under the condition that the male and female avatars displayed composite expressions. Throughout the experiment, participants were not provided with the correctness of their answers.

E. Analysis

1) Recognition rate: The recognition rate, derived from (1), indicates how accurately the facial expressions of the avatars are recognized by the participants under single or composite conditions.

$$recognition\ rate[\%] = \frac{P_{correct}}{n} \times 100 \qquad (1)$$

$P_{correct}$ is the correct answer rate for any emotion, and n is the number of specific facial expressions included in the problem set.

2) Recognizability rate: The recognizability rate, derived from (2), is the percentage of times the expression was presented when a participant responded with that expression.

$$recognizability\ rate[\%] = \frac{P_{correct}}{P_{select}} \times 100 \qquad (2)$$

$P_{correct}$ is the number of correct answers for a given expression, and P_{select} is the number of times the participant responded with a certain facial expression.

For example, if expression A appears twice in a set of questions and a participant correctly answers one of them, the recognition rate is 50%, while the recognizability rate is 100%.

For responses with composite expressions, evaluate each expression independently. For example, if a composite expression of joy and anticipation is displayed, and the participant answers joy and surprise, then joy is correct and surprise is incorrect.

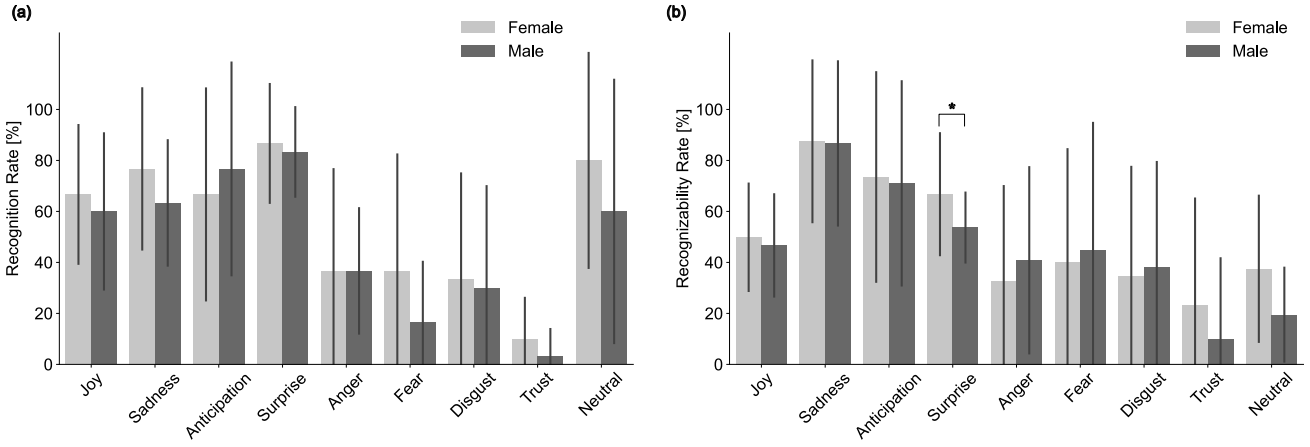
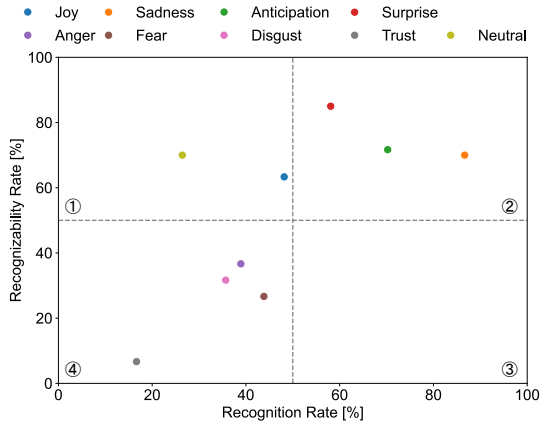


Fig. 5. (a)recognition Rate (mean ± SD) and (b)Recognizability Rate (mean ± SD) for single expressions $*$: $p < 0.05$



① it's often overlooked, but it's the correct answer when participants respond.
② it's rarely overlooked when asked, and mostly answered correctly.
③ participants tended to answer regardless of whether it was displayed.
④ it's often overlooked and is often answered incorrectly.

Fig. 6. Recognition rate vs recognizability rate for single expressions

IV. RESULTS

A. Single expression

Figure 5(a) shows the recognition rates for single expressions of male and female avatars. Higher recognition rates were shown for the expressions Joy, Sadness, Anticipation, Surprise, and Neutral for both male and female avatars compared to other expressions. It can be said that the expressions Joy, Sadness, Anticipation, Surprise, and Neutral were recognized relatively accurately. A t-test was performed to compare the recognition rate of each facial expression against male and female avatars. No significant difference was found between male and female avatars for any facial expression (Joy: $t(9) = 1.50$, $p = 0.17$, Sadness: $t(9) = 1.81$, $p = 0.10$, Anticipation: $t(9) = 1.41$, $p = 0.19$, Surprise: $t(9) = 0.38$, $p = 0.76$, Anger: $t(9) = 0.00$, $p = 1.00$, Fear: $t(9) = 1.62$, $p = 0.14$, Disgust: $t(9) = 0.36$, $p = 0.72$, Trust: $t(9) = 1.50$, $p = 0.17$, Neutral: $t(9) = 1.00$, $p = 0.34$). It can be said that the appearance of avatars has little effect on the facial expression recognition rate.

Figure 5(b) shows the recognizability rate for single expressions of male and female avatars. Higher recognizability rates were shown for the expressions Sadness, Anticipation, and Surprise for both male and female avatars compared to other expressions. It can be said that the expressions Sadness, Anticipation, and Surprise were relatively difficult to misinterpret. A t-test was performed to compare the recognizability rate of each facial expression against male and female avatars. The recognizability rate of Surprise for the female avatar was significantly higher than for the male avatar ($t(9) = 2.29$, $p = 0.04$). No significant differences were found between the male and female avatars in expressions other than Surprise. The appearance of the avatar has little effect on the likelihood of misinterpreting facial expressions, except for Surprise.

Figure 6 shows a scatter plot of the recognition rate and recognizability rate for a single expression, indicating how each expression was recognized. A facial expression plotted on the upper left of the scatter plot (①) means that it is often overlooked, but it is the correct answer when participants respond. A facial expression plotted on the upper right of the scatter plot (②) means that it is rarely overlooked when asked, and mostly answered correctly. A facial expression plotted on the lower right of the scatter plot (③) means that participants tended to answer regardless of whether it was displayed. In other words, the participants answer by taking a wild guess. A facial expression plotted on the lower left of the scatter plot (④) means that it is often overlooked and is often answered incorrectly. Therefore, it can be said that Joy, Anticipation, and Sadness were expressions that were easily recognized by the participants without being overlooked. On the other hand, it can be said that Neutral and Joy were expressions that were easily overlooked by the participants. Other expressions were often overlooked by the participants, suggesting that they were easily mistaken for other expressions.

B. Composite expression

Figure 7(a) shows the recognition rates for composite expressions of male and female avatars. The recognition rates were less than 60% for all expressions except the male avatar's Anticipation expression. This suggests that the

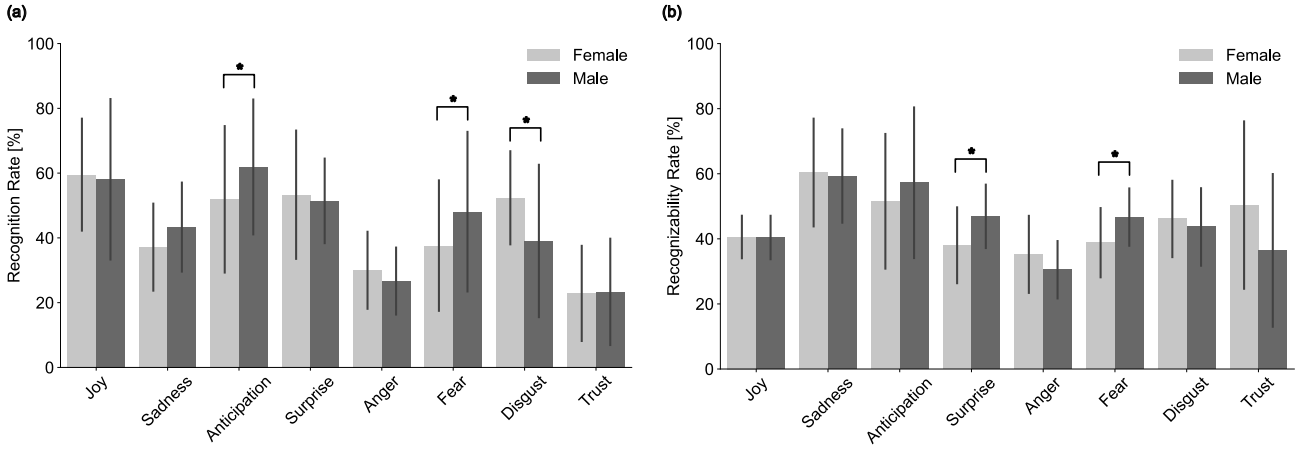
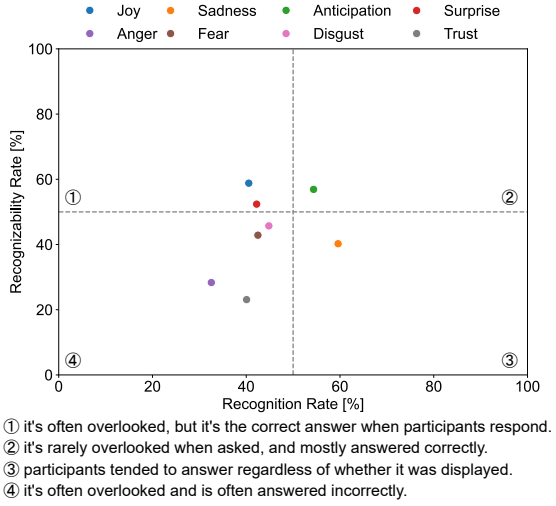


Fig. 7. (a)recognition rates (mean ± SD) and (b)recognizability rate(mean ± SD) for composite expressions $*$: $p < 0.05$



① it's often overlooked, but it's the correct answer when participants respond.
② it's rarely overlooked when asked, and mostly answered correctly.
③ participants tended to answer regardless of whether it was displayed.
④ it's often overlooked and is often answered incorrectly.

Fig. 8. Recognition rate vs recognizability rate for composite expressions

combination of expressions made recognition difficult. A t-test was performed to compare the recognition rate of each facial expression against male and female avatars. Significant differences were found between male and female avatars in terms of Anticipation, Fear, and Disgust ($t(9) = 2.60$, $p = 0.03$; $t(9) = 2.96$, $p = 0.02$; $t(9) = 2.28$, $p = 0.04$). No significant differences were found in the case of single expressions, but such significant differences were found in the case of composite expressions.

Figure 7(b) shows the recognizability rates for composite expressions of male and female avatars. Regardless of the gender of the avatar, the recognizability rates for Joy, Sadness, Anticipation, and Surprise were lower than for single expressions. These three expressions tend to be easily misinterpreted when combined. On the other hand, the recognizability rates for Trust and Disgust increased compared to single expressions. These two tended to be misinterpreted when combined with other expressions.

A t-test was performed to compare the recognizability rate of each facial expression against male and female avatars. The recognizability rates of Surprise and Fear for the male avatar was significantly higher than for the female avatar ($t(9) = 2.57$, $p = 0.03$; $t(9) = 2.75$, $p = 0.02$). Surprise showed a significant difference in recognizability rate between male and female even for single expression (Fig 5(b)). This suggests that the appearance of the avatar may affect the recognizability rate of Surprise. Both the recognition and recognizability rate of Fear showed significant differences between male and female avatars, suggesting the appearance of the avatar may affect.

Figure 8 shows a scatter plot of the recognition rate and recognizability rate for a composite expression, indicating how each expression was recognized. The numbers in Fig. 8 mean the same as those in Fig. 6. Anticipation was an expression that was easy for participants to recognize and difficult to overlook. Joy and Surprise were often overlooked, but these expressions were correctly recognized when participants responded. Although sadness was often recognized correctly, it could also be said that other expressions were often mistaken for sadness. Other expressions were often overlooked by the participants, suggesting that they often gave incorrect answers when they chose the expressions.

V. DISCUSSION

A. Expressions with low recognition rate in single expression

As shown in Fig. 5(a), the recognition rates for Anger, Fear, and Disgust were lower than those for other expressions. Ekman (1961) stated that expressions of anger, disgust, fear, joy, sadness, and surprise are universal, regardless of cultural background. However, it was reported that these expressions were not recognized as accurately in Asian cultures as in Western cultures [22]. In addition, it was reported that Japanese people express facial expressions that differ from the universal expressions described by Ekman, except for Joy and Surprise [23]. Sato et al. (2019) found that expressions of anger are often mistaken for disgust, sadness, or neutral, expressions of disgust are often mistaken for anger or neutral, and expressions of fear are often mistaken for anger, disgust, sadness, or neutral by analyzing Japanese facial expressions based on Ekman's basic emotions. Since

this experiment was conducted with Japanese participants, the recognition rate of Anger, Fear, and Disgust may have been low. In fact, the participants tended to misinterpret the avatar's Anger expression as Disgust or Neutral, Fear expression as Surprise, and Disgust expression as Anger.

The recognition rate for Trust was also low compared to other expressions, and there was a tendency to misidentify it as Joy. As mentioned earlier, it is known that the expressions of joy displayed by Japanese people are similar to the basic expressions described by Ekman [23]. The reason for the tendency to misidentify Trust as Joy may be that the AU that constitutes Joy is part of the AU that constitutes Trust.

B. Increasement of recognizability in composite expression

When avatars displayed composite expressions, the recognizability rates for Trust and Disgust improved compared to when they displayed single expressions. This suggests that combining expressions may make it easier to recognize those that were easily misinterpreted as others. The recognizability rate of expressions that had been misidentified as others tended to improve when expressions of the same polarity were combined (e.g., positive and positive, or negative and negative) compared to a single expression. Combining expressions of the same polarity makes it possible to give avatars rich, easy-to-understand expressions.

C. Impression of avatar's gender gives to participants

The results of the experiment suggest that expressions are perceived differently and with different levels of accuracy depending on the gender of the avatar. Although no statistically significant differences were found, neutral displayed by the female avatar tended to be misinterpreted as anger, while that by the male avatar tended to be misinterpreted as sadness. These results suggest that the gender of avatars may give participants different impressions. However, the number of participants in this experiment was only 10, and the majority were male, so it is necessary to increase the number of participants considering the male-female ratio of the participants in order to conduct a more detailed analysis.

VI. CONCLUSIONS

In this study, we proposed controlling avatar facial expressions by AUs based on Plutchik's basic emotions. We also proposed a composite expression by adding AUs linearly. We conducted an experiment to verify whether the interpretation of human expressions changes depending on the single or composite expressions displayed by the proposal avatar and its gender. The results of the experiment revealed that some expressions are easier for humans to recognize than others. This may be because the experiment was conducted with Japanese participants. This suggests that differences from Western culture, where AU was proposed, may have influenced the results. There tended to be no effect on recognition based on the gender of the avatar in the case of single expressions. However, an effect was observed in the case of complex expressions. Participants may have different impressions based on the gender of the avatar.

REFERENCES

[1] "ChatGPT." https://chatgpt.com/. [Accessed 12-08-2025].

[2] "Google Gemini." https://gemini.google.com/app. [Accessed 12-08-2025].

[3] Nomura Research Institute Ltd., "Nihon no ChatGPT riyodoko (in Japanese)." https://www.nri.com/jp/knowledge/report/20241016_1.html, 2024. [Accessed 01-08-2025].

[4] DENTSU. INC., "Taiwagata ai tono kankeisei ni kansuru ishikityosa (in Japanese)." https://www.dentsu.co.jp/news/release/2025/0703-010908.html, 2025. [Accessed 01-08-2025].

[5] A. A. Abd-alrazaq, M. Alajlani, A. A. Alalwan, B. M. Bewick, P. Gardner, and M. Househ, "An overview of the features of chatbots in mental health: A scoping review," *International Journal of Medical Informatics*, vol. 132, p. 103978, 2019.

[6] A. S. Miner, A. Milstein, and J. T. Hancock, "Talking to machines about personal mental health problems," *JAMA*, vol. 318, pp. 1217–1218, 10 2017.

[7] A. A. Abd-Alrazaq, A. Rababeh, M. Alajlani, B. M. Bewick, and M. Househ, "Effectiveness and safety of using chatbots to improve mental health: Systematic review and meta-analysis," *J Med Internet Res*, vol. 22, p. e16021, Jul 2020.

[8] C. Duffy, " 'There are no guardrails.' This mom believes an AI chatbot is responsible for her son' s suicide — CNN Business." https://edition.cnn.com/2024/10/30/tech/teen-suicide-character-ai-lawsuit. [Accessed 02-08-2025].

[9] D. Archer and R. M. Akert, "Words and everything else: Verbal and nonverbal cues in social interpretation.," *Journal of Personality and Social Psychology*, vol. 35, p. 443–449, June 1977.

[10] A. Mehrabian, *Nonverbal communication*. Routledge, 2017.

[11] P. Ekman, W. V. Friesen, and J. C. Hager, *The Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, 2002.

[12] P. Ekman and W. V. Friesen, "The repertoire of nonverbal behavior: Categories, origins, usage, and coding," *semiotica*, vol. 1, no. 1, pp. 49–98, 1969.

[13] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion.," *Journal of personality and social psychology*, vol. 17, no. 2, p. 124, 1971.

[14] W. V. Friesen, P. Ekman, *et al.*, "Emfacs-7: Emotional facial action coding system," *Unpublished manuscript, University of California at San Francisco*, vol. 2, no. 36, p. 1, 1983.

[15] V. Bettadapura, "Face expression recognition and analysis: the state of the art," *arXiv preprint arXiv:1203.6722*, 2012.

[16] R. Plutchik, *Emotion, a Psychoevolutionary Synthesis*. Harper & Row, 1980.

[17] S. Kusal, S. Patil, J. Choudrie, K. Kotecha, D. Vora, and I. Pappas, "A review on text-based emotion detection–techniques, applications, datasets, and future directions," *arXiv preprint arXiv:2205.03235*, 2022.

[18] A. Seyeditabari, N. Tabari, and W. Zadrozny, "Emotion detection in text: a review," *arXiv preprint arXiv:1806.00674*, 2018.

[19] ParkHwajin, "A study on facial expression design guidelines for digital human modeling - focusing on plutchik's eight primary emotions - ," *International Journal of Advanced Culture Technology*, vol. 13, pp. 201–217, 03 2025.

[20] "MetaHuman — High-Fidelity Digital Humans Made Easy." https://www.metahuman.com/en-US. [Accessed 02-08-2025].

[21] M. Á. Vicente-Querol, A. S. García, P. Fernández-Sotos, R. Rodriguez-Jimenez, and A. Fernández-Caballero, "Development and validation of basic virtual human facial emotion expressions," in *Understanding the Brain Function and Emotions* (J. M. Ferrández Vicente, J. R. Álvarez-Sánchez, F. de la Paz López, J. Toledo Moreo, and H. Adeli, eds.), (Cham), pp. 222–231, Springer International Publishing, 2019.

[22] C. Chen, L. B. Hensel, Y. Duan, R. A. A. Ince, O. G. B. Garrod, J. Beskow, R. E. Jack, and P. G. Schyns, "Equipping social robots with culturally-sensitive facial expressions of emotion using data-driven methods," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, pp. 1–8, 2019.

[23] W. Sato, S. Hyniewska, K. Minemoto, and S. Yoshikawa, "Facial expressions of basic emotions in japanese laypeople," *Frontiers in Psychology*, vol. Volume 10 - 2019, 2019.