This article has been accepted for publication in IEEE Consumer Electronics Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCE.2025.3544816

Toward Believable Emotions: Evaluating FACS Coding for Virtual Human Expressions

Stefano Calzolari, Francesco Strada, Andrea Bottino
Politecnico di Torino

***Abstract*—In interactive computer graphics, the Facial Action Coding System (FACS) is widely used to enhance the emotional expressiveness of Virtual Humans (VHs). By linking specific Action Units (AUs) with facial blendshapes, animators can theoretically reproduce a wide range of human emotions. However, achieving realistic and believable emotional expressions remains challenging, as the same AU intensities do not work equally well across all VHs. This paper explores whether optimal sets of AU intensities can be defined for specific subgroups of VHs, such as those differentiated by gender and visual fidelity, rather than pursuing a one-size-fits-all approach. Through a focused analysis of happiness, sadness, and disgust, we demonstrate that visual fidelity plays a critical role in emotional clarity, while certain emotions require gender-specific adjustments. The findings emphasize the limitations of uniformly maximizing AU intensities across all VHs and offer practical insights for animators, providing a nuanced framework for creating believable emotional expressions in various types of VHs and enhancing realism in interactive applications.**

In the rapidly evolving field of consumer electronics, Virtual Humans (VHs) are increasingly embedded in a wide range of technologies, from virtual assistants and home automation interfaces to gaming, Virtual Reality (VR) and healthcare devices [1], [2], [3]. No longer limited to entertainment or gaming, VHs are now playing a critical role in consumer electronics products designed for home care, fitness and education. Whether integrated into smart home systems, personal wellness devices or VR platforms, VHs provide interactive and personalized experiences that enhance the functionality and user engagement of modern consumer electronics.

The emotional expressiveness of VHs is critical to create believable and meaningful interactions in these devices and applications. The ability of VHs to convey emotion helps users connect more deeply with these virtual characters, enhancing their expe-

Digital Object Identifier 10.1109/MCE.YYYY.Doi Number
Date of publication DD MM YYYY; date of current version DD MM YYYY

Authorized licensed use limited to: Politecnico di Torino. Downloaded on July 01,2025 at 13:10:54 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Consumer Electronics Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCE.2025.3544816

rience [4], [5], especially in scenarios that simulate human behavior, such as medical training or virtual customer service [6], [7]. However, achieving this emotional authenticity is anything but easy. It requires a deep understanding of human emotional expression, especially facial movements, which are one of the most important forms of emotional communication [8].

One of the most important tools for creating realistic emotional expressions in VHs is the Facial Action Coding System (FACS) [9]. FACS breaks down complex facial expressions into individual components called Action Units (AUs), where each AU corresponds to a specific facial muscle movement (e.g. raising the eyebrows, pulling the corners of the mouth). FACS and its extension EMotional FACS (EMFACS) [10] provide a solid framework for encoding basic emotions and animating VH faces with blendshape techniques. However, determining the correct intensity or weighting of each AU to achieve believable emotional expressions remains a challenge. Previous studies have implemented FACS-based emotion coding in specific models [6], [11], [12], but they implicitly assume that the same weightings can be universally applied to VHs with different characteristics (e.g., gender, ethnicity, age). However, creating emotionally realistic VHs across different platforms and devices is challenging, as empirical evidence suggests that even small variations in blendshape weights can lead to significantly different emotional perceptions depending on the VH model used [13]. Therefore, a critical question remains: Are there universal sets of blendshape weights that can be consistently applied to all VH models to effectively convey specific emotions? Or are different sets of weights required for each emotion and for specific types of VHs? To answer these questions, we conducted an exploratory research [14] focusing on the expression of happiness. In this study, we extended our original investigation to include two additional emotions: sadness and disgust. By collecting real human preferences, we investigated how subtle facial expressions influence VHs with different characteristics. This approach provides deeper insights into how the same emotion is perceived in different VHs, simplifies the complex process of encoding emotions, and lays the foundation for creating more realistic and emotionally believable VHs.

RELATED WORKS

The application of FACS to enhance the emotional expressiveness of VHs has been widely studied in both research and industry, particularly in virtual assistants, games, and other human-computer interface technologies [15].

Early studies, such as [12] and [16], demonstrated the technical feasibility of integrating FACS with blendshape techniques for VH animation. These works employed geometric transformations to replicate facial muscle movements, providing a foundation for FACS-based animation methods. However, both studies share a critical limitation: they focused on the mechanical implementation of FACS without evaluating with users the believability of the depicted emotions. They also assumed that facial muscle movements captured in humans could be directly transferred to VHs, neglecting the nuances in users' perception of artificial characters.

The work in [11] advanced the field by simulating emotions more realistically through autonomic physiological control and musculoskeletal modeling, enabling dynamic changes such as skin tone and pupil dilation. While this approach was innovative, it similarly lacked user-centered validation of the generated expressions, presuming that applying FACS to VHs would inherently produce believable emotions. This raises concerns about whether users interpret these expressions the same way as human facial expressions, especially in non-realistic or low-resolution VHs, which are common in consumer electronics.

Beyond these technical approaches, studies on emotional contagion (i.e., the phenomenon where one person's emotions mimic those of others) between users and VHs have contributed valuable insights. For instance, [17] explored emotional contagion through observational analysis, highlighting the variability in human smiles. Although FACS was not explicitly utilized, this study suggested that a broader range of nuanced expressions is essential for creating lifelike VHs. In contrast, [7] demonstrated that VHs displaying emotions based on FACS could elicit emotional contagion in users. Notably, this study employed FACS to design emotional expressions, making it relevant for applications like home automation and AI-powered virtual assistants. However, neither study thoroughly addressed the impact of emotional intensity on user perception. While [17] did not use FACS, [7] integrated it, leaving a gap in understanding how variations in emotional intensity influence user experiences.

Authorized licensed use limited to: Politecnico di Torino. Downloaded on July 01,2025 at 13:10:54 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Consumer Electronics Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCE.2025.3544816

This gap was addressed by [6], who investigated the role of intensity in facial expressions, specifically in expressions of pain. Using controlled experimental setups, the authors found that higher intensities of Action Units (AUs) correlated with stronger perceptions of pain. While their findings provided valuable insights, the generalizability to other emotions or VH models remains unaddressed, highlighting the need for further research on emotional intensity across diverse scenarios.

Recent studies have continued to explore the nuances of emotional expression in VHs, leveraging FACS as the foundation for emotional modeling. In [18], the authors investigated how different levels of visual realism in VHs affect users' empathic responses. They found that higher levels of realism can increase cognitive and affective empathy. Similarly, [19] investigated the perception of micro and macro facial expressions in realistic VHs and found that transferring the expressions of real actors to virtual models can reduce the accuracy of emotion recognition compared to artistically designed animations previously evaluated in [20].

These studies highlight the growing interest in using FACS to enhance the emotional expressiveness of VHs, a crucial aspect in applications and consumer systems that rely on VHs to improve user engagement. However, a significant challenge remains in determining whether a universal set of AU weights can effectively convey emotions across VHs with diverse characteristics, such as gender and visual fidelity. Considering the wide range of potential applications, such as gaming, virtual assistants, customer service, and home automation, finding such a universal solution could be highly beneficial. This underscores the need for optimized FACS-based approaches tailored to specific subsets of VHs to ensure believable emotional expressions in varied consumer electronics applications.

Our work addresses this gap by investigating how FACS can be standardized across different VH models, considering specific characteristics such as resolution and gender, to improve emotional believability and ensure more lifelike and engaging interactions in various systems.

EVALUATION OF EMOTIONAL PERCEPTION IN VH'S PORTRAITS

To investigate how varying the weightings of FACS AUs influences the effectiveness of conveying emotions through different VHs, we collected user feedback on the believability of these emotions. Users were asked to choose which ***portrait*** (i.e., static images of a ***VH*** displaying a specific ***expression*** defined by a set of AU weights) better expressed a given emotion. We focused our analysis on three emotions: **happiness**, **sadness**, and **disgust** because they require the fewest AUs to portray. This reduced number of AUs led to fewer possible combinations of AU weightings, decreasing the number of user evaluations needed. In particular, the following AU combinations were used to convey the studied emotions [10]. **Happiness**: AU6 (*cheek raiser*) and AU12 (*lip corner puller*). **Sadness**: AU1 (*inner brow raiser*), AU4 (*brow lowerer*) and AU15 (*lip corner depressor*). **Disgust**: AU9 (*nose wrinkler*), AU15 and AU17 (*chin raiser*).

Portrait datasets

For each emotion analyzed, we created a ***portrait*** dataset of 100 different ***VHs*** displaying the respective emotion with different ***expressions (e)***. Each expression was created by combining the AUs (associated with each emotion), which could assume three discrete intensity values (0.4, 0.675, and 0.95), avoiding lower values (too neutral) and higher values (prone to rendering artifacts in low-resolution models [14]). The result was a number of different expressions for each VH equal to 3^v, where v is the number of AUs encoding the emotion (Figure 1). To explore how different levels of visual fidelity affect the perception of emotions conveyed by varying AU intensities, we selected VHs from two popular 3D libraries: Rocketbox avatars [21] and MetaHumans[1]. These two libraries represent a significant range in visual fidelity. Rocketbox models have a lower resolution, making them computationally efficient but less detailed, while MetaHumans offer high-resolution models that are more visually expressive but require higher computational resources, which can be challenging for real-time applications. To ensure consistency, all portraits were rendered using UE5 with consistent lighting and camera conditions.

User perception and ranking

To assess how effectively these expressions conveyed the analyzed emotions, we developed a web-based voting application. For each emotion, participants were presented with two clickable portraits of the same VH, each showing a different expression.

[1] https://www.unrealengine.com/en-US/metahuman

Authorized licensed use limited to: Politecnico di Torino. Downloaded on July 01,2025 at 13:10:54 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Consumer Electronics Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCE.2025.3544816

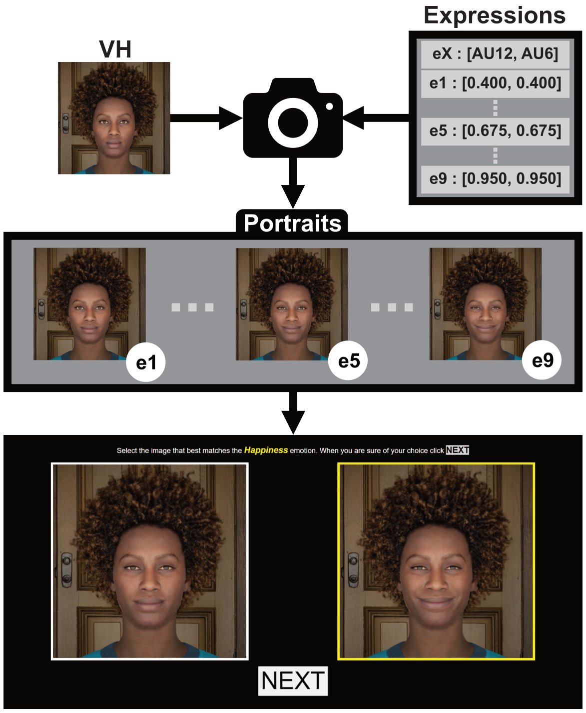

Figure 1. The ***portrait*** dataset generation pipeline and the interface of the web app used for collecting users' votes. As an example, only one VH is shown with expressions of happiness.

The task was to select the expression (i.e., portrait) that best represented the emotion. For each VH, all expressions were compared to one another following a round-robin tournament. A confirmation step ensured that users could reconsider their choice before casting their vote. Once a vote was cast, a win was recorded for the chosen expression and a loss for the other. A new pair of portraits were then presented (Figure 1).

We then calculated a ranking score for each expression using the following formula:

$$s_{c,i} = \frac{w_{c,i} - l_{c,i}}{w_{c,i} + l_{c,i}} \tag{1}$$

where $w_{c,i}$ and $l_{c,i}$ are the number of wins and losses for the expression i of the VH c. This score, normalized between -1 and 1, allowed us to compare the effectiveness of the expressions regardless of the number of votes received.

Finally, we evaluated the performance of each expression within specific VH subgroups by calculating the top-k accuracy $A_{i,g}$, i.e., the frequency with which a given expression i was among the top k choices for each VH in the subgroup g.

EXPERIMENTAL PROTOCOL

In our experiments, we selected 100 VHs subdivided by gender (25 male and 25 females from each VH library). Given the limited diversity of ethnicities and the lack of reference values for the age of the VHs available, we created different ***VH subgroups*** based on their *gender* (male, M, female, F) and *resolution* (low, L, for Rocketbox, and high, H, for MetaHumans).

User evaluations were collected using the voting app and for consistency with the approach used in [14], $k = 3$ was used in computing the top-k accuracies. The total number of comparisons required for each VH is $T = p \cdot (p-1)/2$, where p is the number of expressions for each emotion. To complete a full round-robin, $n \cdot T$, comparisons were needed for n VHs. Therefore, the number of comparisons increases with the number of AUs used to represent an emotion. We note that, to reduce the number of comparisons, we excluded the neutral expression, which is characterized by setting all AUs to 0, because, as shown in [14], it was not selected as the winner in any of the comparisons. This result was confirmed by our initial tests with the new emotions, so we could confidently exclude the neutral expression without jeopardizing the validity of the results. However, while happiness involves two AUs, resulting in $p = 9$ expressions and a manageable number of 3600 comparisons, sadness and disgust each requires three AUs ($p = 27$), resulting in a significantly higher number of 35100 comparisons, which poses a challenge as it requires a considerable number of volunteers and a longer time frame for data collection.

Therefore, we developed a heuristic for sadness and disgust to reduce the number of required matches while still providing meaningful results. Starting from the observation that rather than aiming to produce a complete ranking of all expressions for each emotion, we aim to identify the top-performing expressions (i.e., those that convey the emotion most effectively), we organized the voting process into three rounds (R) (Figure 2):

- **R1 – Initial Screening**: We randomly selected two VHs from each VH subgroup (FL, ML, FH, MH), and compared all expression pairings across two full voting rounds (5616 comparisons). Expressions were then ranked in descending order based on eq. 1.
- **R2 – Grouped Competition**: The top 10 expressions from R1 were divided into two groups: A

Authorized licensed use limited to: Politecnico di Torino. Downloaded on July 01,2025 at 13:10:54 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Consumer Electronics Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCE.2025.3544816

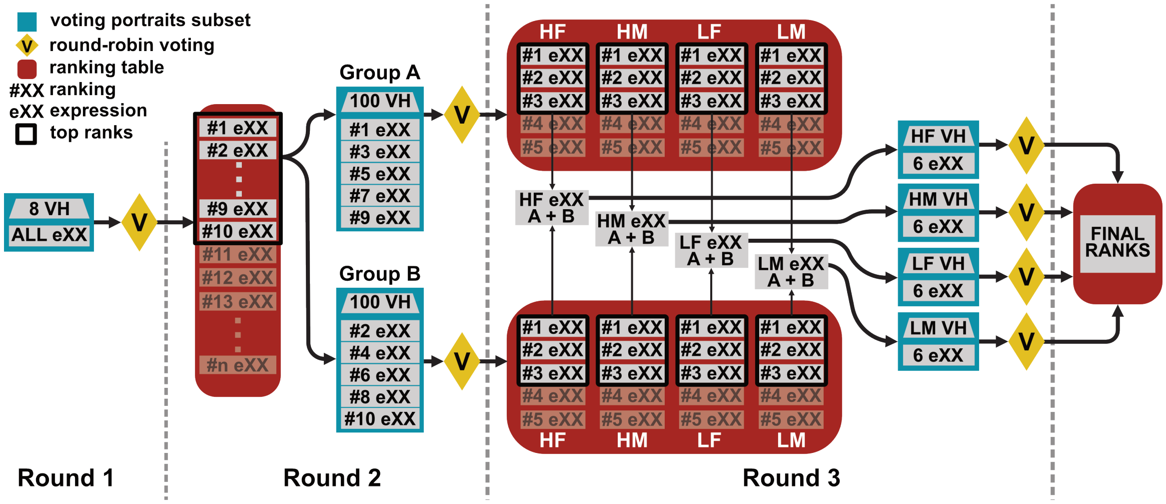

Figure 2. Schema of the three-round experimental protocol followed for sadness and disgust.

(evenly ranked) and B (oddly ranked), ensuring stronger expressions did not immediately compete against each other. Each group underwent round-robin comparisons for each VH, totaling 2000 comparisons.

- **R3 – Final Showdown**: After R2, for each VH subgroup, we selected the top three expressions from group A and B. These formed four final sets of six expressions (one set per subgroup). Each set was compared in the final round, resulting in 1500 comparisons.

Using this heuristic reduces the number of comparisons for sadness or disgust to 9116, representing only 25% of the total comparisons potentially needed for a full round-robin. This reduction makes data collection more manageable within a reasonable time frame. However, as a heuristic approach, the results may be prone to some error since not all expressions are exhaustively compared. Additionally, if an expression is not chosen in R3 for a specific VH subgroup, its top-k accuracy is not computed. This can potentially underestimate its true performance across various VH groupings, such as by resolution and gender.

The app was shared with university employees and students, and informed consent was obtained from all participants. Participation was voluntary, and anonymous votes were recorded, with no personal data collected.

RESULTS

We start by analyzing the accuracy results for each emotion, using top-k accuracy with $k = 3$ for consistency with [14] across all emotions and VH subgroups. We then summarize the findings to provide an overview of the observed patterns.

Happiness

For happiness, we collected 7200 votes from 169 users, which corresponds to two complete rounds of portrait comparisons. Looking at the overall results (group ALL in Table 1), three expressions (e7, e8, e9) clearly stand out from the others in terms of user preference, as they have significantly higher accuracy compared to the other expressions. However, none of them emerged as a clear winner, suggesting that while some expressions are universally effective, there may not be a one-size-fits-all solution for VHs.

For all three top-ranking expressions, the emphasis on smiling (AU12, lip-pulling) is crucial for effectively conveying happiness. It is noticeable that the credibility of the happiness expression decreases when the intensity of AU12 decreases. Interestingly, even an increase in AU6 (cheek-raiser), as in e9, did not necessarily lead to better performance compared to e7 or e8, showing that more intensity does not always equate to greater believability.

When comparing results by gender, we find minimal differences between men and women for the top three expressions, suggesting they are largely gender-neutral. However, resolution had a more significant impact. High-resolution models provided more believ-

Authorized licensed use limited to: Politecnico di Torino. Downloaded on July 01,2025 at 13:10:54 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Consumer Electronics Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCE.2025.3544816

Table 1. Top-3 accuracy of Happiness expression in different groups. The best results in bold underlined, the second ranked results in bold, the third ranked results in underlined.

	AUs weights		ALL	Gender		Resolution		Subgroups			
expression	**AU12**	**AU6**	**k = 3**	**Male**	**Female**	**Low**	**High**	**LM**	**LF**	**HM**	**HF**
e2	0.400	0.675	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
e3	0.400	0.950	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
e1	0.400	0.400	0.01	0.02	0.00	0.02	0.00	0.04	0.00	0.00	0.00
e6	0.675	0.950	0.18	0.22	0.14	0.04	0.32	0.08	0.00	0.36	0.28
e5	0.675	0.675	0.19	0.16	0.22	0.32	0.06	0.24	0.40	0.08	0.04
e4	0.675	0.400	0.24	0.30	0.18	0.38	0.10	0.52	0.24	0.08	0.12
e9	0.950	0.950	0.68	0.64	0.72	**0.72**	**0.88**	0.44	0.52	**0.84**	**0.92**
e8	0.950	0.675	**0.78**	**0.82**	**0.74**	0.68	**0.88**	**0.68**	**0.68**	**0.96**	0.80
e7	0.950	0.400	**0.80**	**0.72**	**0.88**	**0.84**	**0.76**	**0.76**	**0.92**	0.68	**0.84**

able expressions, thanks to features like detailed skin textures and wrinkle maps. In contrast, low-resolution models struggled with expressions involving intense AU6, leading to unperceivable deformations due to the lack of facial detail. This caused expressions to appear more uniform, making e7 (with low AU6 intensity) the best performer, compared to the high-resolution e9 (with high AU6 intensity), as shown in Figure 3.

Sadness

For sadness (and later for disgust), the online voting platform automatically handled the three rounds of heuristic evaluation, moving on to the next round as soon as the total number of votes required to complete the current round was reached. In total, we collected 9116 votes from 166 users. The results are summarized in Table 2, with the note that any expression that was not included in the final set for a given VH subgroup was scored with a precision of zero. Overall, nine out of ten expressions passed R2, with only three of them included in all subgroups.

If we look at the overall results (group ALL), we can identify a set of high-performing expressions (e21, e24 and e27) that consistently show strong activation of AU1 and AU15. While AU1 (with its upward movement of the inner eyebrow, which enhances the impression of vulnerability) and AU15 (which pulls the corners of the lips downward, creating a drooping expression that conveys emotional heaviness) remain key features for the portrayal of sadness, the notable differences in the activation levels of AU4 (brow lowerer), which plays a more nuanced role, highlight the complexity of balancing the different AUs to achieve optimal emotional expression across different VHs.

When we break down the results by gender, we find that there are minimal differences between male and female VHs in the best-performing expressions, indicating again their gender-neutral nature. The dominant factor, as with happiness, is the resolution of the VH. While some expressions, such as e21 and e24, perform equally well on both low and high-resolution models, there are clear differences for others. For example, e27 (Figure 3) is the top performer in high-resolution VHs, but has significantly lower accuracy in low-resolution VHs, and e20 and e12 perform well in low-resolution VHs, but are less effective in high-resolution models: e20 did not reach R3 and e12 is among the lowest ranked expressions.

The main reason for these differences seems to be the lack of detailed facial features in low-resolution models. In particular, the intensity of AU4 plays a crucial role. A high activation of AU4 (as in e27) tends to have a negative effect on the emotional representation in these models. Without fine surface detail, the pronounced downward movement of the eyebrows can distort the intended expression of sadness, sometimes turning it into an expression of anger. It is noteworthy that the optimal expressions for low-resolution models have lower levels of activation of AU4 (0.400 for e21 and e20, and 0.675 for e24).

Disgust

For disgust, we collected 9116 votes from 168 users. Seven out of ten expressions passed R2, four of which were represented in all subgroups. The results are summarized in Table 3.

Authorized licensed use limited to: Politecnico di Torino. Downloaded on July 01,2025 at 13:10:54 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Consumer Electronics Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCE.2025.3544816

Figure 3. Top expressions of the three emotions for high-resolution VHs (top) and low-resolution VHs (bottom).

Table 2. Top-3 accuracy of Sadness expressions across different groups. The 10 expressions are derived from R2, with expressions not included in R3 for a specific VH subgroup marked with '-'. The best results are highlighted as <u>bold underlined</u>, second-ranked as bold, and third-ranked as <u>underlined</u>.

	AUs weights			ALL	Gender		Resolution		Subgroups			
expression	**AU15**	**AU4**	**AU1**	**k = 3**	**Male**	**Female**	**Low**	**High**	**LM**	**LF**	**HM**	**HF**
e9	0.400	0.950	0.950	0.00	0.00	0.00	0.00	0.00	-	-	-	-
e11	0.675	0.400	0.675	0.06	0.12	0.00	0.12	0.00	0.24	-	-	-
e18	0.675	0.950	0.950	0.10	0.20	0.00	0.00	0.20	-	-	0.40	-
e15	0.675	0.675	0.950	0.15	0.08	0.22	0.14	0.16	-	0.28	0.16	0.16
e23	0.950	0.675	0.675	0.21	0.06	0.36	0.06	0.36	0.12	-	-	**0.72**
e20	0.950	0.400	0.675	0.29	0.28	0.30	**0.58**	0.00	0.56	**0.60**	-	-
e12	0.675	0.400	0.950	0.34	<u>0.46</u>	0.22	0.46	0.22	<u>0.60</u>	0.32	0.32	0.12
e27	0.950	0.950	0.950	<u>0.49</u>	<u>0.46</u>	<u>0.52</u>	0.18	<u>**0.80**</u>	-	0.36	<u>**0.92**</u>	<u>0.68</u>
e24	0.950	0.675	0.950	**0.65**	**0.64**	**0.66**	<u>0.56</u>	**0.74**	**0.64**	<u>0.48</u>	**0.64**	<u>**0.84**</u>
e21	0.950	0.400	0.950	<u>**0.71**</u>	<u>**0.70**</u>	<u>**0.72**</u>	<u>**0.90**</u>	<u>0.52</u>	<u>**0.84**</u>	<u>**0.96**</u>	<u>0.56</u>	0.48

Looking at the overall results (group ALL), a number of high-performing expressions can be identified (e27, e24, and e21), all showing the highest activation values for AU17 and AU9. These AUs play a prominent role in the representation of disgust, with AU17 raising the chin to enhance the impression of tension while AU9 pinches the nasal region to emphasize the feeling of disgust. However, the role of AU15 (lip corner depressor) is more variable. The different levels of activation in the different expressions highlight the complexity of achieving an optimal expression of disgust in different VHs.

Authorized licensed use limited to: Politecnico di Torino. Downloaded on July 01,2025 at 13:10:54 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Consumer Electronics Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCE.2025.3544816

Table 3. Top-3 accuracy of Disgust expressions across different groups. The 10 expressions are derived from R2, with expressions not included in R3 for a specific VH subgroup marked with '-'. The best results are highlighted as bold underlined, second-ranked as bold, and third-ranked as underlined.

	AUs weights			ALL	Gender		Resolution		Subgroups			
expression	**AU17**	**AU15**	**AU9**	**k = 3**	**Male**	**Female**	**Low**	**High**	**LM**	**LF**	**HM**	**HF**
e9	0.400	0.950	0.950	0.00	0.00	0.00	0.00	0.00	-	-	-	-
e12	0.675	0.400	0.950	0.00	0.00	0.00	0.00	0.00	-	-	-	-
e20	0.950	0.400	0.675	0.00	0.00	0.00	0.00	0.00	-	-	-	-
e26	0.950	0.950	0.675	0.26	0.30	0.22	0.22	0.30	-	0.44	0.60	-
e15	0.675	0.675	0.950	0.31	0.14	0.48	0.48	0.14	0.28	**0.68**	-	0.28
e23	0.950	0.675	0.675	0.31	0.26	0.36	0.38	0.24	0.40	0.36	0.12	0.36
e18	0.675	0.950	0.950	0.39	**0.62**	0.16	0.30	0.48	**0.60**	-	**0.64**	0.32
e21	0.950	0.400	0.950	0.45	0.52	0.38	**0.52**	0.38	**0.72**	0.32	0.32	0.44
e24	0.950	0.675	0.950	**0.60**	0.48	**0.72**	**0.62**	**0.58**	0.48	**0.76**	0.48	**0.68**
e27	0.950	0.950	0.950	**0.68**	**0.68**	**0.68**	0.48	**0.88**	0.52	0.44	**0.84**	**0.92**

Significant gender differences are evident when examining optimal gender-specific expressions such as e21, e24 and e18, with a delta of up to 0.24 between male and female VHs. These differences challenge the gender-neutral performance observed for happiness and sadness and suggest that gender plays a greater role in how disgust is conveyed and perceived.

Resolution plays a key role in accuracy differences between VHs (Figure 3). In high-resolution VHs, reducing AU intensity (from e27, where all AUs are maximized, to e21) leads to a notable drop in accuracy, while low-resolution VHs are less impacted by these variations. Overall, e24 performs well across both types, although e27 remains the best option for high-resolution VHs.

However, despite these observations, it is difficult to identify clear patterns correlating AU activations, facial detail, and the accuracy of disgust expressions. This suggests that other underlying factors or AU interactions may influence how these expressions are perceived at different resolutions, making the role of facial detail less significant in this case compared to other emotions.

Nevertheless, there is clear evidence that optimal expressions can be defined for certain subgroups (i.e. when both gender and resolution are taken into account). These results show that different emotions require different degrees of adjustment in VH facial traits, but the underlying principle of optimizing expressions for specific subgroups remains valid.

Discussion

Overall, our findings indicate that no single expression fits all VHs universally. However, they do support the existence of expression sets that are effective within specific VH subgroups. Some emotions, like happiness and sadness, can be expressed similarly across genders, while others, like disgust, require gender-specific adaptations to achieve equal emotional clarity. Resolution also plays a crucial role in emotional expressiveness. As observed, the low-resolution models lack the fine facial detail required to accurately reproduce subtle AU activations, especially when their intensity is maximized. However, the results for disgust suggest that factors other than facial details, such as AU interactions, may also play a significant role. This observation suggests that greater caution is needed when establishing direct relationships between AU activations and VH resolution, particularly when the emotion involves the activation of multiple AUs.

Although the results of this study provide valuable insights, it is important to acknowledge some limitations. First, we focused on a limited number of emotions, which may not fully capture the broader range of expressions relevant to different applications. In addition, the use of three intensity levels for the AUs, although practical, may overlook finer emotional nuances. The heuristics-based voting system helped to manage complexity, but certain expressions may not have been fully explored. Finally, the relatively small group of participants and low ethnic diversity in our VH models may limit the generalizability of the re-

Authorized licensed use limited to: Politecnico di Torino. Downloaded on July 01,2025 at 13:10:54 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Consumer Electronics Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCE.2025.3544816

sults. Future studies could benefit from collecting non-anonymous votes along with user-related data such as age, gender, nationality and occupation to enable a deeper analysis of potential perceptual variability across different demographic groups.

CONCLUSIONS

This study investigated how emotions can be expressed by VHs using EMFACS coding in different subgroups defined by gender and resolution. While no universal set of expressions was found, our results show that there are different sets of optimal expressions for specific subgroups. Furthermore, these results challenge the conventional approach of maximizing all AU intensities, which often leads to a less effective representation of emotions. This suggests that a more nuanced approach with tailored modulations for each subgroup is required to achieve believable expressions. Although this study focused on a limited subset of emotions, we believe it offers valuable insights that can help animators and practitioners refine emotional expression in VHs to ensure more realistic and engaging interactions.

REFERENCES

1. L. Cui and J. Liu, “Virtual human: A comprehensive survey on academic and applications,” *IEEE Access*, 2023.
2. B. Lugrin, “Introduction to socially interactive agents,” in *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition*, 2021, pp. 1–20.
3. E. Karuzaki, N. Partarakis, N. Patsiouras, E. Zidianakis, A. Katzourakis, A. Pattakos, D. Kaplanidi, E. Baka, N. Cadi, N. Magnenat-Thalmann *et al.*, “Realistic virtual humans for cultural heritage applications,” *Heritage*, vol. 4, no. 4, pp. 4148–4171, 2021.
4. R. Hortensius, F. Hekele, and E. S. Cross, “The Perception of Emotion in Artificial Agents,” *IEEE Trans. Cogn. Dev. Syst.*, vol. 10, no. 4, pp. 852–864, Dec. 2018.
5. A. Chirico and A. Gaggioli, “When virtual feels real: Comparing emotional responses and presence in virtual and natural environments,” *Cyberpsychology, Behavior, and Social Networking*, vol. 22, no. 3, p. 220–226, Mar. 2019.
6. G. Carbonell and J. Schild, “Conceptual Design of Emotional and Pain Expressions of a Virtual Patient in a Virtual Reality Training for Paramedics,” in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. Christchurch, New Zealand: IEEE, Mar. 2022, pp. 222–226.
7. M. Volonte, Y.-C. Hsu, K.-Y. Liu, J. P. Mazer, S.-K. Wong, and S. V. Babu, “Effects of Interacting with a Crowd of Emotional Virtual Humans on Users’ Affective and Non-Verbal Behaviors,” in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. Atlanta, GA, USA: IEEE, Mar. 2020, pp. 293–302.
8. P. M. Niedenthal and M. Brauer, “Social Functionality of Human Emotion,” *Annu. Rev. Psychol.*, vol. 63, no. 1, pp. 259–285, Jan. 2012.
9. P. Ekman and W. V. Friesen, “Facial action coding system,” *Environmental Psychology & Nonverbal Behavior*, 1978.
10. 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.
11. Y. Tisserand, R. Aylett, M. Mortillaro, and D. Rudrauf, “Real-time simulation of virtual humans’ emotional facial expressions, harnessing autonomic physiological and musculoskeletal control,” in *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. Virtual Event Scotland UK: ACM, Oct. 2020, pp. 1–8.
12. M. H. Alkawaz, D. Mohamad, A. H. Basori, and T. Saba, “Blend Shape Interpolation and FACS for Realistic Avatar,” *3D Res*, vol. 6, no. 1, p. 6, Mar. 2015.
13. D. Kumarapeli, S. Jung, and R. W. Lindeman, “Emotional avatars: Effect of uncanniness in identifying emotions using avatar expressions,” in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 2022, pp. 650–651.
14. S. Calzolari, F. Strada, and A. Bottino, “The quest for believability: Exploring facs adaptations for emotion facial expressions in virtual humans,” in *2024 IEEE Gaming, Entertainment, and Media Conference (GEM)*, June 2024, pp. 1–6.
15. M. Seymour, “Facs at 40: facial action coding system panel,” in *ACM SIGGRAPH 2019 Panels*, 2019, pp. 1–2.
16. M. S. M. Rahim, A. H. Basori, S. M. Saadi, A. E. Rad, and A. Rehman, “Emotional Facial Expression and Tears Simulation: An Analysis & Comparison of Current Approaches,” *3D Res*, vol. 6, no. 2, p. 14, Jun. 2015.

Authorized licensed use limited to: Politecnico di Torino. Downloaded on July 01,2025 at 13:10:54 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Consumer Electronics Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCE.2025.3544816

17. J. Tsai, E. Bowring, S. Marsella, W. Wood, and M. Tambe, "A study of emotional contagion with virtual characters," in *Intelligent Virtual Agents: 12th International Conference, IVA 2012, Santa Cruz, CA, USA, September, 12-14, 2012. Proceedings 12*. Springer, 2012, pp. 81–88.
18. D. Higgins, Y. Zhan, B. R. Cowan, and R. McDonnell, "Investigating the effect of visual realism on empathic responses to emotionally expressive virtual humans," in *ACM Symposium on Applied Perception 2023*, 2023, pp. 1–7.
19. R. H. Montanha, G. N. Raupp, A. C. P. Schmitt, V. F. de Andrade Araujo, and S. R. Musse, "Micro and macro facial expressions by driven animations in realistic virtual humans," *Entertainment Computing*, vol. 52, p. 100853, 2025.
20. R. Montanha, G. Raupp, V. Gonzalez, Y. Partichelli, A. Bins, M. Ferreira, V. Araujo, and S. Musse, "Revisiting micro and macro expressions in computer graphics characters," in *Proceedings of the 22nd Brazilian Symposium on Games and Digital Entertainment*, 2023, pp. 38–45.
21. M. Gonzalez-Franco, E. Ofek, Y. Pan, A. Antley, A. Steed, B. Spanlang, A. Maselli, D. Banakou, N. Pelechano, S. Orts-Escolano, V. Orvalho, L. Trutoiu, M. Wojcik, M. V. Sanchez-Vives, J. Bailenson, M. Slater, and J. Lanier, "The rocketbox library and the utility of freely available rigged avatars," *Frontiers in Virtual Reality*, vol. 1, Nov. 2020.

Stefano Calzolari is currently a PhD student at the Department of Control and Computer Engineering of the Politecnico di Torino where he conducts research under the Computer Graphics and Vision research group of the same university. His current research interests include Virtual Humans for Virtual and Augmented Reality applications. Contact him at stefano.calzolari@polito.it.

Francesco Strada is currently an Assistant Professor at the Department of Control and Computer Engineering of Politecnico di Torino, where he conducts research under the Computer Graphics and Vision research group of the same university. His current research interests include Virtual and Augmented Reality for learning and training, Serious Games, Human-Computer Interaction, and the use of immersive technologies in collaborative scenarios. Contact him at francesco.strada@polito.it.

Andrea Bottino is currently an Associate Professor at the Department of Control and Computer Engineering of the Politecnico di Torino, where he heads the Computer Graphics and Vision research group of the same university. His current research interests include Computer Vision, Machine Learning, Human-Computer Interaction, Serious Games, and Virtual and Augmented Reality. He is a member of the IEEE. Contact him at andrea.bottino@polito.it.

Authorized licensed use limited to: Politecnico di Torino. Downloaded on July 01,2025 at 13:10:54 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.