Face Off: Evaluating Virtual Human Expressions and Non-Tracking Control Methods in VR

Sangeeth Chandran J K
School of Computer Science
Technological University Dublin, ADAPT
Dublin, Ireland
d22128840@mytudublin.ie
*Corresponding author

Marisa Llorens Salvador
School of Computer Science
Technological University Dublin
Dublin, Ireland
marisa.llorens@tudublin.ie

Cathy Ennis
Department of Computer Science
Maynooth University
Maynooth, Ireland
cathy.ennis@mu.ie

2025 11th International Conference on Virtual Reality (ICVR) | 979-8-3503-9273-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICVR66534.2025.11172599

***Abstract*—Social virtual reality (VR) applications have become more ubiquitous in recent years; central to this is the communication pipeline, how users perceive virtual human facial expressions, and how they control them in real time, especially when using VR devices without face-tracking. We investigated both aspects in a set of experiments. Firstly, we compared the perception of virtual human emotions on a traditional 2D screen and in VR. In a second experiment, we used a validated set of stimuli to compare three different control methods for manipulating an avatar's facial expressions in VR. These control methods utilize non-tracking control techniques, which do not rely on real-time face tracking but rely on alternative inputs via the VR controller. Our analysis shows that in VR, the effectiveness ratings for happy, sad, and surprise were significantly higher, and disgust was significantly more recognizable, compared to the screen. These findings contribute to our understanding of virtual human based emotional communication in VR by demonstrating that the perception of facial expression varies between screen and VR. Additionally, we identify raycast selection (point and click) as the most accurate control method, whereas thumbstick labeled (using a controller thumbstick with UI labels for guidance) was the fastest and most preferred method by participants.**

***Index Terms*—avatar, facial expression, emotion perception, control method, VR, virtual reality**



Fig. 1: Facial expression stimuli that represent the six basic human emotions.

I. Introduction

Facial expression is essential for conveying emotions, establishing connections, and displaying mental states. This also applies to virtual humans (VHs) [1]. Studies have shown that facial expressions of VHs have a significant impact on the emotions of those who perceive them, directly influencing their actions in a virtual setting [2] [3]. This signifies the importance of expressive VHs, who can display facial expressions to have meaningful social interaction in virtual worlds [4].

The merits of Virtual Reality (VR) over traditional 2D screens have been studied very well, from which it has been concluded that VR improves immersion, presence, and emotional engagement compared to 2D screens [5]. However, VR's impact on facial expression recognition has not yet been studied extensively. To address this gap, we conducted a comparative study to understand the recognition of facial expressions of VHs in VR.

As VHs become more central to social VR and immersive applications, understanding how facial expressions are interpreted and how they can be controlled becomes critical. Our study investigates VH communication using facial expressions in immersive environments. We pose the following research questions (RQ):

- RQ1: Does VR affect viewer perception of the facial expressions of VHs for a range of emotions?
- RQ2: How do non-tracking control methods for VHs facial expressions perform in terms of usability?

We designed two separate experiments to answer these questions. To understand VH mediated communication in VR, we examined two important aspects of communication: perception (how users interpret emotions) and control (how users express emotions). By investigating these aspects separately, we clearly isolate the factors that influence emotion recognition (Experiment 1) and the usability of avatar expression

This research was conducted with the financial support of Research Ireland at ADAPT, the Research Ireland Centre for AI-Driven Digital Content Technology at TU Dublin [13/RC/2106$_P$2]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Authorized licensed use limited to: Maynooth University Library. Downloaded on November 07,2026 at 15:14:45 UTC from IEEE Xplore. Restrictions apply.

controls (Experiment 2). In the second experiment, the VHs are embodied, so the appropriate term for VHs in the second experiment would be "avatars".

Recent advancements in VR hardware, particularly the incorporation of eye and face tracking devices into Head Mounted Displays (HMDs), have revolutionized the process of tracking users' facial expressions and projecting them onto their avatars. Premium high-end devices or enterprise VR devices such as *Meta Quest Pro*and *HTC Vive Pro 2* have these capabilities, but their low-end counterparts or consumer-grade VR devices, such as the recently launched Meta Quest 3, do not have integrated face-tracking capabilities; this reason gave us the motivation to focus on non-tracking control methods. A recent survey from the digital distribution service Steam [6] shows that over 80% of Steam VR users own a VR device without face-tracking capabilities. This means that capturing facial expressions on widely adopted social VR platforms is a challenge. A survey conducted by Sykownik et al. [7] among social VR users revealed the demand for better tracking technologies among users. Given the absence of face-tracking technology in the most popular Head Mounted Displays, it is increasingly important to explore options that do not involve face tracking. Software-driven approaches that do not require additional hardware offer a more accessible and scalable alternative. Such methods can be integrated into most VR devices without the need for hardware augmentations, making expressive avatars inclusive to a broader user base. By combining both observer and user perspectives, our research contributes insight into VH communication in VR. Our findings highlight the importance of considering the display modality when designing VH facial expressions for VR, as an expression designed on screen can be perceived differently in VR. Beyond perception, we provide a comparative performance analysis of three non-tracking methods for controlling embodied VH (avatar) facial expression based on their accuracy, speed, and user preference. Our contributions can guide the design and development of expressive VHs suited to the constraints of consumer-grade VR devices.

II. Related Work

There have been interesting studies that focus on the effect of VR on emotion perception. To reflect this multifaceted research landscape, we discuss various animation techniques that use different styles of VHs. Tian et al. [8] investigated the emotional impact of VR on users compared to 2D. The study used physiological measurements, such as an electroencephalogram (EEG), and concluded that VR elicited a greater emotional response from users, potentially as a result of its increased immersion. Geraets et al. [9] evaluated the feasibility of using VR tasks for training and evaluating facial emotion recognition and found that VR tasks are effective for training. They compared pictures of people with facial expressions representing basic human emotions to VHs with the same expressions in VR. They found that emotion recognition in VR is comparable to pictures except for disgust and happy. In addition, they also found that there was no difference in emotion recognition between male and female participants. Mal et al. [10] investigated how immersion levels affect the plausibility of VHs from the user's perspective. They found that the higher immersiveness of VR made VHs more plausible and increased the VHs social presence. Kuhne et al. [5] conducted a similar study. They evaluated the comparative impact of immersive VR on the user's cognitive, physiological, and emotional state. The results show that the use of VR leads to increased emotional engagement, presence, and an increase in physiological values such as heart rate and skin conductance. Farizi et al. [11] examined how avatars can convey deception through micro-expressions. They used 2D and 3D representation of avatars in VR to assess deception detection in VR. They found that human observers struggled to detect deception accurately, underscoring the challenges of replicating subtle facial cues in virtual environments. Interestingly, there was no significant difference in the way participants detected deception in 2D and 3D avatars. Although their findings show no difference in emotion detection between 2D and 3D avatars, there is a substantial methodological difference between our study and theirs. As they mainly use VR to test both 2D and 3D scenarios, we use screen to test facial expression recognition of 2D VHs and use VR only to test facial expression recognition of 3D VHs. These studies collectively highlight VR's potential to be more engaging and become a medium for hosting captivating interactive content. However, it is not yet fully understood how VR affects facial expressions when the same stimuli are shown on screen and in VR.

While there are few studies that consider non-tracking methods to manipulate avatars' facial expressions in VR, there are some studies that examine the topic of avatars' expressions and emotions, focusing on its technology and psychology.

Recent advances in avatar facial expression synthesis and animation transfer techniques are highlighted in the studies by Hwang & Park [12] Wei et al. [13] and Yang You et al. [14]. They have made notable contributions in addressing key challenges such as realism, generalization, and robustness in VR environments. These studies propose innovative methods, including combining lip synchronization techniques & deep learning, multi-view image translation, and high-quality animation synthesis systems using depth cameras, to improve the fidelity of avatars' expressions and their influence on social presence in collaborative VR settings. All three studies contribute to developing new tracking algorithms to create avatars' facial animations in VR, but they also highlight the challenges in achieving optimal balance between tracking accuracy, texture detail, and computational power. You et al. [14] proposed a virtual teaching system that captured the speaker's facial expressions and projected them onto an avatar in VR, but the speaker themselves is not immersed in VR. This creates a one-way interaction but limits its applicability in scenarios that require face-to-face communication.

Kullmann et al. [15] addressed the need for realistic and plausible facial expressions in avatars. They examined the importance of facial expressions in VR for natural and plausible experiences by comparing static, synthesized, and

Authorized licensed use limited to: Maynooth University Library. Downloaded on November 07,2026 at 15:14:45 UTC from IEEE Xplore. Restrictions apply.

tracked animation methods to create facial expressions. Their results demonstrated that people perceived the synthesized expressions as more natural than the tracked ones. Tracking inaccuracies and latency issues were affecting the realism and believability of tracked expressions. This underscores the importance of non-tracking methods to control avatars' facial expressions.

Kimmel et al. [16] found that avatars with eye and mouth movements elicited higher levels of co-presence and gaze duration, particularly during verbal explanation tasks. In line with this, Bonsch et al. [17] identified a relationship between the facial expression of a virtual agent, the use of comfort, and interpersonal distance in VR. They observed that users tend to stay farther away from angry agents than happy ones. The facial expressions of avatars also play a crucial role in trust building; as per the findings of Luo et al. [18] positive expressions significantly increased trust and cooperation, while negative expressions led to distrust. These findings reinforce the importance of expressive avatars in social VR environments, accentuating the need for effective control methods that allow users to convey emotions intuitively.

Baloup et al. [19] developed a non-isomorphic interaction technique to express facial expressions that uses emojis or symbols as part of facial expressions to express emotions. Their findings revealed that an excessive number of expression choices can overwhelm users, providing valuable information for our research. They employed more cartoonish avatars and integrated emojis with facial expressions for emoting, whereas we utilized higher-fidelity avatars with more human-like reactions. This is the sole study found to date that examines non-tracking methods for controlling an avatar's facial expressions.

These studies show the relevance of having expressive avatars and the significant progress in studying avatar facial expressions; many challenges remain to be tackled. High-end hardware alone cannot address the problem of achieving impactful facial expressions, with research showing that tracked expressions might not be as effective as previously thought and should be compared and studied against other methods. This highlights the need for exploring non-tracking control methods, such as those investigated in our paper, to provide effective and accessible solutions to animate avatar facial expressions.

III. Hypothesis

Previous works have suggested that VR affects the recognition accuracy of some emotions. Hypotheses 2 and 3 were created on the basis of Sweller's Cognitive Load Theory (CLT) [20] and Fitts' Law [21] respectively. CLT suggests that tasks involving memory recall instead of recognition place great cognitive demands on users, and Fitts's Law predicts that larger and nearer targets are more accessible and quicker to attain. Our hypotheses are as follows:

- **RQ1**
  - **H1:** VR has a significant effect on recognition accuracy of facial expressions.
- **RQ2**
  - **H2:** The control method without emotion labels will be significantly less accurate than the methods with labels.
  - **H3:** Control methods that require the use of only one hand will be significantly faster than the control method that requires both hands.

IV. Experiment Framework

We conducted two experiments to address the two separate yet interconnected aspects of VH-mediated communication. In our first experiment, we investigated how VR affects the perception of VHs' facial expressions, while the second experiment examined non-tracking control methods for manipulating avatar (embodied VHs) facial expressions in VR. The first experiment facilitated the validation of the facial expressions employed in the second experiment.

A. Experiment 1: Emotional Perception

We first conducted an experiment to understand how VR affects the perception of VHs' facial expressions by participants.

1) Stimuli & Procedure: The six basic universal expressions (happy, sad, anger, fear, disgust, and surprise) (see figure 1) were chosen as stimuli in this experiment [22] [23]. We created the stimuli based on Ekman and Friesen's research on facial cues [24]. The muscle cues associated with each facial expression were adjusted to create two variations of each emotion. This helped standardize the expressions and ensured that the expressions used for the studies were consistent.

The VHs were created using Ready Player Me and imported into Blender, where facial rigs were added to the models. The facial rigs were then used to move individual facial muscles to match the facial cues of each expression [24]. Rendered images of each expression were taken from the front view to be used as stimuli for the 2D screen, and 3D models with facial expressions were used as stimuli for VR. The expressive VHs were exported to Unity 2022.4.2. Unity was used to develop the experiment for 2D and VR deployments. We used a WebGL plugin to build the study for the screen and the Oculus Platform SDK for Unity to build the study for VR.

We used two VHs, a male and a female. They expressed six different emotions, each with two variations, resulting in a total of 24 combinations (images), with two repetitions of each condition. During the study (on screen and VR), participants viewed each facial expression produced by two VHs. Thus, each participant viewed a total of 48 images, arranged in random order. The stimuli were presented to them for five seconds as per previous studies [25], after which they answered the following questions:

- Q1. Which of the following emotions did that expression represent?
  - a.Happy b.Sad c.Fear d.Anger e.Disgust f.Surprise
- Q2. Please rate how effective the image was in representing the emotion you chose.

Authorized licensed use limited to: Maynooth University Library. Downloaded on November 07,2026 at 15:14:45 UTC from IEEE Xplore. Restrictions apply.

– a.1 b.2 c.3 d.4 e.5 (1 - highly ineffective, 5 - highly effective)

2) Outcome Measures: We examined two key measures to assess the perception and effectiveness of emotions displayed by VHs. They are in line with previous studies [26] [9] [27]. The measurements are as follows:

Perceived Emotion, the choice of emotion that participants believed the observed facial expression represented.

Perceived Effectiveness of Emotion Expression, the efficacy of the selected emotion as seen by the observer.

3) Participants: We implemented a between-subject design and recruited a total of 40 participants. Twenty participants (11 female, 9 male) aged between 22 and 66 years ($M = 36, SD = 12.84$) were recruited online via Prolific as participants for the 2D screen study. Using the university mailing system and word of mouth, we recruited 20 more participants (11 male and 9 female) aged between 18 and 31 ($M = 24.68, SD = 7.26$) to be part of the VR participant group. Before starting the experiment, participants received information about the study, instructions on what to expect, and their informed consent. The on-screen and VR studies took 12-15 minutes each. Participants who were recruited through Prolific for the on-screen study were compensated according to Prolific standards, and participants recruited for the VR study were compensated according to university standards. Ethical approval for this experiment was obtained from the university's ethics committee to ensure that all procedures met the required ethical standards.

B. Experiment 2: Facial Expression Control

We developed and evaluated three control methods to emulate existing non-tracking control methods in one of the most popular social VR platforms, VRChat [28]. These methods rely on the VR controller and do not use tracking similar to that in VRChat, representing a variety of widely used interface designs in social VR.

Raycast Selection (RS): This method allows users to control their avatar's facial expressions by pointing a ray from one hand and selecting expressions by pressing labeled buttons from a 3D menu attached to the other (see figure 2). This method was selected because it mimics commonly used UI paradigms, in which users interact through point-and-click mechanisms.

Thumbstick Control: A more dynamic method in which participants use the thumbsticks on VR controllers to select predefined facial expressions; a visual representation of the thumbstick is attached to the avatar's hand to aid in understanding the position of the thumbstick and the radial position of each facial expression. The position of each expression was determined by their positiveness; the most positive emotions (happy and surprise) were at the top, while the most negative ones (fear and anger) were at the bottom. Sad was placed next to happy, and disgust was placed next to surprise. There are two thumbstick-based controls. They are as follows:

Thumbstick Non-Labeled (TN): The positions of facial expressions within the thumbstick visualization are not labeled

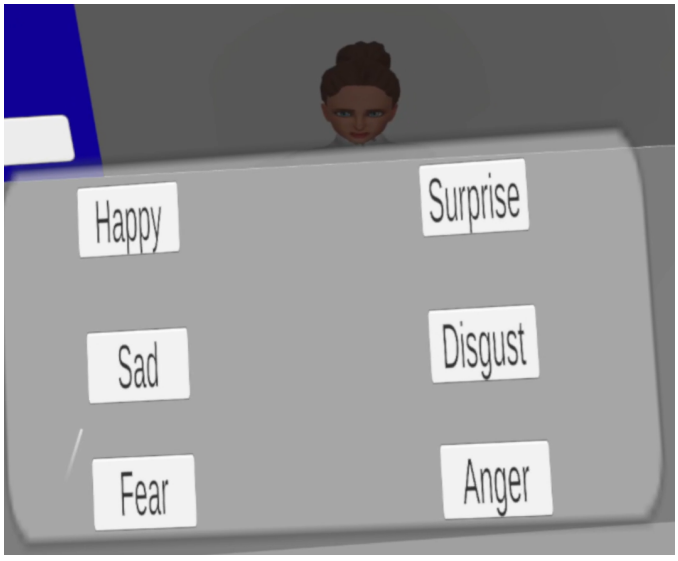


Fig. 2: Raycast Selection control method.

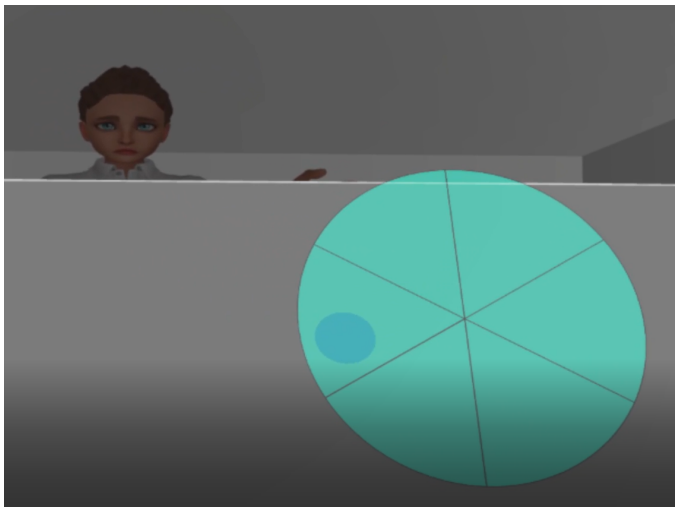


Fig. 3: Thumbstick Non-Labeled control method.

(see figure 3). This method has a low visual load design that requires only one hand and relies primarily on spatial memory and cognitive effort. This method represents use cases such as fast-paced social interactions where reliance on UI and labels is minimal, thus encouraging the user to fully engage in the interaction without any distraction.

Thumbstick Labeled (TL): The positions of facial expressions within the thumbstick visualization are labeled (see figure 4). This method preserves one-handed input while incorporating visual labeling to minimize memory load.

The order in which we presented the participants with each experiment block (control method) was semi-randomized. The order remained the same between thumbstick-based methods for all trials; TN was first, and TL was second. This was done to prevent learning effects between blocks. We randomized the order between thumbstick-based and raycast selection methods.

Authorized licensed use limited to: Maynooth University Library. Downloaded on November 07,2026 at 15:14:45 UTC from IEEE Xplore. Restrictions apply.

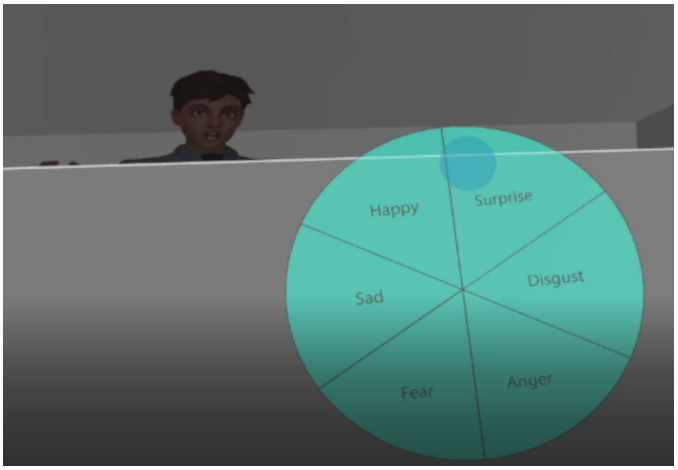


Fig. 4: Thumbstick Labeled control method.

1) Stimuli & Procedure: Upon arrival, participants completed a demographic questionnaire that also included a self-assessment of VR expertise. Participants embodied avatars corresponding to their gender, according to previous work [29] [30]. Before starting the experiment with a control method, participants received a two-minute tutorial on that control method. During the experiment, participants were instructed to look at their selected facial expression in a mirror placed in front of their avatar. A label (expression label) located on the side of the mirror displays the expression to be selected. After selecting an expression, the participant pressed the 'next' button on the expression label to display the next expression to be selected. The experiment was followed by a post-experiment questionnaire.

2) Outcome Measures: To evaluate the effectiveness of different non-tracking methods for controlling facial expressions, we measured quantitative and qualitative values that would give insight into the practicality of these control methods. The measures are as follows:

Selection speed, is the time taken from the time the expression label shows the expression to an expression chosen by the participant.

Accuracy, refers to the precision of the selection made by the participant.

Selection speed and accuracy were chosen as core performance indicators based on standard practices to evaluate interaction techniques in HCI. These metrics are in line with previous studies exploring non-tracking facial control methods in VR [19].

3) Participants: We implemented a within-subject design. Twenty participants were recruited, 11 men and 9 women aged 18 to 31 ($M = 24.68, SD = 7.26$), via the university's mailing system and word of mouth; The same participants from the VR section of the first experiment were also used for this experiment. Participants received information on the study and instructions on what to expect. Their informed consent was also obtained for this experiment. Each session lasted between 10-15 minutes, and participants were compensated for their time according to university standards. All participants were right-handed. Ethical approval for this experiment was obtained from the university's ethics committee to ensure that all procedures met the required ethical standards.

C. Data Collection and Logging

We recorded all quantitative and qualitative values in real time during the experiment in a cloud-based spreadsheet and a local device, respectively.

V. Results

For experiment 1, the data were ordinal and categorical, while for experiment 2, the data did not follow normality. For these reasons, non-parametric tests were used to evaluate the results for both experiments. Cramer's V was used to calculate effect sizes.

1) Emotional Perception: **Accuracy in emotion recognition:** We used a chi-square test to compare the recognition accuracy of each emotion in VR and on screen and applied the Bonferroni correction for Type I errors in multiple comparisons. The results revealed that there is a significant difference in the way participants perceived disgust; it has a higher recognition rate in VR ($\chi^2 = 7.569$, $p = 0.005$, V = 0.15). Further investigation of the impact of VH's gender on emotion recognition accuracy in VR showed that participants were significantly more accurate in identifying male VH's disgust($\chi^2 = 9.525$, $p = 0.002$,V = 0.24) than female VH's disgust. While female VH's fear ($\chi^2 = 11.512$, $p = 0.0006$,V = 0.26) and surprise ($\chi^2 = 5.4813$, $p = 0.019$,V = 0.18) were more accurately recognized compared to male VH's fear and surprise (see figure 5).

Effectiveness Rating: A Wilcoxon-Ranked Sum test revealed that during accurate emotion recognition, participants rated happy ($W = 800, p = 0.005$), surprise ($W = 8253, p = 0.026$) and sad ($W = 7851, p = 0.011$) significantly more effective in VR than on screen (see figure 6).

2) Facial Expression Control: A Shapiro-Wilk test was performed to assess the normality of the speed data in three control methods. The results indicated that the assumption of normality was violated for all methods. RS [$W = 0.6606$, $p <$ .001], TN [$W = 0.7490$, $p <$ .001], and TL [$W = 0.8417$, $p <$ .001]. Therefore, non-parametric tests were used for subsequent analyses. Pairwise comparisons were performed using Wilcoxon Signed-Rank tests, and Bonferroni correction was applied to account for multiple comparisons. Adjusted p-values are reported alongside original test results.

TABLE I: Wilcoxon Signed-Rank Test Results for Accuracy Comparison Between Control Methods with Bonferroni Correction

Comparison	W	p	Adjusted p
RS vs TL	67.0	0.140	0.420
RS vs TN	156.5	0.002	0.006
TL vs TN	96.0	0.006	0.020

Accuracy: Participants made the most accurate selections using the RS method. The average accuracy of participants was

Authorized licensed use limited to: Maynooth University Library. Downloaded on November 07,2026 at 15:14:45 UTC from IEEE Xplore. Restrictions apply.

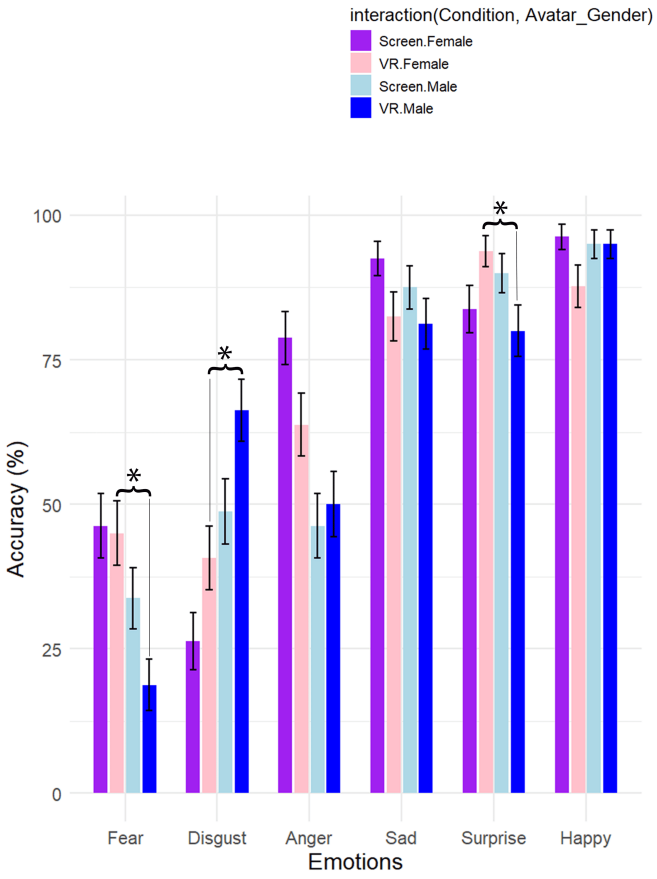


Fig. 5: Accuracy of emotion recognition per VH group.

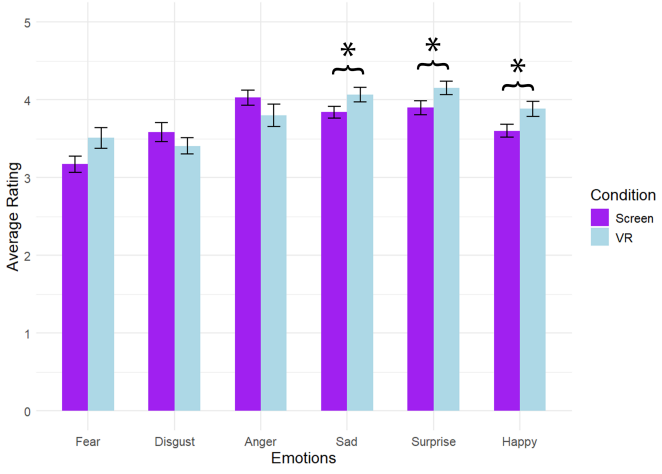


Fig. 6: Effectiveness rating for accurate emotion recognition between screen and VR.

TABLE II: Wilcoxon Signed-Rank Test for Speed Comparison Between Control Methods with Bonferroni Correction

Comparison	W	p	Adjusted p
RS vs TN	189	$p < .001$	$p < .001$
RS vs TL	190	$p < .001$	$p < .001$
TN vs TL	151	0.023	0.069

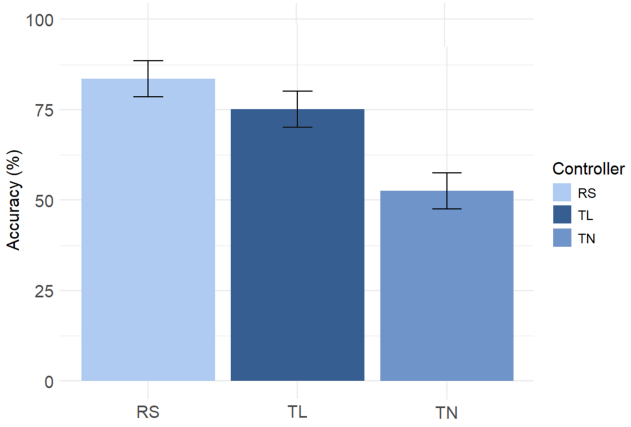


Fig. 7: Accuracy of each control method in percentage.

83.51% for RS, followed by the TL method at 75.2%, while the accuracy was lowest with the TN method at 52.6%(see table I)(see figure 7).

Speed: A Wilcoxon-Signed Rank test showed significant speed differences between the RS and both thumbstick options, with the thumbstick methods significantly faster than the RS method (see table II). However, the VR experience level of the participants was important in the TN method. The results show that experienced participants were 0.62 seconds faster than novices in making a correct selection, which was proven to be significant through a Wilcoxon-Ranked Sum test ($W = 1175$, $p = 0.011$). Participants were the fastest in making selections using the TL method (overall median speed = 1.38 s and median speed for correct responses = 1.39 s), followed by the TN method (overall median speed = 1.46 s and median speed for correct responses = 1.55 s). In contrast, the participants were the slowest with the RS method, with median speeds of 3.13 s overall and 3.06 s while being correct (see Figure 8).

User Preference: Of the participants, 65% preferred TL, 25% preferred RS and only 10% preferred the TN method.

VI. Discussion

Our study examined the impact of VR on facial expression recognition of VHs. A deeper analysis of individual emotions in our study revealed that recognition accuracy of disgust is significantly higher in VR, while Geraets et al. [9] reported disgust being harder to identify in VR compared to a 2D screen. This suggests that the manifestation of individual emotions in VR differs from traditional 2D screens, supporting our H1.

Authorized licensed use limited to: Maynooth University Library. Downloaded on November 07,2026 at 15:14:45 UTC from IEEE Xplore. Restrictions apply.

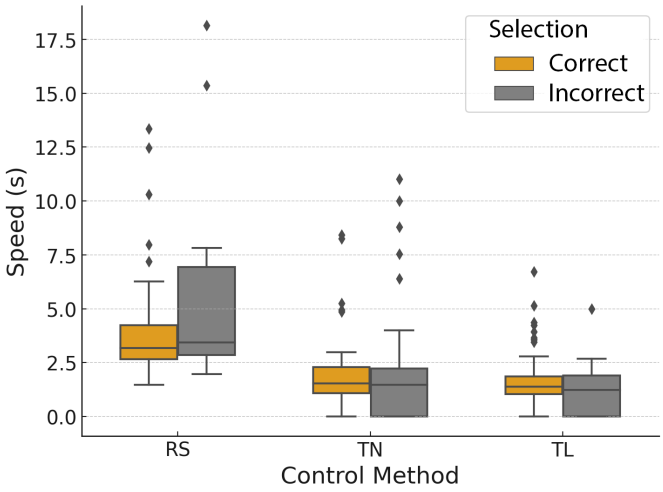


Fig. 8: Speeds of each control method while making correct vs incorrect selection.

Possible reasons for this discrepancy could be the higher visual fidelity of our VHs. Geraets et al. [9] used images of real humans for the 2D condition and VHs for the VR condition. They also added ambient street noise as a distraction for participants in their VR condition, and we did not include any such distraction in our experiments. Combined, these factors could have contributed to the difference in outcomes between their study and ours. A further analysis of the accuracy of each emotion and its relationship with avatar gender gave some more interesting results. Our results highlight that the accuracy of emotion recognition and its effectiveness are influenced by the gender of the virtual character. Specifically, male VH displaying disgust was more accurately recognized in VR compared to female VH displaying disgust. In contrast, female VH's expression of fear and surprise was more accurately recognized than male VH's in VR, Kapitanović et al. [31] found that emotion recognition was more accurate when the model was female, regardless of the emotion displayed. However, their study only included fear, sad, and anger in a 2D medium while employing a larger number of participants, while we included six emotions in 2D and VR with a smaller participant pool. This difference could be due to gender bias in emotion perception, as more dominant emotions such as anger and disgust are more associated with men, while submissive emotions such as fear and sadness are associated with women [32]. Although data analysis has shown that VR has an effect on emotion perception, the smaller effect sizes indicate that the magnitude of this effect is modest. An analysis of the effectiveness rating showed that happy, surprise, and sad are significantly highly rated in VR compared to the screen.

Thus, our study provides evidence that VR and the gender of VH have an impact on emotion perception. Disgust was more accurately recognized in VR, and easily recognizable emotions such as happy, surprise and sad were more amplified in VR. These results suggest that immersion provided by VR can influence the clarity and strength of certain emotions.

Using emotion expression as a task, we successfully compared 3 different non-tracking control methods regarding their usability by testing the speed, accuracy, and user preference. Our results provided strong evidence of a performance difference between the three methods. Participants were able to express emotions more accurately using the RS method, closely followed by TL. In contrast, participants performed poorly while using TN, achieving precision just about half the time, thus supporting H2. This outcome aligns with the Cognitive Load Theory [20], which suggests that tasks involving memory recall instead of recognition place great cognitive demands on users, thus making the task more prone to errors. TN requires participants to memorize and recall positions of expression, significantly increasing the cognitive load compared to RS and TL, which have labels of emotion acting as visual guidance. In a conversational context, searching for visual representations of controls may be disruptive; therefore, for long-term users, labels may become redundant once they have memorized the positions of all expressions, allowing them to intuitively manipulate the thumbstick to control their avatar's facial expressions without visual guidance. TL was the fastest overall, while RS, although more accurate, was the slowest, nearly twice as slow as TL. Consequently, this supports H3. Fitts' law [21] predicts that larger and nearer targets are more accessible and quicker to attain. Thumbstick-based controls exhibit a minimal distance from the origin to the destination, whereas RS methods require greater movement from the origin to the destination. Participants could improve reaction time if the size of the buttons of RS were larger in size. The participants preferred TL over RS despite RS's higher accuracy. The participants had a mirror in front of their avatar to check their facial expression after choosing one, but they still preferred the TL over the RS. This could be due to how TL requires minimal physical and cognitive effort, with only one hand needing to control it, while RS needs both hands, one hand to bring the 3D menu in front of them and another to point and click.

The importance of participants' VR experience was highlighted in the TN method. The novices were more accurate than the experienced participants when using this method, but the experienced users were significantly faster than the novices when making a correct selection. This could be because the novices were more meticulous in trying to make the correct choice. Experienced users, on the other hand, may have relied on muscle memory and quick action, potentially leading to overshooting or misjudgment. The cognitive load required for this method may make participants with better memory retention more accurate. So, by chance, novice participants in this group might have had better memory retention. The most preferred control method is TL, but developers may make it more accurate by adding a feature that requires the user to move the thumbstick to the target and hold it there to select an expression. This has the potential to decrease the selection speed while increasing the accuracy. The accuracy-speed trade-off here might not be as severe as in RS. Our results align with those of Baloup et al. [19] in demonstrating

Authorized licensed use limited to: Maynooth University Library. Downloaded on November 07,2026 at 15:14:45 UTC from IEEE Xplore. Restrictions apply.

the importance of intuitive, structured input methods. Our study looked only at menu-based selection methods, while they examined various interaction methods, including gestures and voice control. They found that menu-based techniques had faster selection times and lower error rates. However, a key distinction between our studies lies in the type of facial expressions used. We used subtle and more realistic human expressions, suitable for formal and informal social VR situations, while they incorporated cartoon-like augmentations to express emotions in a more stylized manner suitable for more informal situations.

VII. Limitations & Future Work

There are some limitations to our experiment. The viewing distance between the participants and the virtual human may have varied between the conditions on the screen and in VR; while the viewing distance was constant in VR, the viewing distance from the screen and the participant was uncontrolled as the study was conducted via an online platform. The participants in the screen-based experiments were older than those in the VR-based experiment. Previous research [33] has shown that older adults (senior citizens) tend to exhibit reduced accuracy in recognizing certain facial expressions, such as anger, but the difference in mean age between the groups for this study was more than 50, while the difference in mean age in our study is 11.32 (mean age of 36 and 24.68 for screen condition and VR condition, respectively). We only employed a single type of VH that had a non-photorealistic design with blendshape-driven expressions. Thus, our findings may not generalize to photorealistic, hyper-stylized avatars, nor are they generalizable to avatars of all ages and ethnicities.

The same participants participated in the VR section of experiment 1 and experiment 2. The control method, which involved pointing and clicking to answer pop-up questions in experiment 1, was similar to the RS control method used in experiment 2. This exposure could have affected their performance in the RS control method.

The study effectively measured speed, accuracy, and user preference in an isolated environment, but facial expressions are primarily used for social interactions. Future work will examine these control methods in an interactive social scenario.

VIII. Conclusion

We studied VH-based emotional interaction in VR by conducting a dual investigation into emotional perception and emotional expression of VHs. Our findings reveal that immersive environments not only amplify the perceived intensity of easily recognizable emotions but also improve the recognition accuracy for subtle expressions like disgust. However, these improvements in perception caused by VR depend on the gender of the VH as well. So, there is a need to consider the visual identity of a VH while designing visually expressive avatars. Our results also demonstrate that performance does not always match user preferences. The RS method yielded the highest accuracy, while TL was the fastest and most preferred. So, convenience rather than output was the popular choice.

A key contribution of our results lies in addressing both the user as observer and user as author roles in VH communication. By studying perception, we demonstrate that expression perception is influenced by the display system. Furthermore, by conducting a performance evaluation of three non-tracking control methods, we provide evidence of the impact of speed and accuracy trade-offs in user preference when controlling avatar facial expressions. Our research highlights that low-cost, non-tracking input methods can still facilitate emotionally rich VH communication. Thus, our work informs the design and development of VH-based communication systems in social VR platforms. Future research should explore dynamic viewing angles and social scenarios to improve expressive communication in social VR environments.

References

[1] C. Oh Kruzic, D. Kruzic, F. Herrera, and J. Bailenson, "Facial expressions contribute more than body movements to conversational outcomes in avatar-mediated virtual environments," *Scientific Reports*, vol. 10, no. 1, p. 20626, November 2020. [Online]. Available: https://doi.org/10.1038/s41598-020-76672-4

[2] M. Rapuano, T. Iachini, and G. Ruggiero, "Interaction with virtual humans and effect of emotional expressions: Anger matters!" *Journal of Clinical Medicine*, vol. 12, no. 4, p. 1339, February 2023.

[3] L. O. H. Kroczek, A. Lingnau, V. Schwind, and C. T. Haas, "Observers predict actions from facial emotional expressions during real-time social interactions," *Behavioural Brain Research*, vol. 442, p. 114689, 2024. [Online]. Available: https://epub.uni-regensburg.de/58590/1/1-s2.0-S0166432824002821-main.pdf

[4] N. Hube, K. Vidačković, and M. Sedlmair, "Using expressive avatars to increase emotion recognition: A pilot study," in *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, ser. CHI EA '22. Association for Computing Machinery, 2022. [Online]. Available: https://doi.org/10.1145/3491101.3519822

[5] C. Kuhne, E. D. Kecelioglu, S. Maltby, R. J. Hood, B. Knott, E. Ditton, F. R. Walker, and M. G. Kluge, "Direct comparison of virtual reality and 2d delivery on sense of presence, emotional and physiological outcome measures," *Frontiers in Virtual Reality*, vol. 4, 2023. [Online]. Available: https://www.frontiersin.org/journals/virtual-reality/articles/10.3389/frvir.2023.1211001

[6] Steam, "Steam hardware and software survey," Steampowered.com, 2024. [Online]. Available: https://store.steampowered.com/hwsurvey/Steam-Hardware-Software-Survey-Welcome-to-Steam

[7] P. Sykownik, L. Graf, C. Zils, and M. Masuch, "The most social platform ever? a survey about activities & motives of social vr users," in *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, 2021, pp. 546–554.

[8] F. Tian, M. Hua, W. Zhang, Y. Li, and X. Yang, "Emotional arousal in 2d versus 3d virtual reality environments," *PLoS One*, vol. 16, no. 9, p. e0256211, September 2021. [Online]. Available: https://doi.org/10.1371/journal.pone.0256211

[9] C. Geraets, S. Klein Tuente, B. Lestestuiver, M. van Beilen, S. Nijman, J. Marsman, and W. Veling, "Virtual reality facial emotion recognition in social environments: An eye-tracking study," *Internet Interventions*, vol. 25, p. 100432, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214782921000725

[10] D. Mal, E. Wolf, N. Döllinger, M. Botsch, C. Wienrich, and M. E. Latoschik, "From 2d-screens to vr: Exploring the effect of immersion on the plausibility of virtual humans," in *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '24. New York, NY, USA: Association for Computing Machinery, 2024. [Online]. Available: https://doi.org/10.1145/3613905.3650773

[11] F. D. Farizi, S. Bangay, and S. McKenzie, "Facial cues for deception detection in virtual reality based communication," in *Proceedings of the 3rd International Conference on Big Data and Internet of Things (BDIOT 2019)*. ACM, 2019. [Online]. Available: https://doi.org/10.1145/3361758.3361782

[12] J. Hwang and K. Park, *Full-Face Animation for a Virtual Reality Avatar*, 07 2023, pp. 199–206.

Authorized licensed use limited to: Maynooth University Library. Downloaded on November 07,2026 at 15:14:45 UTC from IEEE Xplore. Restrictions apply.

[13] S.-E. Wei, J. Saragih, T. Simon, A. W. Harley, S. Lombardi, M. Perdoch, A. Hypes, D. Wang, H. Badino, and Y. Sheikh, "Vr facial animation via multiview image translation," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 4, pp. 67:1–67:16, 2019.

[14] Y. You, L. Song, and Y. Yang, "High-quality facial expression animation synthesis system based on virtual reality," in *VR/AR and 3D Displays*, W. Song and F. Xu, Eds. Singapore: Springer Singapore, 2021, pp. 21–32.

[15] P. Kullmann, M. Botsch, T. Menzel, and M. E. Latoschik, "An evaluation of other-avatar facial animation methods for social vr," in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '23. Association for Computing Machinery, 2023. [Online]. Available: https://doi.org/10.1145/3544549.3585617

[16] S. Kimmel, F. Jung, A. Matviienko, W. Heuten, and S. Boll, "Let's face it: Influence of facial expressions on social presence in collaborative virtual reality," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23. New York, NY, USA: Association for Computing Machinery, 2023. [Online]. Available: https://doi.org/10.1145/3544548.3580707

[17] A. Bönsch, S. Radke, H. Overath, L. M. Asché, J. Wendt, T. Vierjahn, U. Habel, and T. W. Kuhlen, "Social vr: How personal space is affected by virtual agents' emotions," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2018, pp. 199–206.

[18] L. Luo, D. Weng, N. Ding, J. Hao, and Z. Tu, "The effect of avatar facial expressions on trust building in social virtual reality," *The Visual Computer*, vol. 39, pp. 5869–5882, 2023. [Online]. Available: https://doi.org/10.1007/s00371-022-02700-1

[19] M. Baloup, T. Pietrzak, M. Hachet, and G. Casiez, "Non-isomorphic interaction techniques for controlling avatar facial expressions in vr," in *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: https://doi.org/10.1145/3489849.3489867

[20] J. Sweller, "Cognitive load during problem solving: Effects on learning," *Cognitive Science*, vol. 12, no. 2, pp. 257–285, 1988. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1207/s15516709cog1202_4

[21] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *Journal of Experimental Psychology*, vol. 47, no. 6, pp. 381–391, 1954.

[22] P. Ekman and W. V. Friesen, *Unmasking the Face: A Guide to Recognizing Emotions from Facial Expressions*. Cambridge, MA: Malor Books, 1975.

[23] P. Ekman, "An argument for basic emotions," *Cognition & Emotion*, vol. 6, pp. 169–200, 1992. [Online]. Available: https://api.semanticscholar.org/CorpusID:11771973

[24] P. Ekman and W. V. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto, CA: Consulting Psychologists Press, 1978.

[25] G. Fairchild, Y. Stobbe, S. H. M. van Goozen, A. J. Calder, and I. M. Goodyer, "Facial expression recognition, fear conditioning, and startle modulation in female subjects with conduct disorder," *Biological Psychiatry*, vol. 68, no. 3, pp. 272–279, August 2010. [Online]. Available: https://doi.org/10.1016/j.biopsych.2010.02.019

[26] C. Ennis, L. Hoyet, A. Egges, and R. McDonnell, "Emotion capture: Emotionally expressive characters for games," in *Proceedings of Motion on Games*, ser. MIG '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 53–60. [Online]. Available: https://doi.org/10.1145/2522628.2522633

[27] K. Limbrecht-Ecklundt, A. Scheck, L. Jerg-Bretzke, S. Walter, H. Hoffmann, and H. C. Traue, "The effect of forced choice on facial emotion recognition: a comparison to open verbal classification of emotion labels," *Psycho-social Medicine*, vol. 10, p. Doc04, 2013, published online 2013 Jun 17. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3687244/

[28] D. Maloney, G. Freeman, A. Robb, H. Pfotenhauer, and M. Slater, "Social virtual reality as a mental health tool: How people use vrchat to support social connectedness and wellbeing," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–13. [Online]. Available: https://dl.acm.org/doi/10.1145/3544548.3581103

[29] T. D. Do, C. I. Protko, and R. P. McMahan, "Stepping into the right shoes: The effects of user-matched avatar ethnicity and gender on sense of embodiment in virtual reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, no. 5, pp. 2434–2443, 2024.

[30] V. Schwind, P. Knierim, C. Tasci, P. Franczak, N. Haas, and N. Henze, ""these are not my hands!": Effect of gender on the perception of avatar hands in virtual reality," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 1577–1582. [Online]. Available: https://doi.org/10.1145/3025453.3025602

[31] A. Kapitanović, A. Tokić, and N. Šimić, "Differences in the recognition of sadness, anger, and fear in facial expressions: The role of the observer and model gender," *Archives of Industrial Hygiene and Toxicology*, vol. 73, no. 4, pp. 308–313, 2023. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9985350/

[32] E. A. Plant, J. S. Hyde, D. Keltner, and P. G. Devine, "The gender stereotyping of emotions," *Psychology of Women Quarterly*, vol. 24, no. 1, pp. 81–92, 2000. [Online]. Available: https://doi.org/10.1111/j.1471-6402.2000.tb01024.x

[33] T. Ruffman, J. D. Henry, V. Livingstone, and L. H. Phillips, "A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging," *Neuroscience & Biobehavioral Reviews*, vol. 32, no. 4, pp. 863–881, 2008. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2859895/

Authorized licensed use limited to: Maynooth University Library. Downloaded on November 07,2026 at 15:14:45 UTC from IEEE Xplore. Restrictions apply.