Check for updates

Computer Animation and Virtual Worlds

WILEY

SPECIAL ISSUE CASAXR 26: International Conference on Computer Animation, Social Agents, and Extended Reality

SPECIAL ISSUE PAPER OPEN ACCESS

Making Faces: Evaluating Facial Control Methods in VR for Live Conversations

J. K. Sangeeth Chandran[1,2] | Marisa Llorens Salvador[1] | Cathy Ennis[2,3]

[1]School of Computer Science, Technological University Dublin, Ireland | [2]ADAPT Research, Ireland, Ireland | [3]Department of Computer Science, Maynooth University, Ireland

Correspondence: J. K. Sangeeth Chandran (sangeeth.jayachandrankala@tudublin.ie)

Received: 17 April 2026 | **Revised:** 12 May 2026 | **Accepted:** 14 May 2026

Keywords: avatars | control methods | controller-based interaction | co-presence | face tracking | facial expression | nonverbal behavior | virtual reality | VR

ABSTRACT

Facial expressions are a crucial part of affective communication in our daily lives, and the same applies to virtual humans. However, achieving a full range of believable and naturalistic emotional expression via VR devices still remains a challenge, particularly when constrained by the capabilities of midrange consumer VR devices without face tracking as opposed to the more expensive versions with facial tracking capabilities. In this study, we evaluated and compared three methods to control avatar facial expression in VR: continuous face tracking (FT); thumbstick label (TL), which is a VR controller's thumbstick-based selection method with visually labeled expression; and thumbstick label with confirmation button (TB), which is a variation of TL with an additional step of pressing a button. Thirty participants took part in the experiment. Participants found FT to be significantly more usable than TL and TB. While TL was comparable to FT on co-presence and several items of expressiveness, with significant gaps emerging mainly for naturalness and smoothness in expression change. These findings suggest that a well-designed control method with low latency and low motor and cognitive load can serve as an alternative method of avatar control in VR devices without facial tracking technology.

1 | Introduction

In virtual reality (VR), virtual humans (VHs) serve as the primary tool for social interaction, usually either via avatars (embodied and controlled by the user) or agents (autonomous and semi-autonomous). Through embodiment, VHs enable users to express their identity, communicate intent, and engage emotionally with others in virtual spaces [1, 2]. However, achieving a full range of believable and naturalistic emotional expression via VHs still remains a challenge, particularly when constrained by the capabilities of midrange consumer VR devices without the capacity for face tracking.

Facial expression plays a critical role in emotional communication and social understandings. As a nonverbal cue, facial expression helps to convey affect, build trust, and create interpersonal dynamics [3–5]. Replication of the richness of basic human facial expressions in VR using avatars is a challenging task. As a result, most VR users are compelled to rely on alternative methods such as controller-based interfaces or indirect mappings such as voice-based controls to convey emotions through avatars' facial expressions.

These gaps present a challenge that requires the design of effective non-tracking avatar controls for facial expression in VR. Our study addresses this gap by comparing three methods to control avatar facial expression: continuous face tracking (FT), controller thumbstick-based selection with visual labels (TL), and controller thumbstick with an additional confirmation button (TB).

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Computer Animation and Virtual Worlds* published by John Wiley & Sons Ltd.

These control methods are compared against each other in terms of co-presence, usability, and expressiveness, as these measures capture complementary aspects of social interaction and human-computer interaction. Co-presence reflects the perceived sense of social connection and awareness. Usability evaluates the effectiveness and interaction effort associated with each control method. Expressiveness assesses the extent to which users can successfully convey and perceive emotions using the control methods. Which leads to our research question:

RQ. *How do non-tracking control methods compare to face tracking in natural conversations for facilitating presence, usability, and expressiveness?*

Through this study, we aim to investigate whether the thumbstick-based methods can be used in mid-range consumer-grade VR devices and still produce face-tracking-like effects. The findings from this study can serve as practical guidelines for user experience and user interaction designers as well as developers working with social virtual reality (SVR).

2 | Related Work

Facial expressions have always been identified as extremely important for interpersonal communication and have been shown to contribute more to emotional outcomes than body cues [3, 6]. Studies comparing different avatar heads and expression levels in mixed or virtual reality similarly find that richer or higher-intensity facial animation can enhance perceived performance and influence aspects of social presence and user experience, although not always uniformly across tasks [7, 8].

2.1 | Co-Presence and Trust

Experiments on collaborative tasks have shown that animated eye and mouth movements, when integrated with verbal explanations and realistic avatars, increase social presence and co-presence [7, 9]. Avatars with tracked facial expressions have been shown to improve task performance, social presence, and interpersonal attraction compared to avatars with low expressiveness [10]. Facial expression also influences trust and cooperation with positive expressions when combined with lip-sync. This increases trust in the avatar and improves the likelihood of collaboration. Negative expressions were found to have the opposite effect of positive expressions, while expressionless avatars were less impactful in decision-making tasks [5].

2.2 | Control Methods for Avatar Facial Expressions

Three main approaches exist for controlling avatar facial expressions in SVR. (1) Face tracking, (2) expression generation from speech/audio, and (3) using manual controls. Face tracking attempts to mirror the user's own facial expressions through infrared/RGB-D sensors [11] or physiological sensors such as electromyography (EMG) or electrooculography (EOG); these estimates, combined with tracking software estimates, blend shape parameters or facial landmarks in real time [12]. However, they require dedicated hardware and careful calibration and are sensitive to occlusion, lighting, and tracking instability, especially on consumer devices [10, 13]. Devices integrated with facial tracking sensors are expensive and are not as used or accessible as mid-range devices without face tracking.

A second approach to implementing facial expressions is through synthetic generation of expressions from audio and speech context, where the system infers lip motion and broader facial expressions directly from the user's voice and sometimes additional cues such as text content or gaze [14–16]. Deep learning models can map raw waveforms or spectrograms to 3D facial vertices or avatar blendshapes, jointly modeling speech articulation and emotion to produce expressive co-speech animation in real time. These audio-driven methods are attractive to users because they do not require additional hardware or special attention during an SVR session, as expressions are automated and not something that needs explicit user attention. Yet user studies have shown that these methods lag behind facial tracking for conveying subtle and negative expressions like sadness and disgust and often focus more on the lip-sync through mouth articulation than the full face [14, 17, 18]. They also falter when the tempo of the speech exceeds a certain limit [16]. Users have also been shown to prefer manual control methods over audio methods due to the high error rate in expression and low usability and satisfaction [19].

Finally, manual control of avatar facial expressions usually uses the VR controller and treats it as an explicit interaction task, mapping thumbsticks, buttons, or interaction menus to a set of expressions whose timing and intensity are explicitly controlled by the user [13, 19, 20]. Previous studies have shown that carefully designed interfaces allow quick selection and continuous modulation while speaking, and user studies identify methods involving thumbstick-based control methods as both efficient and preferred. They have pointed towards a speed and accuracy trade-off when choosing facial expression control methods, as a quicker method of selection could have less accuracy in the selection of expression, and they also recommended adding a confirmation step to increase the accuracy of manual controls [20]. It has also been advised to have a lower number of expressions as choices to be given to users, since having too many options was found to increase selection time and error rate [19].

From a cognitive perspective, operations with interfaces, such as the selection of expression and the confirmation of selection, may introduce additional decision-making and monitoring processes. Thus, increasing cognitive load by consuming working memory resources, particularly when performed alongside a primary task such as conversation [21]. In terms of motor load, human-computer interaction theories such as Fitt's law [22] and the keystroke-level model [23] predict that additional input operations increase movement time and motor effort. Prior work on avatar facial expressions has often evaluated control techniques in isolated, task-driven, or highly constrained scenarios, such as recognizing emotions or performing specific interaction tasks. These settings provide experimental control but do not fully capture the demands of live, free-flowing social interaction, where timing, motor load, and cognitive load are critical. As a result, it remains unclear how different facial expression control methods affect users' experience of co-presence, the usability of the system, and the perceived naturalness and expressiveness of communication when people simply converse with each other in VR.

1546427x, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cav.70137 by National University Of Ireland Maynooth, Wiley Online Library on [07/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3 | Methodology

In order to determine the co-presence, usability, and expressiveness of three methods to control avatar facial expressions, we conducted a study that employed a within-subjects design. Prior studies have focused on isolated and task-driven environments to evaluate avatar facial expression control methods in VR. Here we use dyadic live interactions between participant pairs to measure how avatar facial expression control methods perform in terms of co-presence, usability, and expressiveness. This examines the efficacy of facial expression control methods to be used in an uncontrolled and unscripted free-flowing dyadic conversation. Our research focus is more on intentional user-controlled input methods, which is why we did not consider audio-based controls in this study, as they rely more on system inference.

Three facial expression control methods were compared: (1) continuous face tracking, (2) intentional thumbstick-based selection with visible labels, and (3) thumbstick-based selection requiring explicit confirmation (Figure 2). The design was intended to isolate the effects of interaction friction and intentional control on social presence, usability and perceived expressiveness.

Through this evaluation, we aim to provide design insights on how facial expression control methods should be deployed to support natural and emotionally engaging social interactions in Social VR (SVR).

3.1 | Control Methods

We examined three different control methods in this experiment; they are as follows:

3.1.1 | Face Tracking (FT)

In the FT method, facial expressions were animated using real-time facial tracking provided by Meta Quest Pro. Facial movements were mapped directly to the avatar's 72 facial blendshapes, enabling spontaneous and continuous expression during conversation.

3.1.2 | Thumbstick Labeled (TL)

In the TL method, participants used the controller's thumbstick to select facial expressions from a labeled radial interface. Directional input from the thumbstick corresponded to predefined expressions labeled in the UI, allowing participants to select expressions without additional confirmation steps. Expressions were activated immediately upon the movement of the thumbstick and would return to the neutral position upon release of the thumbstick. TL was selected as a lightweight non-tracking control method because prior VR work has argued that avatar facial-expression control in social settings should be fast, easy to access, and minimally disruptive to the primary task of talking [19], and TL meets these criteria based on prior work [20].

3.1.3 | Thumbstick Label With Confirmation Button (TB)

The Thumbstick Label with Confirmation Button used the same thumbstick-based selection mechanism as TL but required participants to press an additional button (the grab button on the controller using the middle finger) to confirm expression selection. Holding the button locked the expression in place and releasing the button reverted the expression to neutral. TB was included as a design variation of TL to reduce the selection error with the thumbstick, which was an issue reported in prior work [20].

Thumbstick conditions were augmented by the use of visemes to provide lip-sync to the dialogues. The thumbstick conditions offered six expressions: happy, surprise, sad, anger, fear, and disgust, which are designated as the basic human expressions by Ekman [24] (Figure 1). Under thumbstick conditions, the available expressions were restricted to six, as excessive options can lead to performance difficulties due to heightened cognitive load [19] and the six basic human expressions were chosen as the default set of expressions. FT provided a wide range of expressions that were only limited by the users' facial expressiveness; we decided not to limit the expression set of the FT in order to accurately represent the practical use case of this method.

3.2 | Apparatus and Virtual Environment

The experiment was created in Unity, version 6.0; the face tracking was applied using Meta's Movement SDK. The character for the experiment used was the sample realistic character for the Movement SDK provided by Unity. A brief training phase preceded the main experiment and the character for the training phase was created using Reallusion's Character Creator 4. Facial expressions for TL and TB were created using the blendshapes/action units as recommended by the facial action coding system (FACS) [25].

The study was conducted using Meta Quest Pro, with an HMD and two controllers. Participants occupied two separate rooms but shared the same virtual space, where each participant embodied a full-body avatar. Both participants embodied the same virtual character. This was to minimize any confounding effects and to provide the same perception for both participants, and the participants were unaware that they were using the same character during the experiment. A different character was used for the training phases to avoid any self-association effects that might occur if participants later observed their own avatar being embodied by their partner. Participants were positioned facing each other at a fixed interpersonal distance (2 units in unity) within the virtual environment, which remained constant throughout all conditions (Figure 2).

3.3 | Procedure

Within-subjects design was chosen to control for individual variability in expressiveness across all control methods. Each experimental session involved one dyad and lasted ~1 h. For each control condition, participants first completed a two-minute training phase immediately before the corresponding

1546427x, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cav.70137 by National University Of Ireland Maynooth, Wiley Online Library on [07/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

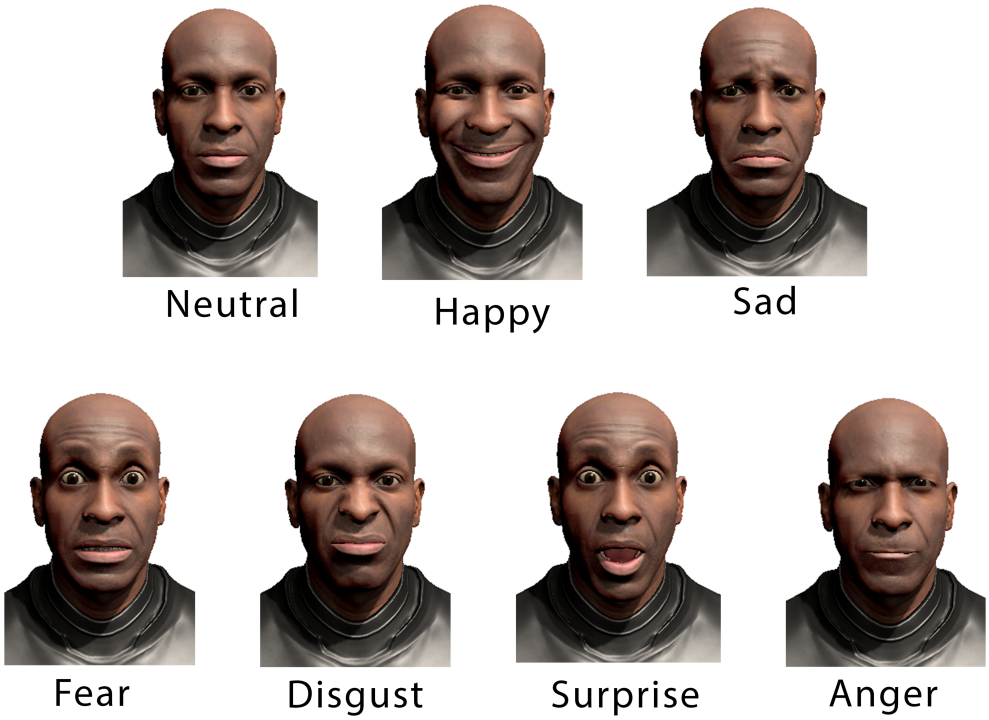


FIGURE 1 | Neutral and six basic human expressions.

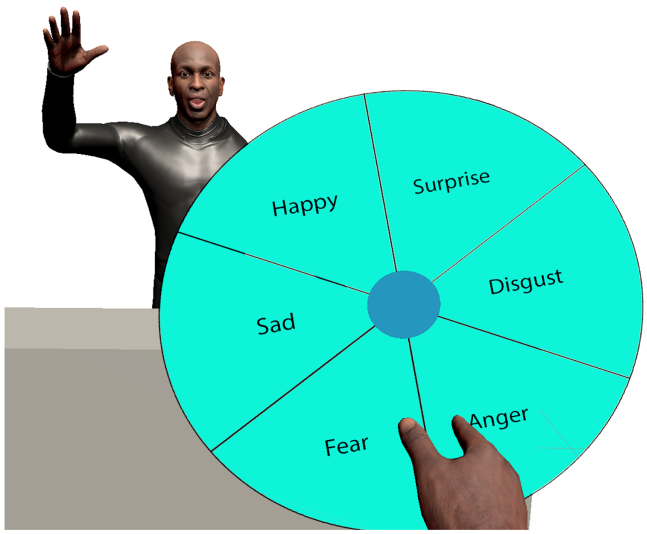


FIGURE 2 | Experiment phase: Interaction using TL method for avatar facial expression control (point of view of the participant).

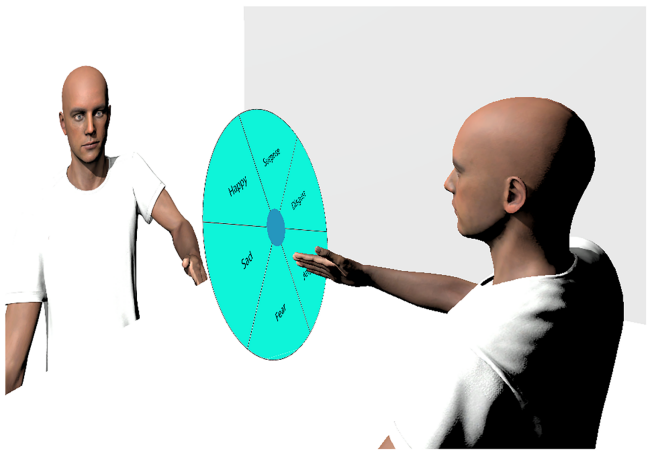


FIGURE 3 | Training phase (side view).

experimental phase to get familiarized with each control method. This phase involved getting familiar with control methods in front of a virtual mirror (Figure 3). The training phase was followed by a free-flowing dyadic conversation lasting up to 10 min. Conversation topics consisted of three open-ended prompts designed to encourage more natural dialogue than task-oriented interaction:

- Is AI good or bad for society? Express your opinion.
- How has social media changed our society? Express your opinion.
- Do yoga and meditation have any actual benefits? Express your opinion.

The order of control methods and conversation topics was counterbalanced across dyads using a Latin-squared design to mitigate order effects.

3.4 | Participants

Thirty participants (all male, aged 18–38 years) took part in the study, forming 15 dyads; all male participants were chosen due to the male avatar being used. Dyads were chosen in a way so that both participants in a dyad had little to no prior relation with each other. All participants provided informed consent prior to

1546427x, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cav.70137 by National University Of Ireland Maynooth, Wiley Online Library on [07/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 | Expressiveness questionnaire items.

Item	Statement
Q1	My partner's facial expressions were clear and distinguishable.
Q2	My partner's expressions felt natural.
Q3	My partner's expression changes appeared smooth.
Q4	My partner seemed to understand the emotion I was expressing.
Q5	The interaction felt emotionally authentic.
Q6	I felt like their avatar's expressions matched their emotions.
Q7	My partner responded appropriately to my expressions.
Q8	I felt that I was genuinely expressing myself through the avatar.

participation and the study protocol was approved by the institutional ethics committee.

3.5 | Measures

Following each experimental block/condition, participants completed questionnaires to assess the following measures:

- Co-presence: Measured using a set of 7-point Likert-scale items using the co-presence section of the Networked Minds Social Presence Inventory [26], focusing on the feeling of shared space and mutual awareness.
- Usability: Assessed using 5-point Likert-scale items adapted from the VR System Usability Questionnaire [27], capturing perceived ease of use, effort and control clarity.
- Expressiveness: Evaluated using custom 5-point Likert-scale items measuring perceived multiple measures of expressiveness (Table 1). A 5-point Likert scale was chosen over a 7-point scale to reduce the cognitive load of the study. The questions were phrased as agreement-based declarative statements. The responses recorded ranged from *1—strongly disagree* to *5—strongly agree*.

4 | Results

All data were subjected to a Shapiro–Wilk normality test before analysis, and the Friedman test was used to analyze non-normal data, and effect sizes were calculated using Kendall's W Where significance was observed, pairwise comparisons were done using the Wilcoxon signed-rank test, and effect sizes are reported alongside.

4.1 | Co-Presence

We found (Figure 4 and Table 2) that FT sustained co-presence the most (based on mean ratings), but analysis indicates that there was no significance between FT and TL while both FT and

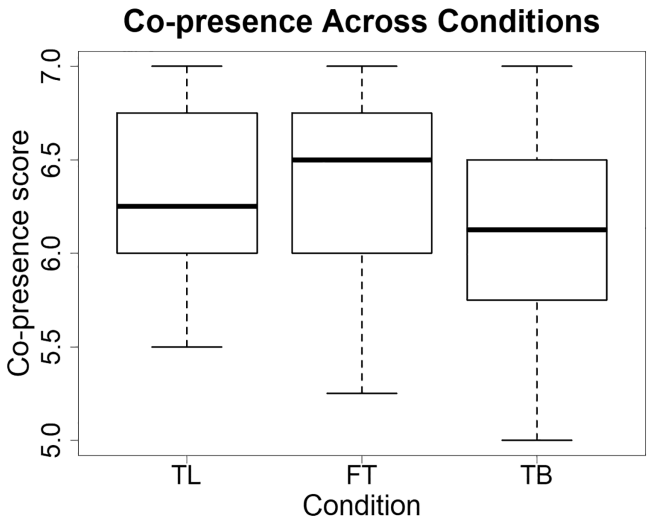


FIGURE 4 | Mean co-presence rating among the three control methods.

TABLE 2 | Post hoc Wilcoxon signed-rank test results for co-presence.

Comparison	W	p_{Holm}	r	Sig.
TL vs. FT	15	0.057	0.35	No
TL vs. TB	28	0.033	0.44	Yes
FT vs. TB	65	0.013	0.52	Yes

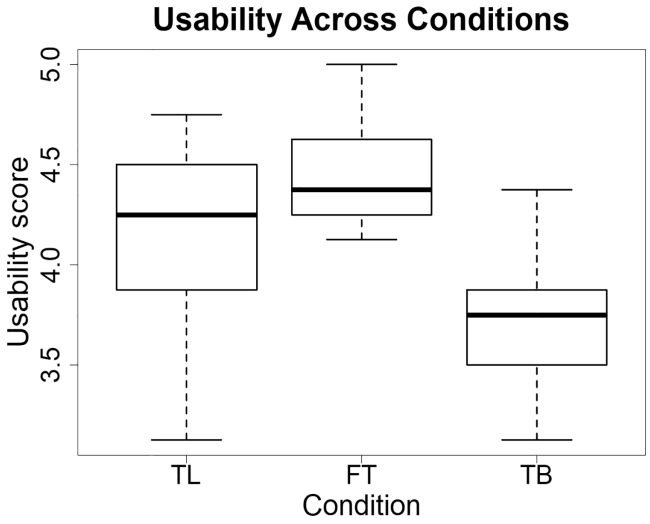


FIGURE 5 | Mean usability rating among the three control methods.

TL enhanced the sense of co-presence significantly more than TB. Participants reported high levels of co-presence across all control methods. A Friedman test revealed a significant effect of the control method, $\chi^2(2) = 10.09$, $p = 0.006$, with a small effect size (Kendall's $W = 0.17$). Post hoc Wilcoxon signed-rank tests with Holm correction showed:

4.2 | Usability

The results (Figure 5 and Table 3) show that FT was perceived as significantly more usable than both TL and TB. TL was rated significantly more usable than TB. A Friedman test revealed a highly

1546427x, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cav.70137 by National University Of Ireland Maynooth, Wiley Online Library on [07/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 | Post hoc Wilcoxon signed-rank test results for usability.

Comparison	W	p_{Holm}	r	Sig.
TL vs. FT	71	0.004	0.52	Yes
TL vs. TB	425	0.013	0.82	Yes
FT vs. TB	459	< 0.001	0.85	Yes

TABLE 4 | Post hoc Wilcoxon signed-rank test results for expressiveness.

Comparison	W	p_{Holm}	r	Sig.
TL vs. FT	75.5	0.019	0.43	Yes
TL vs. TB	382.5	< 0.001	0.75	Yes
FT vs. TB	408.0	< 0.001	0.75	Yes

significant effect of the control method in usability $\chi^2(2) = 37.78$, $p < 0.001$, with a large effect size (Kendall's $W = 0.63$). To further examine pairwise differences, Wilcoxon signed-rank tests with Holm correction were conducted:

4.3 | Expressiveness

Expressiveness was assessed using an 8-item custom questionnaire 1; we analyzed each item individually due to the questionnaire being a custom one and not standardized. This also enabled us to get a more nuanced insight into the effectiveness of each control method in achieving expressiveness. A Friedman rank-sum test revealed a significant effect of control method on expressiveness, $\chi^2(2) = 25.77$, $p < 0.001$, with a medium-to-large effect size (Kendall's $W = 0.43$). To determine which control methods differed significantly across expressivity-related measures, Wilcoxon signed-rank tests were conducted for all pairwise comparisons, with Holm correction applied to control for multiple testing (Table 4).

The results (Figure 6, Tables 5 and 6) show that FT performed consistently better than TL and TB. By observing individual measures, it is notable that TL was comparable to FT in every measure except naturalness and smoothness in change of expression, while TB constantly underperformed in the majority of the measures.

5 | Discussion

Our study investigated how facial expression control methods perform and influence a dyadic interaction in terms of co-presence, usability, and expressiveness. By situating the experiment in a free-flowing conversation in VR, we extend the prior works focusing on isolated or task-based environments and examine how facial expression control methods perform under the attentional demands of natural social interaction. The three methods differ in how intentional the expression production is. FT supports a continuous and wide range of expressions with varying intensities, which are limited only by the expressiveness of the user's face. While thumbstick-based methods require

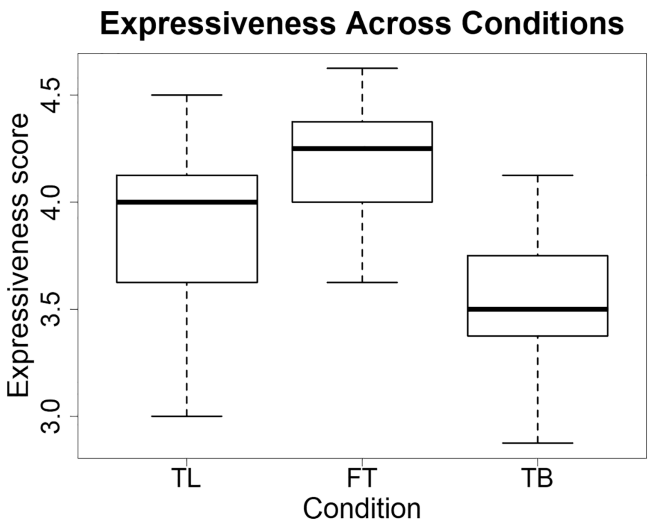


FIGURE 6 | Mean expressiveness rating among the three control methods.

TABLE 5 | Friedman tests for individual expressiveness items (Q1–Q8) across control methods with Holm's correction applied.

Item	$\chi^2(2)$	p_{Holm}	Kendall's W	Sig.
Q1	2.06	0.358	0.03	No
Q2	27.79	< 0.001	0.46	Yes
Q3	24.60	< 0.001	0.41	Yes
Q4	0.97	0.617	0.01	No
Q5	25.07	< 0.001	0.42	Yes
Q6	15.76	< 0.001	0.26	Yes
Q7	7.00	0.030	0.12	Yes
Q8	2.80	0.247	0.05	No

deliberate selection of only the six basic facial expressions from a radial menu, with only one level of intensity. Our results suggest that, although FT provides the most natural and seamless communication experience, a manual control method such as TL with low interaction friction (low cognitive and motor load) can achieve comparable levels of co-presence and communicative clarity, making it a viable alternative for VR devices without facial tracking.

5.1 | Co-Presence

The participants reported that co-presence for TL is comparable to that of FT, while TB was rated significantly lower for co-presence. This indicated that continuous tracking is not a necessity when it comes to achieving a high sense of mutual awareness and shared space in SVR. A heightened co-presence can be achieved and maintained using other methods of facial expression controls as long as there are not too many steps in controlling the expressions. Having multiple steps for selection of an expression would increase the motor load and could take the attention away from the person at the other end and towards controlling your own expressions. Here, we suspect that the additional step of confirming the selected expression while using TB

1546427x, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cav.70137 by National University Of Ireland Maynooth, Wiley Online Library on [07/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 6 | Post hoc Wilcoxon signed-rank tests for expressiveness items with significant Friedman effects. Holm-corrected p-values and effect sizes (r) are reported.

Item	Comparison	W	p_{Holm}	r	Sig.
Q2	TL vs. FT	6.0	< 0.001	0.65	Yes
	TL vs. TB	15.0	0.317	0.18	No
	FT vs. TB	223.5	< 0.001	0.71	Yes
Q3	TL vs. FT	47.5	0.009	0.52	Yes
	TL vs. TB	132.5	0.009	0.51	Yes
	FT vs. TB	313.0	< 0.001	0.75	Yes
Q5	TL vs. FT	7.5	0.260	0.21	No
	TL vs. TB	136	< 0.001	0.66	Yes
	FT vs. TB	149.5	< 0.001	0.64	Yes
Q6	TL vs. FT	36.5	0.082	0.32	No
	TL vs. TB	251.5	0.029	0.45	Yes
	FT vs. TB	213.0	0.002	0.63	Yes
Q7	TL vs. FT	1.0	0.276	0.20	No
	TL vs. TB	10.0	0.127	0.34	No
	FT vs. TB	15.0	0.102	0.39	No

introduced a friction during interaction that broke the flow and rhythm of a free-flowing conversation, likely making the partners feel less jointly engaged.

5.2 | Usability

Usability ratings given by participants for each control method show a clear hierarchy. FT was the best rated, followed by TL and then TB, the worst rated among all three. FT's advantage likely stems from its direct mapping of natural facial movements to avatar expressions, eliminating the need for conscious selection, thus having no motor or cognitive load on the users. TL and TB, while expressing the same predefined expressions, had differing usability ratings, likely due to the additional motor and cognitive load associated with TB compared to TL and the delay in expressions. This aligns with existing knowledge that in social contexts, even small interaction costs can be important because they compete with attentional resources needed for listening, planning speech, and monitoring the partner's reactions [28].

5.3 | Expressiveness

Participant's ratings of the Expressiveness Questionnaire reveal a more nuanced pattern. FT was consistently rated higher for naturalness of expression and smoothness of expression change. These results underline the importance of continuous fine-grained facial motion for conveying subtle affective cues that are lacking in thumbstick control methods [19, 29], thus enabling users to feel embodied and in emotional sync with their avatars.

However, despite the expressive richness that FT offers, on several items of expressiveness, TL was comparable to FT, including clarity of expressions, emotional authenticity of the interaction, and partner's understanding of emotions, with significant gaps emerging mainly for naturalness and smoothness of change in facial expression. This points towards an idea that, in the context of unstructured, opinion-based conversations, a small, well-chosen set of expressions may be sufficient to support communication and mutual understanding [19, 20], while the added expressive nuances of FT mainly enhance subjective experience (feeling natural and smooth) rather than basic communicative success, which can exchange ideas and basic emotions.

Interestingly, although TL and TB offered identical expressions, TB here has scored lower in almost every metric compared to FT and TL, except in clarity of expression (Q1).

5.4 | Expression Detail Versus Expression Timing

The lack of significant differences across all control methods for expression clarity, emotional response, and perception of partner's responsiveness suggests that recognizability of expression states can be preserved even when expressive flow and smoothness are reduced.

It is also important to note here that TB, despite having the exact same expressions as TL, constantly underperformed compared to TL in co-presence, usability, and multiple measures of expressiveness, which points to the timing and syncing of expressions being critical factors in a conversational context over the accuracy of expressions. This indicates a design choice that SVR systems should prioritize low-latency and low-step methods for facial expression controls more than purely visual details in expressions. The confirmation step for TB was introduced to improve the accuracy of selection; however, our findings suggest that in conversational contexts, the timing coordination between speech and facial expression may be more critical than accuracy. Conversational meanings heavily rely on the synchronization of verbal

1546427x, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cav.70137 by National University Of Ireland Maynooth, Wiley Online Library on [07/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and nonverbal cues; even the extra step of pressing a button may reduce perceived clarity and authenticity [22].

6 | Limitations and Future Work

We have only evaluated the performance of the control methods in a realistic-looking male avatar and male participants, so our results may not generalize across all avatar genders and user genders. Our participant pool contained strangers or acquaintances who had opinion-based yet polite conversations under 10 min. This resulted in the underutilization of more intense and negative expressions like anger and fear. Additionally, our study focused on the subjective experience of the participants on expressiveness and interaction quality; we did not utilize objective measurements.

Future studies should use a diverse participant pool under different social contexts and different relationship dynamics between users over longer periods of interaction that are more emotionally charged. Future researchers could examine the frequency of each expression used and the duration of expressions to better understand how control methods influence conversational dynamics.

7 | Conclusion

We evaluated and compared three methods to control avatar facial expression in VR in a free-flowing dyadic conversation in VR. Investigating beyond isolated or task-based evaluations, our findings demonstrate that factors influencing interaction quality in VR are not only the expressive richness but also the interaction mechanics that govern how expressions are implemented or deployed.

Crucially, introducing additional interaction steps may degrade the quality of interaction. This highlights a design trade-off between expressive detail and interaction cost. In real-time avatar-to-avatar conversations, minimizing latency and motor load may be more important than maximizing expressive richness and nuances. These findings suggest that while face tracking offers the most natural and self-expressive experience, a well-designed control method with low latency and limited cognitive and motor load can serve as a viable alternative method of avatar facial expression control in VR devices without facial tracking technology.

Author Contributions

J. K. Sangeeth Chandran: conceptualization, data curation, formal analysis, methodology, visualization, writing – original draft, writing – review and editing. **Marisa Llorens Salvador:** methodology, supervision, validation, writing – review and editing. **Cathy Ennis:** conceptualization, funding acquisition, methodology, supervision, validation, writing – review and editing.

Acknowledgments

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 under the grant agreement no. 13/RC/2106P_2.

Funding

This work was supported by ADAPT — Centre for Digital Content Technology (13/RC/2106P_2).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

1. N. Hube, M. Reinelt, K. Vidackovic, and M. Sedlmair, "A Study on the Influence of Situations on Personal Avatar Characteristics," *Visual Computing for Industry, Biomedicine, and Art* 7, no. 1 (2024): 23.

2. H. H. S. Ip, S. W. L. Wong, D. F. Y. Chan, et al., "Enhance Affective Expression and Social Reciprocity for Children With Autism Spectrum Disorder: Using Virtual Reality Headsets at Schools," *Interactive Learning Environments* 32 (2024): 1012–1035.

3. C. O. 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* 10 (2020): 20626.

4. C. N. W. Geraets, S. Klein Tuente, B. P. Lestestuiver, et al., "Virtual Reality Facial Emotion Recognition in Social Environments: An Eye-Tracking Study," *Internet Interventions* 25 (2021): 100432.

5. L. Luo, D. Weng, N. Ding, J. Hao, and T. Ziqi, "The Effect of Avatar Facial Expressions on Trust Building in Social Virtual Reality," *Visual Computer* 39 (2022): 5869–5882.

6. C. Ennis, L. Hoyet, A. Egges, and R. McDonnell, "Emotion Capture: Emotionally Expressive Characters for Games," in *Proceedings of Motion on Games, MIG '13* (Association for Computing Machinery, 2013), 53–60.

7. T. Combe, R. Fribourg, L. Detto, and J.-M. Norm, "Exploring the Influence of Virtual Avatar Heads in Mixed Reality on Social Presence, Performance and User Experience in Collaborative Tasks," *IEEE Transactions on Visualization and Computer Graphics* 30 (2024): 2206–2216.

8. S. A. Aseeri and V. Interrante, "The Influence of Avatar Representation on Interpersonal Communication in Virtual Social Environments," *IEEE Transactions on Visualization and Computer Graphics* 27 (2021): 2608–2617.

9. S. Kimmel, F. Jung, A. Matviienko, W. Heuten, and S. C. J. 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* (Association for Computing Machinery, 2023).

10. Y. Wu, Y. Wang, S. Jung, S. Hoermann, and R. Lindeman, "Using a Fully Expressive Avatar to Collaborate in Virtual Reality: Evaluation of Task Performance, Presence, and Attraction," 2 (2021).

11. S. Bai, T. Wang, C. Li, et al., "Universal Facial Encoding of Codec Avatars From Vr Headsets," *ACM Transactions on Graphics* 43 (2024): 1–22.

12. C. Kim, H.-S. Cha, J.-H. Kim, H. Kwak, W. Lee, and C. Im, "Facial Motion Capture System Based on Facial Electromyogram and Electrooculogram for Immersive Social Virtual Reality Applications," *Sensors* 23 (2023): 3580.

13. K. Krug, X. Song, and W. Büschel, "Face Off: External Tracking vs. Manual Control for Facial Expressions in Multi-User Extended Reality," in *Proceedings of the 2025 Mensch Und Computer*, vol. 2025 (Association for Computing Machinery, 2025).

1546427x, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cav.70137 by National University Of Ireland Maynooth, Wiley Online Library on [07/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14. Y. Pan, S. Tan, S. Cheng, Q. Lin, Z. Zeng, and K. Mitchell, "Expressive Talking Avatars," *IEEE Transactions on Visualization and Computer Graphics* 30 (2024): 2538–2548.

15. A. Richard, C. S. Lea, S. Ma, J. Gall, F. D. L. Torre, and Y. Sheikh, "Audio- and Gaze-Driven Facial Animation of Codec Avatars," in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)* (IEEE, 2020), 41–50.

16. T. Karras, T. Aila, S. Laine, A. Herva, and J. Lehtinen, "Audio-Driven Facial Animation by Joint End-To-End Learning of Pose and Emotion," *ACM Transactions on Graphics* 36 (2017): 1–12.

17. H. Kim, G. Ali, and J.-I. Hwang, "Lufa: Lightweight Upper-Face Animation for Vr/Mr Avatars," in *2025 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (IEEE, 2025), 841–842.

18. A. Visconti, D. Calandra, and F. Lamberti, "Comparing Technologies for Conveying Emotions Through Realistic Avatars in Virtual Reality-Based Metaverse Experiences," *Computer Animation and Virtual Worlds* 34, no. 3-4 (2023): e2188.

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* (Association for Computing Machinery, 2021).

20. J. K. S. Chandran, M. L. Salvador, and C. Ennis, "Face Off: Evaluating Virtual Human Expressions and Non-Tracking Control Methods in Vr," in *2025 11th International Conference on Virtual Reality (ICVR)* (IEEE, 2025), 126–134.

21. J. Sweller, "Cognitive Load During Problem Solving: Effects on Learning," *Cognitive Science* 12, no. 2 (1988): 257–285.

22. P. M. Fitts, "The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement," *Journal of Experimental Psychology* 47, no. 6 (1954): 381–391.

23. S. Card, T. Moran, and A. Newell, "The Keystroke-Level Model for User Performance Time With Interactive Systems," *Communications of the ACM* 23 (1980): 396–410.

24. P. Ekman, "An Argument for Basic Emotions," *Cognition & Emotion* 6 (1992): 169–200.

25. P. Ekman and W. V. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement* (Consulting Psychologists Press, 1978).

26. F. Biocca and C. Harms, Guide to the Networked Minds Social Presence Inventory v. 1.2 (2003).

27. Y. M. Kim and I. Rhiu, "Development of a Virtual Reality System Usability Questionnaire (VRSUQ)," *Applied Ergonomics* 119 (2024): 104319.

28. T. W. Boiteau, P. Malone, S. Peters, and A. Almor, "Interference Between Conversation and a Concurrent Visuomotor Task," *Journal of Experimental Psychology. General* 143, no. 1 (2014): 295–311.

29. Z. Sheng, L. Nie, M. Liu, Y. w. Wei, and Z. Gao, "Toward Fine-Grained Talking Face Generation," *IEEE Transactions on Image Processing* 32 (2023): 5794–5807.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Video S1:** Demonstration of the evaluated facial control methods in VR.

1546427x, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cav.70137 by National University Of Ireland Maynooth, Wiley Online Library on [07/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License