This is an author and pre-print version. The official publication can be found at https://doi.org/10.1145/3743049.3748590

# Face Off: External Tracking vs. Manual Control for Facial Expressions in Multi-User Extended Reality

Katja Krug*
katjakrug@acm.org
Interactive Media Lab Dresden, TUD
Dresden University of Technology
Dresden, Germany

Xiaoli Song
xiaoli.song@mailbox.tu-dresden.de
Interactive Media Lab Dresden, TUD
Dresden University of Technology
Dresden, Germany

Wolfgang Büschel
bueschel@acm.org
VISUS, University of Stuttgart
Stuttgart, Germany

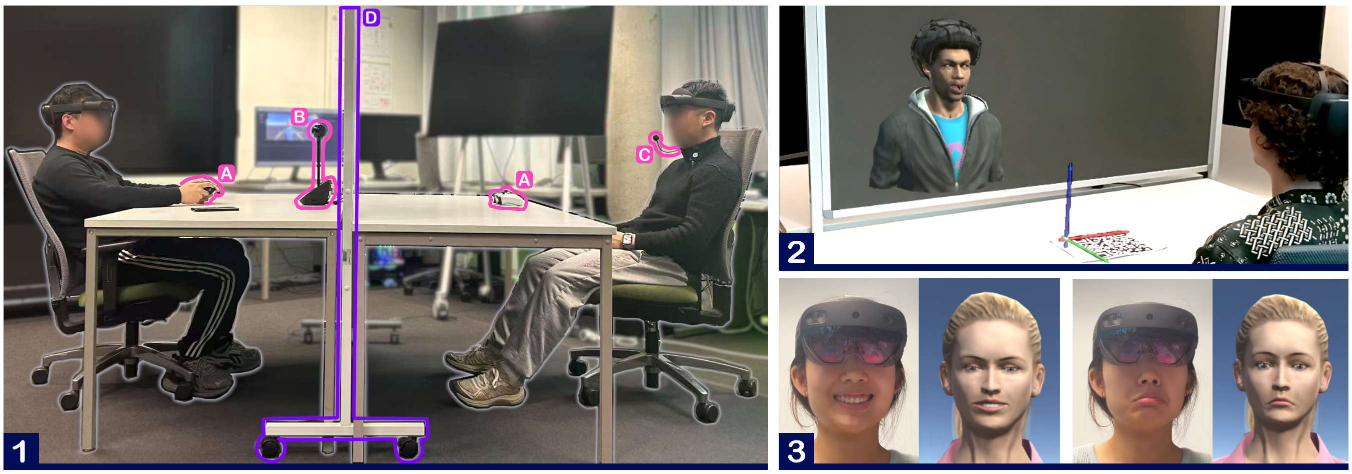


**Figure 1: We studied the differences between externally tracked (3) and manually controlled facial expressions in AR. In our study setup, both participants are visually separated by a whiteboard (1D). They see the conversation partner's 3D avatar in AR through their HMDs (2). Depending on the condition, face tracking via webcam (1B) or manual input (1A) controls the expressions shown by the avatars.**

## Abstract

In distributed multi-user XR spaces, avatar facial expressions are usually enabled by built-in sensors in high-end HMDs. Motivated by the diverse landscape of devices without these capabilities, we investigate two alternative methods to execute facial expressions. In a study with 18 participants collaborating in dyads, we compared (1) external webcam-based face tracking and (2) manually triggered preset expressions, exploring trade-offs between less reliable, video-based tracking of partially obscured faces and less natural manual control. Our results show that participants prefer manual triggering over unstable face tracking, as the latter leads to significantly higher task load and effort, while the former did not negatively influence interpersonal communication and was easier to use.

*Katja Krug is also with the Centre for Tactile Internet with Human-in-the-Loop (CeTI)


*MuC '25, Chemnitz, Germany*

ACM ISBN 979-8-4007-1582-2/25/08
https://doi.org/10.1145/3743049.3748590

## CCS Concepts

• **Human-centered computing → Mixed / augmented reality**; *Interaction techniques*.

## Keywords

Augmented Reality, Facial Expressions, Face Tracking

**ACM Reference Format:**
Katja Krug, Xiaoli Song, and Wolfgang Büschel. 2025. Face Off: External Tracking vs. Manual Control for Facial Expressions in Multi-User Extended Reality. In *Mensch und Computer 2025 (MuC '25), August 31-September 03, 2025, Chemnitz, Germany*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3743049.3748590

## 1 Introduction & Background

In avatar-mediated, multi-user Extended Reality (XR), expressive 3D avatars can support non-verbal communication. Especially in tasks requiring face-to-face discussions, facial cues can have a greater influence on conversational outcomes than bodily cues [16], because avatars with facial expressions significantly increase user satisfaction and social presence [25], as well as interpersonal attraction [28]. Realistic avatars can increase comfort [14] and interpersonal trust [1], and the mood of these facial expressions can further influence this trust positively or negatively [20]. However, most head-mounted displays (HMDs) lack high-quality face-tracking capabilities, excluding users without access to flagship HMDs from fully expressing themselves and socially participating in multi-user XR spaces. Designing these spaces in a more device-inclusive manner could narrow this gap [15].

One approach towards more inclusive design is allowing the substitution of high-end tracking sensors with cheaper and more accessible hardware, like external webcams. However, this could destabilize the expression of emotional cues, triggering the uncanny valley effect and a feeling of eeriness [4, 18]. The quality of face tracking is generally highly user-dependent due to factors such as facial hair, glasses, and even makeup [9, 26]. In multi-user XR, face tracking using external cameras is additionally limited by the HMDs partially covering the faces, requiring artificial substitution of the expressions in the missing upper-half of the face to decrease the sense of unnaturalness [27].

In light of these drawbacks, alternative control methods for facial expressions should be considered. Manually triggered emotional expressions can help users express their intentions and feelings in virtual environments [12] and enable emotional consensus and trust building during virtual collaboration [13]. However, explicitly triggering expressions, e.g, via controllers or hand gestures, can feel unnatural and disruptive in conversations [2] and, ultimately, hinder communication. Further, synthetic emotions must be carefully designed and smoothly animated to avoid the uncanny valley [3, 21].

Both external tracking and manually controlled expressions can make multi-user XR spaces more inclusive for users without access to flagship face tracking technology. Prior work, such as Kullmann et al. [17], already demonstrated the potential of synthesized expressions in comparison to tracked ones, but relied on pre-recorded animations without real-time interaction. To evaluate the potential of these alternatives to built-in face tracking in a dynamic conversational setting, we conducted a user study with 18 participants, comparing facial expressions based on camera-based, external live tracking with expressions triggered by explicit controller input. Our study setup paired users in a remote, HMD-based augmented-reality (AR) scenario. The external tracking of HMD-covered faces led to expected instability, with tracking issues increasing task load and varying widely between users. Our results show that, under adverse tracking conditions, manually triggered expressions cause lower perceived task load, do not reduce task performance or interpersonal trust, and are a viable alternative when built-in face tracking is unavailable for tasks that rely heavily on facial expressions.

## 2 Study

Our study employed a within-subject experimental design with two conditions: (1) real-time *tracked facial expressions* captured with an external web-cam and (2) *manually triggered preset expressions*. Condition sequences were counterbalanced across groups to minimize order effects. Participants alternated between the two roles in our task, guesser and answerer (see subsection 2.1), to ensure balanced exposure to both perspectives. Based on this design and indications from prior literature, we defined the following hypotheses:

**H1:** Tracked facial expressions will lead to a **higher sense of social presence** compared to manually triggered preset expressions, due to their real-time responsiveness.

**H2:** Tracked facial expressions will **enhance interpersonal trust** compared to manually triggered preset expressions, due to their perceived authenticity.

**H3:** Manually triggered preset expressions will **enhance communication satisfaction** compared to tracked facial expressions, due to their smoothness and stability.

**H4:** Tracked facial expressions will lead to a **greater task load** compared to manually triggered preset expressions, as they require actively making facial expressions rather than simple button presses.

### 2.1 Task & Participants

Prior research by Le Tarnec et al. [19, 25] shows that manual tasks strongly distract participants' attention from each other's faces in multi-user XR, causing them to miss facial expressions and nonverbal cues. To ensure focus on facial expressions, we adapted the game *"Who am I?"* into a dialogue-based scenario in which guessers had to identify their hidden character by interpreting non-verbal responses from the answerer's avatar. In this guessing game, the guesser asks a series of "yes/no" questions about their assigned character. Instead of answering verbally, answerers only respond through facial expressions (Happy - positive, Sad - negative, Surprised - unexpected, and Disgusted - strong negative), and guessers have to interpret the meaning of each response. In the tracking condition, we animated the HMD-covered upper half of the face when emotions were recognized, ensuring a fairer comparison by avoiding the uncanny valley effect caused by missing upper-face movement [27].

The study included 18 unpaid volunteer participants (16 men, 2 women) aged between 23 and 28 years (M = 25.5), recruited from the local university and grouped into 9 dyads. 11 had experience with VR/AR devices, 9 had used the HoloLens 2, and 7 had previously interacted with avatars.

### 2.2 System & Setup

During the study, two participants wearing AR HMDs sat face-to-face at a table divided by a whiteboard so they could not see each other's physical bodies (See Figure 1;1) and instead only see each other's avatars in front of them (See Figure 1;2). In this asymmetric setup, the guesser used a microphone to capture their voice and animate their avatar's lip movements accordingly while talking (See Figure 1;1C). During the tracking condition, the answerer's face was tracked by a webcam (See Figure 1;1B), and during the manual trigger condition, they used an Xbox controller to select the expressions (See Figure 1;1A). The study system was developed in Unity. External face tracking was achieved using OpenSeeFace [7]. Because the upper face was obscured, the four pre-determined reactions (Happy, Sad, Surprised, Disgusted) were inferred solely from mouth movements, specifically, whether it was opened and if the movement of mouth corners exceeded personalized thresholds in four directions (Up, Down, In, Out). Informed by the work of Du et al. [5], who define a combination of Action Units (AU) for compound facial expressions based on the Facial Action Coding System (FACS) [6, 8], corresponding blendshapes for full avatar facial animation were then triggered.

This is an author and pre-print version. The official publication can be found at https://doi.org/10.1145/3743049.3748590

To animate the guesser's avatar's lips, we used Oculus Lipsync [23]. Networking was handled using Colibri [11].

## 2.3 Procedure

After signing the informed consent form, participants were introduced to the game rules and guided through the HoloLens 2 usage and eye calibration process. Then, using the Xbox controllers, both participants selected their avatar characters from a diverse set of options. Once in the main game scene, participants experienced one of the two conditions. In the tracking condition, the answerer first completed a brief calibration to personalize the expression tracking. During calibration, they performed various facial expressions, monitoring their avatar's reaction, and self-assessing the perceived difficulty of achieving them. Based on this, we adjusted sensitivity thresholds for the respective expressions accordingly until users reported them to be comfortable and controllable. Each game round of *"Who am I?"* had a 5-minute time limit. In the beginning, the answerer viewed their target character's image and name on a screen. Then, the guesser started asking a series of "yes/no" questions, and the answerer responded only using facial expressions, guiding the guesser to identify the character. If the guesser was unable to determine the character, they could say "pass" or "next". The goal was to guess as many characters as possible within the time limit. After each round, participants completed questionnaires. Roles were then switched, and the process was repeated until all participants experienced both conditions.

## 2.4 Measurements

We logged the facial expression usage during the task. The collected data further included social presence, using the Multimodal Presence Scale (MPS) [22]; task load, via the standard NASA Task Load Index (NASA-TLX) [10]; interpersonal trust, using a slightly adapted self-developed questionnaire by Aseeri and Interrante [1]; and communication satisfaction, measured using adapted items from Suh [24]. Except for the NASA-TLX, all items were rated on 7-point Likert scales ranging from strongly disagree (1) to strongly agree (7), using the median to aggregate multiple items. All questionnaires are provided in the supplementary material for this paper.

# 3 Results

For the statistical analyses, we used the Wilcoxon signed-rank test. To control the false discovery rate, we applied Benjamini-Hochberg correction ($\alpha = .05$) and report corrected $p$-values. To estimate effect sizes for significant results, we used the matched pairs rank-biserial correlation ($r$).

## 3.1 Measures of Social Interaction

The questionnaire about **social presence** shows an acceptable internal consistency with Cronbach's $\alpha = .76$. The test showed no statistically significant differences between conditions ($W = 20.0, p = .964$), see Figure 2. These findings do not offer sufficient statistical evidence to support H1. The questionnaire about **interpersonal trust** shows good internal consistency, with Cronbach's $\alpha = .83$. Similarly to the social presence, the test showed no statistically significant difference between the two conditions ($W = 13.5, p = .507$). As shown in Figure 2, the distribution of trust scores in the tracking condition was slightly more concentrated (IQR = 1.125, Q1 = 5.00, Q3 = 6.125) than in the preset condition (IQR = 2.0, Q1 = 5.00, Q3 = 7.00). These findings do not provide sufficient statistical evidence to support H2. The questionnaire on **communication satisfaction** showed good internal consistency, with Cronbach's $\alpha = .87$. Again, the Wilcoxon signed-rank test did not show a statistically significant difference between conditions ($W = 21.5, p = .428$) and therefore there is insufficient evidence to support H3.

## 3.2 Performance and Usability Metrics

For **task load**, we found a statistically significant difference between the conditions for physical task load ($W = 5.5, p = .022, r = .879$) and effort ($W = 9.0, p = .022, r = .829$). As shown in Figure 2, physical task load scores under the tracking method were more widely dispersed than in the manual method. This suggests that the participants' perceived physical task load varied more in the tracking condition. These results provide statistical support for H4, confirming that using real-time facial tracking expressions causes a much higher task load than manually triggered preset expressions.

To further evaluate this, we also looked into the **ease of use** for expressions in both methods. Participants rated them on a 7-point Likert scale from "very difficult" (1) to "very easy" (7). As Figure 2 shows, the manual expressions were easier to use, achieving a median score of 7 and tightly distributed scores. In contrast, the tracking method was rated significantly harder to use with a median of 3 ($W = 2.0, p = .016, r = -0.962$) and showed a wider distribution of ratings, suggesting a more inconsistent feeling of the tracking expressions across participants. According to participants' feedback, a total of 12 participants reported experiencing difficulties, all of which were related to the tracking method. Their feedback can be grouped into three types of difficulty: *control difficulty* (n = 6), *response instability* (n = 4), and *physical fatigue* (n = 2).

We explored two aspects of participants' **preferences**: (1) which method they preferred to use for *self-expression*, and (2) which method they *preferred to observe* in others' facial animations. As shown in Table 1, for self-expression, 6 of 16 participants preferred the tracking method, 8 favored the preset method, and 2 reported no significant difference (NSD) between the two methods. For viewing others' avatars, 4 participants preferred the tracking method, while 10 favored preset expressions. Again, 2 indicated no clear preference.

**Table 1: Preference for expression methods for the participants' self-expression and as the observer.**

| Role | Tracking | Manual | No sig. Diff. |
|---|---|---|---|
| Self-Expression | 37.5% | 50.0% | 12.5% |
| Observer | 25.0% | 62.5% | 12.5% |

Using the timestamps and corresponding expression usages in the log files, we calculated the **duration of each expression state** by computing the time differences between consecutive timestamps. We compared expression usage across conditions using the relative percentage of time spent in each expression (See Table 2).

This is an author and pre-print version. The official publication can be found at https://doi.org/10.1145/3743049.3748590

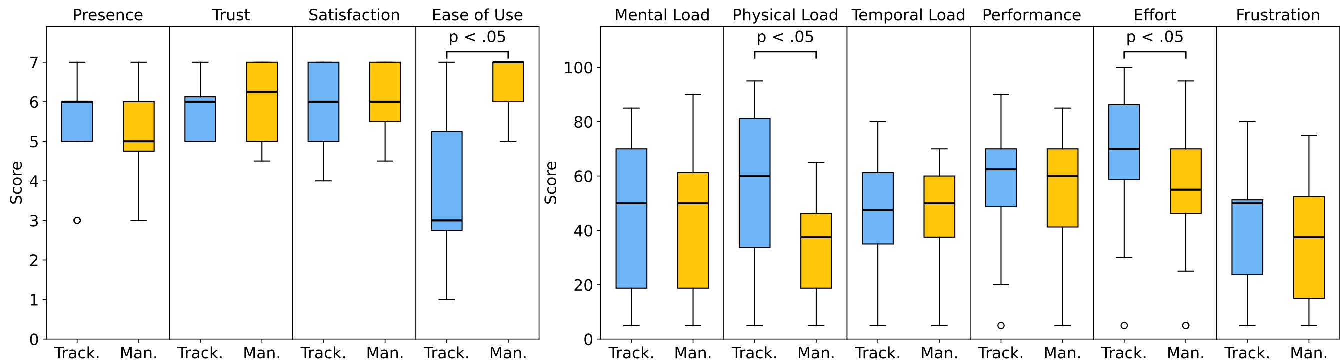


**Figure 2: From left to right: Scores for Social Presence, Interpersonal Trust, Communication Satisfaction, and Ease of Use, all in a range from 1 to 7, with larger scores being better. Second group: Raw TLX scores for Mental Load, Physical Load, Temporal Load, Performance, Effort, and Frustration, all in a range from 0 to 100, with larger values being better. The boxes show the 25th and 75th percentiles, the horizontal bars mark the median, and the whiskers show the smallest and largest values up to 1.5 times the interquartile range; outliers are marked individually. Brackets mark significant differences.**

**Table 2: Percentage distribution of expression duration by condition.**

| Method | Neutral | Happy | Sad | Surprised | Disgusted |
|---|---|---|---|---|---|
| Preset | 69.89% | 11.48% | 13.19% | 1.20% | 4.24% |
| Tracking | 36.99% | 19.56% | 41.10% | 0.85% | 1.50% |

## 4 Discussion

Our study compared alternatives to built-in face-tracking for emotional expression in HMD XR, each with distinct limitations: tracking was less reliable due to the partially occluded faces, and manual triggering required conscious effort, potentially reducing naturalness. We wanted to investigate the user's attitude towards these factors and their influence on communication. Significant differences emerged in user experience: participants reported higher physical task load and effort for the face-tracking condition, leading us to accept H4. One contributing factor may be the variability in tracking quality across participants depending on facial structure, facial hair, and accessories [9, 26]. 12 out of 18 participants reported such difficulties, requiring exaggerated facial gestures to ensure accuracy, potentially increasing cognitive and physical demands, which correlated with lower individual ease-of-use scores. The expression log (see Table 2) offers another perspective. In the manual condition, neutral expressions occupied 7̃0% of time versus only 3̃7% in the tracking condition, suggesting an over-sensitivity to involuntary minor movements or specific facial characteristics. The *Sad* expression was triggered more than three times more often in the tracking condition, probably due to many people's resting face having slightly down-turned mouth corners, while *Disgusted* was triggered around three times more often in the manual condition, possibly indicating difficulties in triggering it through tracking or participants being more open to express strong emotions manually.

Although our scenario revolved around collaboration through emotional expressiveness, our results showed that the type of control had no significant impact on interpersonal communication, leading us to reject H1-H3. Despite knowing whether their partner used tracked or manually triggered expressions, participants did not perceive manually triggered expressions as less trustworthy or satisfying. We speculate that the higher reliability of the manually triggered expressions outweighed their potential unnaturalness, as evidenced by strong user preferences for the manual method for both self-expression and observation. While the former can be traced back to the significantly increased task load and effort, the latter can not be connected to any of our measurements.

## 5 Limitations

While our findings offer insights into the usability of the presented methods, there are limitations to our study. First, the unbalanced gender distribution (16 male, 2 female) restricts generalizability. Further, our study task forced participants to rely solely on facial expressions for communication, which is not a reflection of a natural, dynamic conversation. Finally, while representative of current external face-tracking technology, the limitations of our tracking setup may have exaggerated the task load associated with this condition. Improved or alternative external tracking methods could yield different outcomes and warrant further investigation.

## 6 Conclusion

We present the results of a study comparing two low-tech alternatives to built-in face tracking for emotional expression in avatar-mediated multi-user XR collaboration: external webcam-based tracking and manually triggered expressions via controller. Given the limited distribution of face-tracking sensors in current HMDs and motivated by the need to make multi-user XR spaces more inclusive, we explored the trade-offs between less reliable video-based tracking of half-obscured faces and less natural manual control of emotional expressions. Our findings show that participants prefer manual triggering, as unstable tracking significantly increased task load and effort. We conclude that unless an alternative real-time face tracking solution can be consistently reliable, it may hinder rather than enhance collaboration. Since high tracking quality strongly depends on the available hardware, we recommend manual expression control as a viable, accessible alternative that did not compromise communication in our scenario. Future research should refine manual expression interfaces to support diverse input methods and a variety of collaborative contexts.

## Acknowledgments

This work was funded by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany's Excellence Strategy – EXC 2050/1 – Project ID 390696704 – Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop" (CeTI) of Technische Universität Dresden. This work was also supported by the Alexander von Humboldt Foundation, funded by the German Federal Ministry of Research, Technology and Space.

## References

[1] Sahar Aseeri and Victoria Interrante. 2021. The Influence of Avatar Representation on Interpersonal Communication in Virtual Social Environments. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (May 2021), 2608–2617. doi:10.1109/TVCG.2021.3067783

[2] Marc Baloup, Thomas Pietrzak, Martin Hachet, and Géry Casiez. 2021. Non-Isomorphic Interaction Techniques for Controlling Avatar Facial Expressions in VR. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*. ACM, Osaka Japan, 1–10. doi:10.1145/3489849.3489867

[3] Niccolò Casiddu, Francesco Burlando, Claudia Porfirione, and Annapaola Vacanti. 2021. Designing Synthetic Emotions of a Robotic System. In *Human Systems Engineering and Design III*, Waldemar Karwowski, Tareq Ahram, Darko Etinger, Nikola Tanković, and Redha Taiar (Eds.). Springer International Publishing, Cham, 148–155. doi:10.1007/978-3-030-58282-1_24

[4] Theo Combe, Rebecca Fribourg, Lucas Detto, and Jean-Marie Norm. 2024. 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, 5 (May 2024), 2206–2216. doi:10.1109/TVCG.2024.3372051

[5] Shichuan Du, Yong Tao, and Aleix M. Martinez. 2014. Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences* 111, 15 (2014), E1454–E1462. doi:10.1073/pnas.1322355111 arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1322355111

[6] Paul Ekman and Wallace V Friesen. 1978. Facial action coding system. *Environmental Psychology & Nonverbal Behavior* (1978).

[7] Emiliana. 2021. *OpenSeeFace*. https://github.com/emilianavt/OpenSeeFace

[8] E Friesen and Paul Ekman. 1978. Facial action coding system: a technique for the measurement of facial movement. *Palo Alto* 3, 2 (1978), 5.

[9] Geof Givens, J. Ross Beveridge, Bruce A. Draper, Patrick Grother, and P. Jonathon Phillips. 2004. How features of the human face affect recognition: a statistical comparison of three face recognition algorithms. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, Vol. 2. II–II. doi:10.1109/CVPR.2004.1315189

[10] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (Oct. 2006), 904–908. doi:10.1177/154193120605000909

[11] Sebastian Hubenschmid, Daniel Immanuel Fink, Johannes Zagermann, Jonathan Wieland, Harald Reiterer, and Tiare Feuchtner. 2023. Colibri: A Toolkit for Rapid Prototyping of Networking Across Realities. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 9–13. doi:10.1109/ismar-adjunct60411.2023.00010

[12] Masahiro Ide, Shoji Oshima, Shingo Mori, Masato Yoshimi, Junko Ichino, and Shunichi Tano. 2021. Effects of Avatar's Symbolic Gesture in Virtual Reality Brainstorming. In *Proceedings of the 32nd Australian Conference on Human-Computer Interaction (OzCHI '20)*. Association for Computing Machinery, New York, NY, USA, 170–177. doi:10.1145/3441000.3441081

[13] Allison Jing, Michael Frederick, Monica Sewell, Amy Karlson, Brian Simpson, and Missie Smith. 2023. How Visualising Emotions Affects Interpersonal Trust and Task Collaboration in a Shared Virtual Space. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 849–858. doi:10.1109/ISMAR59233.2023.00100

[14] Jonah-Noël Kaiser, Simon Kimmel, Eva Licht, Eric Landwehr, Fabian Hemmert, and Wilko Heuten. 2025. *Get Real With Me: Effects of Avatar Realism on Social Presence and Comfort in Augmented Reality Remote Collaboration and Self-Disclosure.* 18 pages. doi:10.1145/3706598.3713541

[15] Katja Krug, Julián Méndez, Weizhou Luo, and Raimund Dachselt. 2024. Don't Leave Me Out: Designing for Device Inclusivity in Mixed Reality Collaboration. *arXiv preprint arXiv:2409.05374* (2024).

[16] Catherine Kruzic, David Kruzic, Fernanda Herrera, and Jeremy Bailenson. 2020. Facial Expressions Contribute More than Body Movements to Conversational Outcomes in Avatar-Mediated Virtual Environments. *Scientific Reports* 10 (Nov. 2020). doi:10.1038/s41598-020-76672-4

[17] Peter Kullmann, Timo Menzel, Mario Botsch, and Marc Erich Latoschik. 2023. 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 (CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3544549.3585617

[18] Dilshani Kumarapeli, Sungchul Jung, and Robert W. Lindeman. 2022. 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)*. 650–651. doi:10.1109/VRW55335.2022.00176

[19] Hugo Le Tarnec, Olivier Augereau, Elisabetta Bevacqua, and Pierre De Loor. 2023. Improving Collaborative Learning in Virtual Reality With Facial Expressions. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers (UbiComp/ISWC '22 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 303–307. doi:10.1145/3544793.3560387

[20] Le Luo, Dongdong Weng, Ni Ding, Jie Hao, and Ziqi Tu. 2022. The Effect of Avatar Facial Expressions on Trust Building in Social Virtual Reality. *Vis. Comput.* 39, 11 (Oct. 2022), 5869–5882. doi:10.1007/s00371-022-02700-1

[21] Meeri Mäkäräinen, Jari Kätsyri, and Tapio Takala. 2014. Exaggerating Facial Expressions: A Way to Intensify Emotion or a Way to the Uncanny Valley? *Cognitive Computation* 6, 4 (Dec. 2014), 708–721. doi:10.1007/s12559-014-9273-0

[22] Guido Makransky, Lau Lilleholt, and Anders Aaby. 2017. Development and validation of the Multimodal Presence Scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior* 72 (July 2017), 276–285. doi:10.1016/j.chb.2017.02.066

[23] Meta Developers. 2025. *Unity Development Using Oculus Lip Sync.* https://developers.meta.com/horizon/documentation/unity/audio-ovrlipsync-unity/

[24] Kil Soo Suh. 1999. Impact of communication medium on task performance and satisfaction: an examination of media-richness theory. 35, 5 (May 1999), 295–312. doi:10.1016/s0378-7206(98)00097-4

[25] Hugo Le Tarnec, Elisabetta Bevacqua, Olivier Augereau, and Pierre De Loor. 2023. Effect of Avatar Facial Expressiveness on Team Collaboration in Virtual Reality. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents (IVA '23)*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/3570945.3607330

[26] Philipp Terhörst, Jan Niklas Kolf, Marco Huber, Florian Kirchbuchner, Naser Damer, Aythami Morales Moreno, Julian Fierrez, and Arjan Kuijper. 2022. A Comprehensive Study on Face Recognition Biases Beyond Demographics. *IEEE Transactions on Technology and Society* 3, 1 (March 2022), 16–30. doi:10.1109/TTS.2021.3111823

[27] Angela Tinwell, Debbie Nabi, and Andrew Williams. 2011. Facial Expression of Emotion and Perception of the Uncanny Valley in Virtual Characters. *Computers in Human Behavior* 27 (March 2011), 741–749. doi:10.1016/j.chb.2010.10.018

[28] Yuanjie Wu, Yu Wang, Sungchul Jung, Simon Hoermann, and Robert Lindeman. 2021. Using a Fully Expressive Avatar to Collaborate in Virtual Reality: Evaluation of Task Performance, Presence, and Attraction. *Frontiers in Virtual Reality* 2 (April 2021). doi:10.3389/frvir.2021.641296