VR Charades

Studying the Importance of Facial Expressions and Gestures on Social Communication in Virtual Environments

Daniel Pauls
Visual Computing and Artificial Intelligence
Coburg University
Coburg, Germany
daniel.pauls@stud.hs-coburg.de

Stefan Kunz
Visual Computing and Artificial Intelligence
Coburg University
Coburg, Germany
stefan.kunz@stud.hs-coburg.de

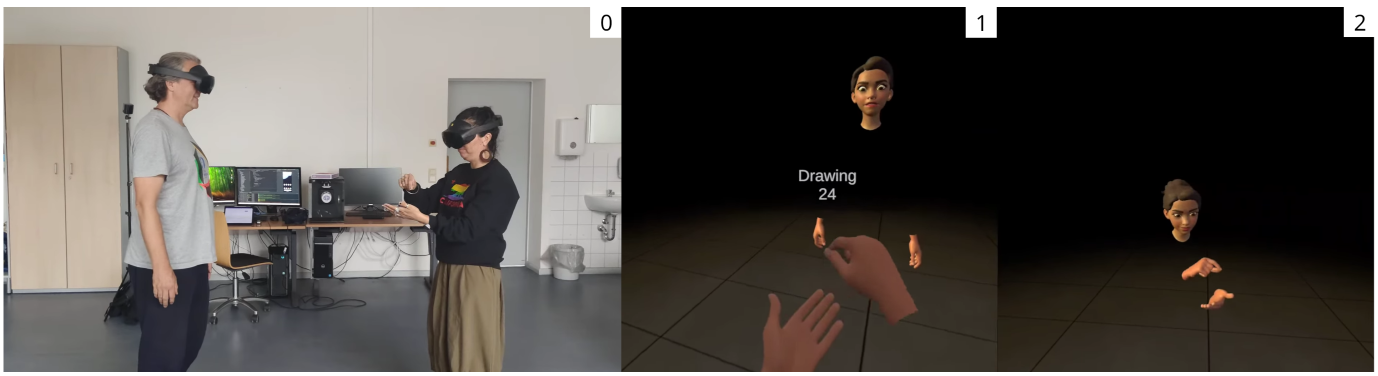


Fig. 1: Two participants taking part in the study. 0) shows the participants in real life. The person on the left is the guesser, and the person on the right is the actor. 1) shows the view from the perspective of the actor. 2) shows the view from the perspective of the guesser.

***Abstract*—This research explores non-verbal communication in virtual reality through a charade-style game utilizing face and hand tracking. Participants played the game as dyads under six different conditions varying the availability of facial expressions and gestures. A minimal avatar representation was used, consisting of a head and hands, with facial expressions driven by blendshape data and hand movement via tracking input. Participants alternated between acting and guessing roles in each condition. The results were analyzed using boxplots without inferential statistics, showing that conditions with dynamic face and hand tracking were perceived more positively in terms of communication effectiveness and social presence. This study demonstrates the potential of enhanced non-verbal cues in improving VR-based interaction and highlights avenues for future research in VR communication.**

***Index Terms*—virtual reality, communication, collaboration, gestures, avatars**

I. Introduction

Virtual reality (VR) has become an increasingly important medium for remote social interaction. However, many current VR systems still struggle to capture the richness of non-verbal communication, which plays a central role in human interaction. Facial expressions and gestures are especially critical in conveying emotions, intentions, and attention.

This study investigates the importance of facial expressions and hand gestures for communication in VR, using a charades-style game as a controlled experimental setting. By comparing dynamic and static representations of face and hand movements, we aim to better understand how different input modalities affect social presence, embodiment, and communication effectiveness. The goal is to contribute to ongoing research in VR interaction and human-computer communication, highlighting the value of non-verbal signals in virtual environments.

II. Previous Work

Prior research has explored how different input modalities influence social interaction in virtual environments. Adkins et al. [3] investigated the effects of hand tracking versus controller input on social presence and communication quality in VR, specifically in combination with audio channels. They found that hand tracking led to more natural interaction and increased feelings of presence compared to traditional controllers.

Wu et al. [2] also used a charades-style game to study social interaction in VR, but their implementation relied on controller-based input and did not include face or eye tracking. Their findings demonstrate the value of gesture-based interaction for shared understanding in virtual games, but offer limited insight into more expressive non-verbal channels.

While these studies focused on input devices and verbal communication, it inspired us to extend this line of research by examining non-verbal communication in the absence of speech. In contrast to Adkins et al., our study isolates non-verbal cues by removing verbal communication, focusing on the impact of facial expressions and gestures on social presence and interaction quality in a VR charades game.

A. Communication via Facial Expressions and Gestures

Facial expressions and gestures serve as crucial non-verbal cues for communication, enhancing emotional clarity and intent interpretation. Non-verbal communication, which comprises the majority of communication at any given time [7], contributes to the quality and emphasis of interactions. In a charades-based game, these subtle signals become primary tools for expression, compensating for the lack of verbal language.

While VR allows for communication between individuals through text and voice chat, it also enables expressive non-verbal interactions through facial expressions, gestures and proxemics. Through face tracking and hand motion capture, VR systems allow users to communicate effectively without speech, making them particularly useful in games like charades, where gesture-based expression is key.

III. Experimental Design

This study uses a within-subjects design to evaluate the effects of non-verbal communication under the conditions seen in Table I. We measure social presence, embodiment, communication effectiveness, game enjoyment, and game difficulty using questionnaires. In addition, we measure performance using the number of correct guesses.

A. Hypotheses

Our hypotheses are that dynamic conditions perform better than their static counterparts in terms of social presence, embodiment, and communication effectiveness. We expect that the condition *Hands and Face* will outperform the condition of only Hands and only Faces. Furthermore, we expect that social presence will be highest in conditions where the face is dynamic.

B. Participants

There were a total of 10 participants (5 dyads). 80% reported themselves as male and 20% as female. The participants' ages ranged from 19 to 48 (*mean* = 28, *std* = 9.2). Participants were recruited via e-mail and word of mouth. The experiment was announced as a group task. Each participant was given a sequential identification number. For each condition, the participants performed two games per dyad with roles reversed: once as the actor and once as the guesser. In each round, one participant acted out the given words while the other guessed them. This ensured that every participant experienced both roles under each of the six conditions. Dyads were in the same room. Only the actor was instructed not to communicate verbally, while the guesser was free to speak their guesses out loud.

C. Experimental Conditions

The communication elements are divided into three components: *Face*, *Hands*, *Hands and Face*. Furthermore, we divide these components into their dynamic and static interaction capabilities resulting in a total of six conditions. The overall avatar used in this experiment comprises of a head and hand models.

1) Face - Dynamic: The avatar they control exists only as a head model. The position and rotation are tracked by the sensors of the head-mounted display (HMD). Eye and facial tracking features are turned on.

2) Hands - Dynamic: The avatar they control exists only as the hand models. Hand tracking is done using the *Quest Pro's* inbuilt hand tracking sensors.

3) Hands and Face - Dynamic: The avatar consists of both the dynamic face and the dynamic hand variant.

4) Face - Static: The avatar they control exists only as a head model. The position and rotation are tracked by the HMD sensors. Eye and facial tracking features are turned off.

5) Hands - Static: The avatar they control exists only as the hand models. Instead of hand tracking, the participants use motion controllers to move their virtual hands. These are limited to different hand and finger positions by pressing certain buttons on the controller. Otherwise, the fingers remain in a static position.

6) Hands and Face - Static: The avatar consists of both the static face and the static hand variant.

D. Prototype and VR Environment

The prototype was developed in Unity version 6000.0.49f1 as a standalone application running on the participants' Meta Quest Pro headsets. The avatars used are generic female models consisting only of a head and two hands. We use the same models for both participants. Facial expressions are rendered using blendshapes, controlled by floating-point values at a maximum frequency of 60Hz. For the individual bones of the hands and head, we transmit the current position and rotation of each bone at a maximum frequency of 60Hz.

A UDP network connection links each HMD to a local server running on a nearby desktop computer, which handles the transmission of blendshape and hand data. The application is based on a sample scene from the Unity-Movement package provided by Meta [1], chosen to facilitate the integration of face tracking. No noticeable latency occurred during testing.

Only the actor sees a minimal UI with the current word to act out. The word list used in the game consists of 100 activity-based terms, generated using an AI-based text generation model and subjectively reviewed for appropriate difficulty. Researchers manually trigger the transition to the next word after a successful guess or once the 30 second limit is reached.

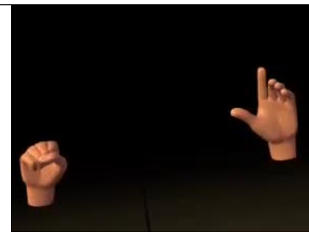
		Communication Elements		
		Face	**Hands**	**Hands and Face**
Interaction Capability	**dynamic**			
	static			

TABLE I: Table of conditions. From left-to-right and top-to-bottom: *dynamic - Face*, *dynamic - Hands*, *dynamic - Hands and Face*, *static - Face*, *static - Hands*, *static - Hands and Face*

E. Procedure

Two participants arrive at the laboratory and are asked to complete consent forms, a demographic survey, and information disclosure agreements. They receive an explanation of the study and the charades game rules, followed by an instruction on how to wear the headset. Participants are also informed about potential cybersickness and told that they can cancel their participation at any time without consequences.

Researchers measured and adjusted the interpupillary distance (IPD) for each participant if needed and hand out a Meta Quest Pro HMD. Participants put on the headset themselves and are guided to a clear area in the center of the room to avoid collisions during the study.

To familiarize themselves with the virtual environment and avatar, participants spend about one minute in a virtual room where they can observe and test facial and eye tracking with their partner. A brief testing phase ensures that the actor can see the word UI and hear the audio cues. The experiment then starts: the actor has 90 seconds to act out as many words as possible for the guesser to identify. Each word has a maximum duration of 30 seconds. If this limit is reached, a sound cue and a researcher will signal that a new word is shown. When a word is correctly guessed, the researcher will press a button to show the next word.

After 90 seconds, the roles are reversed, and a new set of words is used. This procedure is repeated for each of the six experimental conditions. Participants are informed before each round, which condition they are about to play. To counterbalance order effects, each dyad follows a different predefined condition sequence. After each condition, participants complete a questionnaire before continuing with the next round. The entire session takes approximately 45 to 50 minutes per dyad.

F. Measurements

After each condition, the participants undergo a questionnaire about the social presence, avatar embodiment, communication through body language, game enjoyment, and difficulty. To evaluate the social presence we used predefined questions from Nowak et al. [4] and for the embodiment T. Peck et al. [5]. The questions are listed in Table II. All measurements use the 7-pt Likert scale. Additionally, the time it took the dyad to guess the words correctly is also measured as well as their number of correct guesses made during the experiment.

IV. Results

The results were analyzed based on descriptive statistics and visual inspection of boxplots for each measured dimension across the six conditions. No inferential statistical tests were conducted due to the small number of participants. However, the visualizations suggest that certain conditions consistently lead to higher ratings in dimensions such as social presence, embodiment, and communication effectiveness.

A. Performance

Performance was measured by the amount of correct guesses and the time it took to guess the word. Figure 7 shows that all *Dynamic* conditions perform better than their *Static* counterparts. Additionally *Hands* + *Face* for both *Dynamic* and *Static* performed the best out of all conditions. This is in line with our hypotheses.

Figure 8 shows the average time it took to guess a word under the six conditions. Here the best conditions are *Hands + Face Static*, *Hands + Face Dynamic* and *Hands Dynamic*. Both *Face* variants perform the worst. Surprisingly to see, it is only the *Hands* condition where the *Dynamic* outperforms the *Static* variant.

TABLE II: Questionnaire

Measurement	Questions	Scale (1-7)
Social Presence	To what extend was this like a face-to-face meeting?	Not like face-to-face meeting - A lot like face-to-face meeting
	To what extend did the the virtual character seem real?	Not real at all - Very real
	To what extend could you understand the other person's feelings through their body language?	Not understandable at all - Very understandable
	To what extend did you feel you could get to know someone that you met only through this system?	Strongly disagree - Strongly agree
Embodiment	I felt as if the virtual body was my own	Strongly disagree - Strongly agree
	I felt as if the movements of the virtual body were influencing my own movements	Strongly disagree - Strongly agree
	My hand, eye and face movement were reflected naturally in the virtual environment	Strongly disagree - Strongly agree
Communication Effectiveness	To what extend were you aware of the other person paying attention to you?	Not aware at all - Very aware
	I was able to communicate my intentions clearly to my partner	Strongly disagree - Strongly agree
	My partner could understand my body language	Strongly disagree - Strongly agree
Game Enjoyment	I enjoyed playing the game	Strongly disagree - Strongly agree
	I felt immersed in the game	Strongly disagree - Strongly agree
	The game was engaging and fun	Strongly disagree - Strongly agree
Game Difficulty	It was difficult to act out words	Strongly disagree - Strongly agree
	It was difficult to guess the words	Strongly disagree - Strongly agree

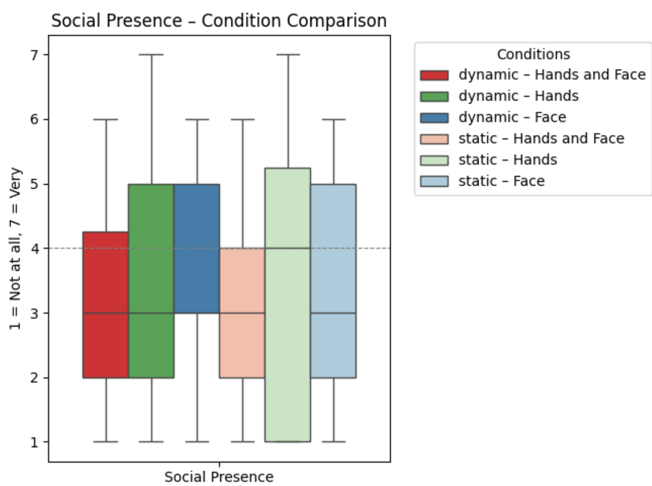


Fig. 2: The Social Presence in comparison between each condition.

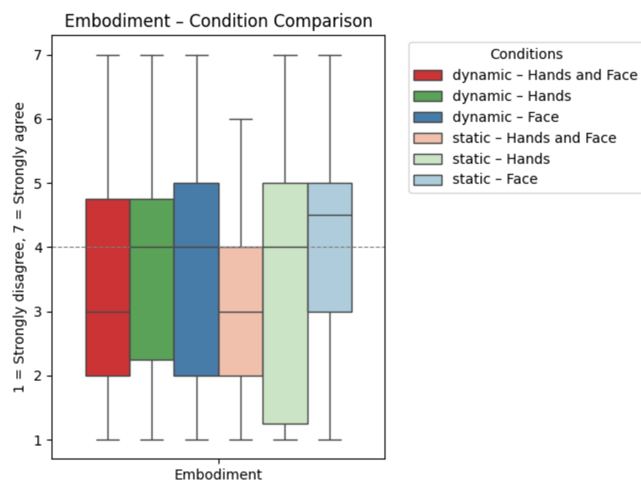


Fig. 3: The Embodiment in comparison between each condition.

B. Qualitative Results

The visualizations show the combined result of each questions for the given category (see table II). The *Hands + Face Static* condition showed the least sense of embodiment, although not significant. The communication effectiveness for the *Hands Dynamic* condition proved to be better compared to its static counterpart. Interestingly, figure 5 and 6 show the *Hands + Face Dynamic* condition as the least enjoyable game experience whilst also being perceived as the most difficult one. Surprisingly, the *Dynamic* conditions show an equal median score with the rest of the conditions. Although these observations cannot be confirmed statistically, they indicate potential trends that warrant further investigation in future studies with larger sample sizes.

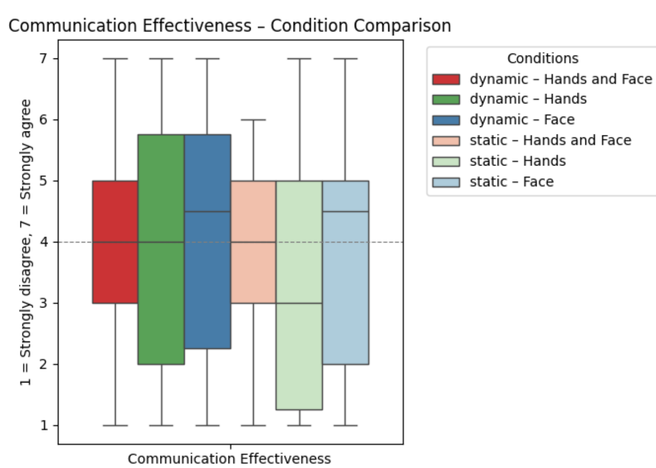


Fig. 4: The Communication Effectiveness in comparison between each condition.

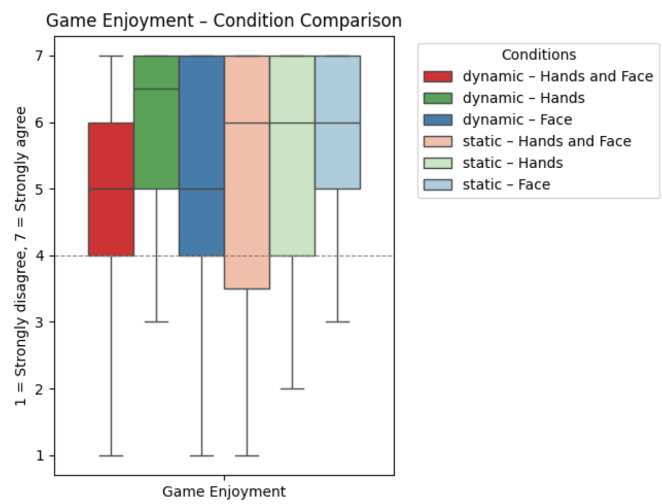


Fig. 5: The Game Enjoyment in comparison between each condition.

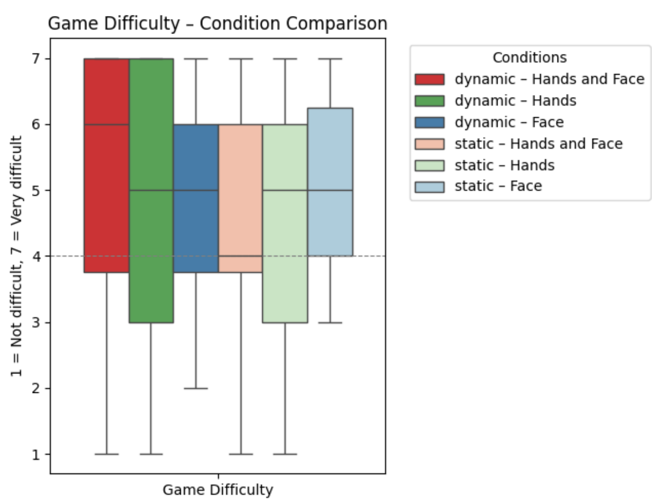


Fig. 6: The Game Difficulty in comparison between each condition.

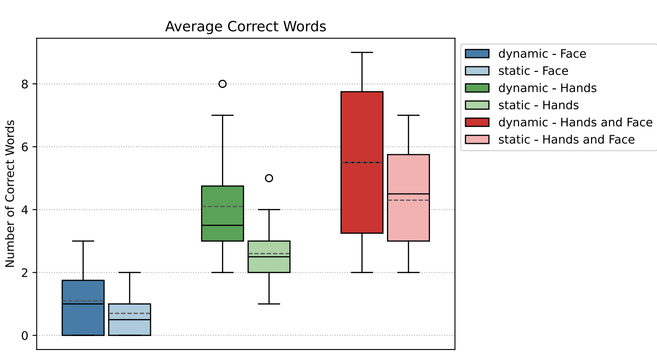


Fig. 7: Box plot showing the average number of correct words for each condition. Bigger is better.

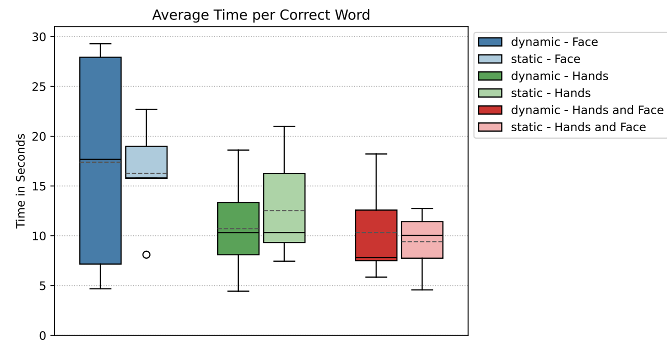


Fig. 8: Box plot showing the average time per correct word in seconds for each condition. Smaller is better.

V. Discussion and Limitations

Since the participants were in the same room during the experiment, some verbal communication was bound to happen. During challenging charades-tasks, the actor would sometimes respond with a mumbled yes or no to the guesser. A way to prevent accidental verbal communication would be to separate the individuals in separate rooms.

A limitation in this study was an uneven distribution of gender. With 80% of participants identifying as male and 20% as female and most of the dyads were familiar with their partner. This imbalance might influence the results, since research indicates that gender can have an effect on collaboration strategies in virtual environments [8].

A technical limitation is the range of the hand tracking the Quest Pro is capable of. Moving the hands too close to the face or above and behind the head results in a tracking loss. Once the tracking is lost, the virtual hands will jump to the default position in the virtual environment.

Another small issue during experiment was some sporadic and short-lived connection loss during the experiment. The reason behind this is unknown as it only occurred twice. We speculate that this was caused by the distance between the HMDs and the relay server.

VI. Conclusion

This exploratory study investigated how different levels of non-verbal communication, specifically facial expressions and hand gestures, affect interaction in VR-based charade games. Although no inferential statistics were applied, the visual analysis of participant feedback suggests that conditions with dynamic face and hand tracking can enhance the perceived quality of communication and social presence.

The findings highlight the importance of rich embodiment in VR communication, especially in tasks that depend on non-verbal cues. While technical factors were not central to this study, the prototype demonstrated that even lightweight avatar representations can facilitate meaningful interaction when augmented by face and hand tracking.

Future research should expand on these results with larger samples and statistical analysis to further evaluate the impact of non-verbal channels on communication in VR.

REFERENCES

[1] Meta, *Unity-Movement*, Version 76.0.1, 2025. [Online]. Available: https://github.com/oculus-samples/Unity-Movement

[2] Y. Wu, Y. Wang, S. Jung, S. Hoermann, and R. W. Lindeman, "Using a fully expressive avatar to collaborate in virtual reality: Evaluation of task performance, presence, and attraction," *Frontiers in Virtual Reality*, vol. 2, 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frvir.2021.641296. doi: 10.3389/frvir.2021.641296

[3] A. Adkins, A. Normoyle, L. Lin, Y. Sun, Y. Ye, M. Di Luca, and S. Jörg, "How important are detailed hand motions for communication for a virtual character through the lens of charades?," *ACM Transactions on Graphics*, vol. 42, no. 3, Article 27, May 2023. doi: 10.1145/3578575

[4] K. L. Nowak and F. Biocca, "The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments," *Presence: Teleoperators and Virtual Environments*, vol. 12, no. 5, pp. 481–494, Oct. 2003. doi: 10.1162/105474603322761289

[5] T. Peck and M. Gonzalez-Franco, "Avatar embodiment. A standardized questionnaire," *Frontiers in Virtual Reality*, vol. 1, Feb. 2021. doi: 10.3389/frvir.2020.575943

[6] K. Eves and J. Valasek, "Adaptive control for singularly perturbed systems examples," Code Ocean, Aug. 2023. [Online]. Available: https://codeocean.com/capsule/4989235/tree

[7] A. Mehrabian, "Nonverbal Communication"

[8] Amal Yassien, ElHassan B. Makled, Passant Elagroudy, Nouran Sadek, and Slim Abdennadher. 2021. Give-Me-A-Hand: The Effect of Partner's Gender on Collaboration Quality in Virtual Reality. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama, Japan) (CHI EA '21). Article 392, 6 pages. https://doi.org/10.1145/3411763.3451601