Communication Behavior in Embodied Virtual Reality

Date de création	@March 28, 2023 4:15 PM
Love	♥♥
Link	Communication Behavior in Embodied Virtual Reality | Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (acm.org)
Mots-clés	Collaboration, Communication, Embodiment, Presence
Personne	Harrison Jesse Smith, Michael Neff
Technologie	VR
# Year of publication	2018

Subject : how embodied virtual reality (VR) can support communication around a spatial task

Communication Behavior in Embodied Virtual Reality - Google Slides

Principle :

The Department of Energy estimates that roughly eight percent of US energy is used to support passenger transport to enable face-to-face communication [1].

Embodiment as a communication tool.

Embodied VR → a person's movements are tracked and then used to drive an avatar in a shared virtual world.

To render participants' bodies : use of optical motion capture to create simple 3D meshes of the bodies (eyebrowridge+nose, but no others facial expressions) =

manual gesture, locomotion and verbal dialog, but limited hand movement and no facial expressions.

Collaborative task : role-play being new roommates 1) allocate rooms 2) agree on a furniture arrangement (need manipulation of the environnement)

This echoes earlier work comparing face-to-face, audio/video, and audio-only communication.

Most real-time collaboration mediums can be grouped into three different categories: face-to-face, tools that support audio communication only, such as a telephone, and tools that support audio and visual communication, such as video conferencing.

→ but video does not usually seem to be enough :

they cannot fully observe their partner (visibility) or their partner's physical context (visual co-presence), most video feeds are stationary, positioned to provide closeup representations of the remote partner's face or upper torso (Such positioning can hinder movement-centric tasks, and prevent transmission of posture and gesture cues), Offsets between screens and cameras make it difficult or impossible to establish mutual gaze (Eye contact is important for conversational management and establishing intimacy between partners, and its absence can reduce the perceived social presence of the communication tool)

→ When considering remote collaboration tools with a visual component, it is helpful to draw distinctions between shared visual information pertaining to the state of a task (shared
visual workspace) and visual information depicting the state of the remote partner.

Evaluating Communication Tools
Two commonly used metrics are the quality of task solutions achieved and their times-to-completion
More nuanced insights can be gained by annotating and analyzing interactions: length, frequency, and initiation of conversational turns [37], gaps between turns [25], gaze patterns [4], and overlapping turns [43]
More subjective measures are obtained through user surveys.

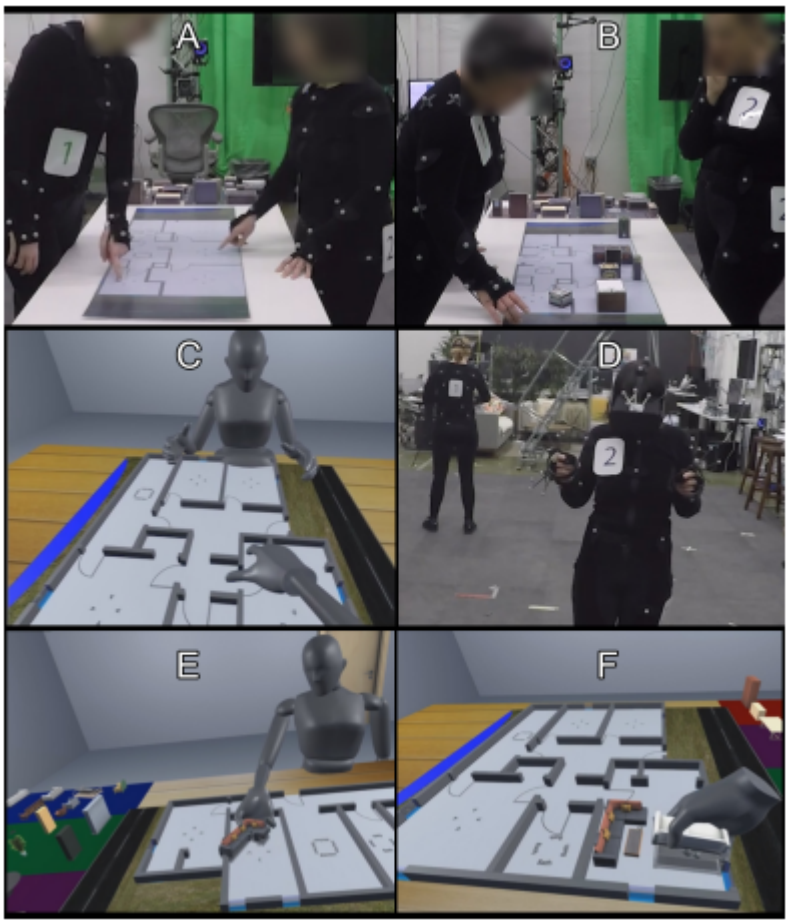
Gesture Type	Description
Reference Object or Location	Deictic (or pointing) gesture to an object or location.
Reference Person	Deictic gesture at self or interlocutor.
Spatial or Distance	Gestures conveying more complex spatial or distance information, such as a path through the apartment.
Backchannel	Acknowledgments of interlocutor, including head nods and manual gestures.
Representation	Metaphoric and iconic hand movements, illustrative of an idea (but not fitting in “Spatial or Distance").
Emotional or Social	Gestures conveying strong emotions or other social information.
Beat	Small movements of the hand in rhythm with the spoken prosody.
Self-adaptor	Self-manipulations not designed to communicate, such as nose scratches.

Table 1. Annotators would apply one or more of these tags to each observed gesture.

Speech Data	Description
Utterance	A section of speech. A sentence or comparable.
Pragmatic	Task related suggestions and discussion.
Social or Emotional	Strongly social or emotional utterances, such as “I’m very excited."
Non-task Discussion	Discussion not related to the task.
Backchannel	Verbal acknowledgements that indicate listening, such as “uh huh".
Complete Reference	Fully qualified references that can be completely understood from the utterance, like “I’d like room A".
Reference Pronoun	The use of terms like “this" or “that" to refer to things, such as “I’d like this room."
Conversational Turn	The duration for which one person holds the floor before the other takes over. Labeled with how the person gets the turn.
Interruption	The person takes the floor by interrupting the other.
No Marker	No clear indication of how the floor was obtained.
Verbal Hand Over	The interlocutor verbally passed the floor to the speaker.
Nonverbal Hand Over	The interlocutor nonverbally passed the floor to the speaker.

Table 2. Speech is tagged in the two levels specified, with individual tags listed below each level.

In the Face-To-Face condition, these annotations were made based on audio and video feeds from three camera angles. For the virtual reality conditions, annotations were made based on video feeds, audio and color-coded audio waveforms, POV footage for each participant, and multiple scene reconstructions in which avatars were always visible and color-coded.



Results :

Embodied VR provides a high level of social presence with conversation patterns that are **very similar to face-to-face interaction**. In contrast, providing only the shared environment was generally found to be lonely and appears to lead to degraded communication.

This suggests that participants employed similar communication patterns in F2F and embodied virtual reality.

Importants concepts :

- Les personnes se coupent plus la parole en face-to-face qu'en embodied-VR

ANNOTATED PARTICIPANT BEHAVIOR

Statistical method to study the datas :

Repeated measures ANOVAs were run to determine if each dependent value varied significantly across the three conditions of F2F, embodVR and no_embodVR. Mauchly's test for sphericity was run on all data and correction by Greenhouse-Geiser and Huynh-Feldt were applied as needed (both of these always succeeded). Type II error was corrected for using False Discovery Rate correction. When significant variation was found in the ANOVA, Bonferronicorrected pairwise t-tests were run to determine which factors varied.

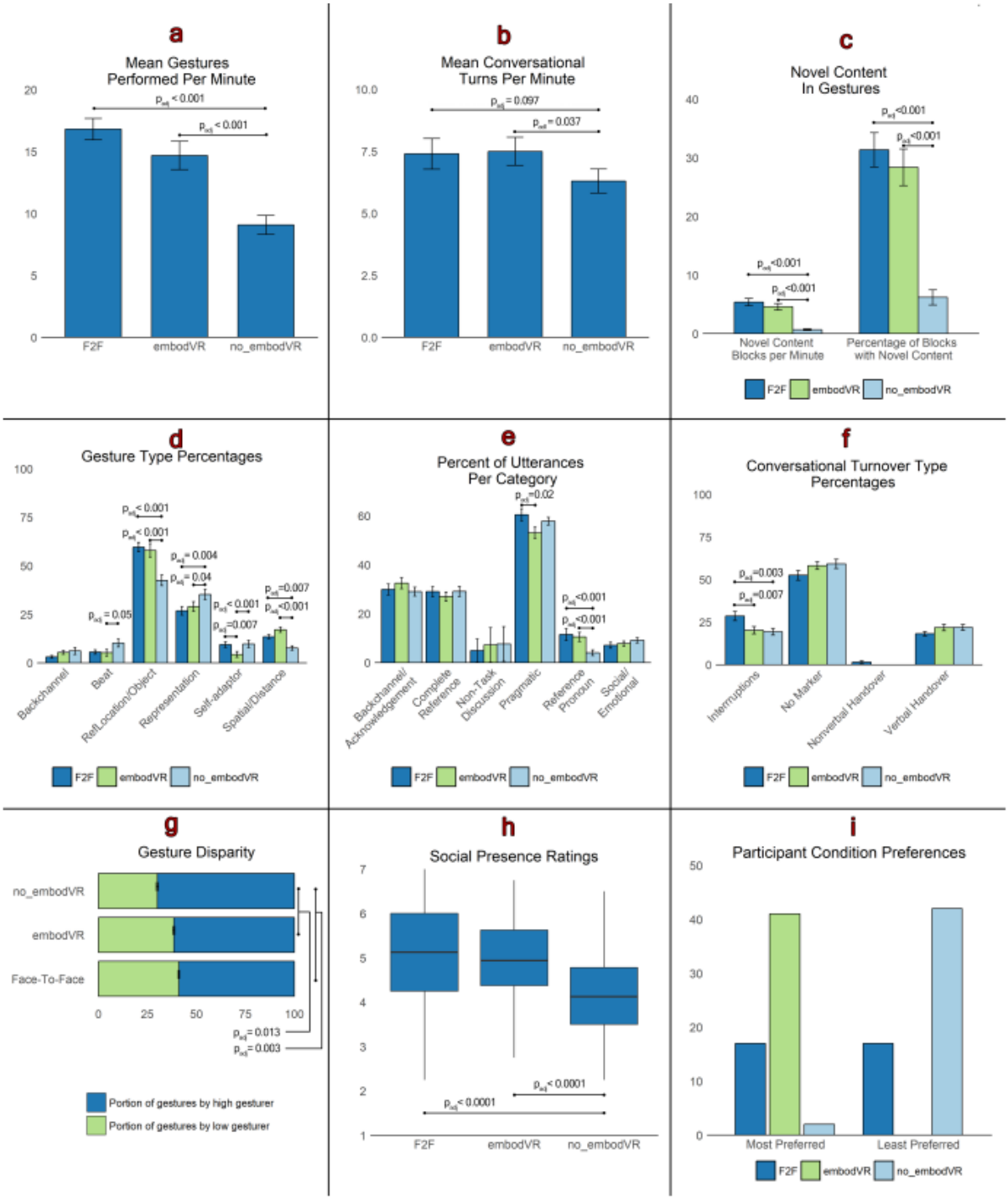


Figure 2. Sub figure a shows the average number of gestures performed per minute for each condition. Subfigure b shows the percentage of gestures that fall into each annotated category (note that, because some gestures fit multiple categories, totals for each condition can add up to over 100%). Subfigure c shows the rate and percentage of gestures which introduced novel content into the discussions (for example, point at a location while referring to it by a referential pronoun). Subfigure d shows the mean number of conversational turns taken per minute. Subfigure e shows the percent of utterances that fall in each annotated category (note that, because some gestures fit multiple categories, totals for each condition can add up to over 100%). Subfigure f shows the frequencies of the manners by which conversational turns were started. Subfigure g shows the ratio of gestures performed by the more frequent gesturer and less frequent gesturer in each dyad. Subfigure h shows the mean social presence scores, with standard errors of the mean, as measured by the semantic difference questionaire. Subfigure i shows the most and least favorite conditions, as reported by participants at the end of the experiment. All error bars show standard error of the mean.

SEMANTIC DIFFERENCE MEASURE OF SOCIAL PRESENCE

Results of a repeated measures ANOVA, followed by pairwise comparisons using paired t-tests, indicate that both F2F and embodVR showed significantly higher perceptions of social presence
than no_embodVR (medium effect size, Cohen's d of 0.62 and 0.65 respectively). There was no significant difference between F2F and embodVR (negligible effect size).

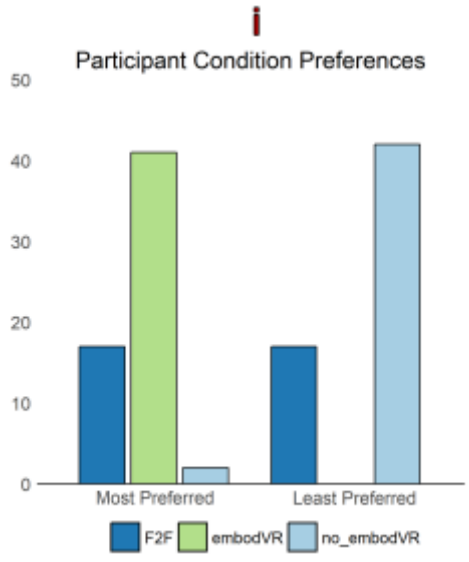
NETWORKED MINDS MEASURE OF SOCIAL PRESENCE

Factor analysis underlined : Clarity of Communication, Social Awareness, Conversation Management, and Disconnection to Partner.

Posthoc analysis showed that there was no significant difference between embodVR and F2F on Clarity of Communication, Conversation Management and Disconnection to Partner. For
Conversation Management, embodVR performed significantly better than no_embodVR. F2F and embodVR performed significantly better than no_embodVR on the other two factors, showing the same pattern as with the semantic difference measure.

PARTICIPANT PREFERENCES AND EXIT INTERVIEWS

EmbodVR was most preferred by 39 participants and least preferred by 0.
F2F was most preferred by 15 and least preferred by 17.



GLOBAL CONCLUSION

Having a tracked body in the virtual world seems to help people feel that they are really interacting with another person: all but one subjective measure showed no significant difference for social presence between F2F and embodVR, with lower social awareness possibly reflecting the lack of facial information.
There was a clear preference for including a body in the experience as people felt “alone" in no_embodVR and ratings dropped. Removing the body decreased referential pronoun
usage and lowered the frequency with which participants took conversational turns.

LIMITATIONS

1) a particular context in which users have a shared visual work space. The activities included a negotiation task and a design task. Behavior may vary for different environments and different activities.

2) while we measure conversational behavior and subjective experience, we don’t measure the effectiveness of the conversation. For example, it would be interesting to examine social conversation to see whether facial motion plays a more dominant role here.

Method :

30 dyads complete two tasks using a shared visual workspace: negotiating an apartment layout and placing model furniture on an apartment floor plan. Dyads completed both tasks under three different conditions: **face-to-face**, **embodied VR with visible full-body avatars**, and **no embodiment VR**, where the participants shared a virtual space, but had no visible avatars.

To control for variations within participant pairs, we employed a within-subjects design where each dyad performed the tasks under each of the three conditions. In order to prevent subjects from reusing previous decisions in subsequent conditions, three different floor plans were utilized.

To start, they had a tutorial + Before the first task of each condition, participants played a short bluffing game (1 minute) to familiarize themselves with interacting in the condition.

Evaluation : participants’ subjective impressions and a detailed analysis of their actual verbal and nonverbal communication behavior.

Datacollection :
For all conditions, subjects were recorded with three GoPro cameras. Audio was recorded either with lapel microphones or through the HMD microphone. For the VR conditions, the
POV (Point of View video) of each participant was recorded during the interaction. In addition, the various transformations of each object within the scene were recorded at 20 frames per
second, allowing us to create videos post-hoc of the interaction with color-coded avatars, including making avatars visible that had been hidden during the No Embodiment condition.

Ideas **Questions/Remarks**