Communication Behavior in Embodied Virtual Reality

Harrison Jesse Smith, Michael Neff

CHI 2018

How embodied virtual reality (VR) can support communication around a spatial task ?

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

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

Most real-time collaboration mediums :

- Face to face
- tools that support audio communication only, such as a telephone
- tools that support audio and visual communication, such as video conferencing.

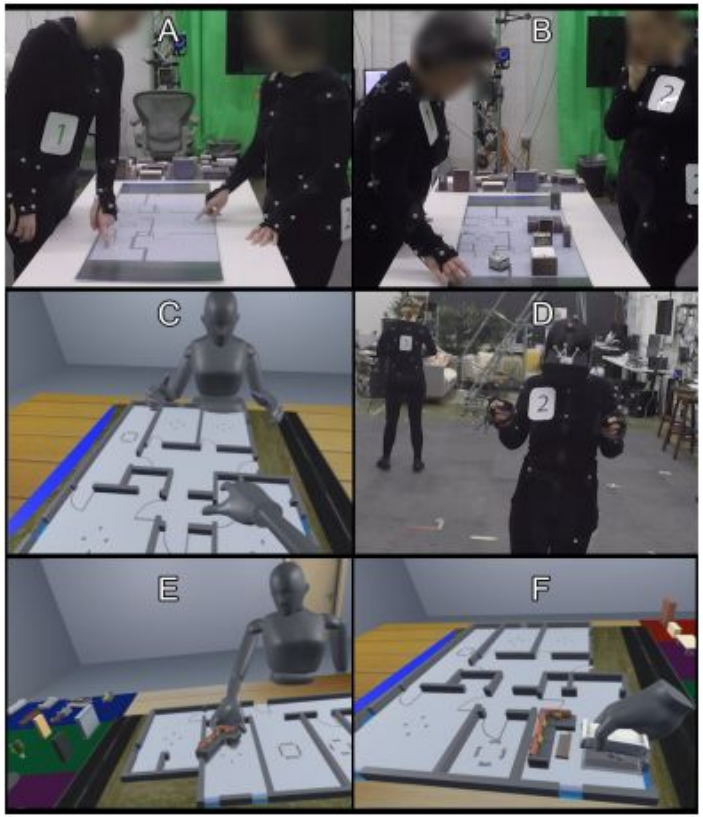
But,

- Lack of visibility and visual co-presence
- Stationary video feeds
- Almost impossible to establish mutual gaze

Collaborative task :

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

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.



01

Face to Face

02

Embodied VR

03

Non-Embodied VR

Method

- 30 dyads

- each dyad performed the tasks under each of the three conditions

- three different floor plans were utilized

- tutorial then bluffing game then tasks

Evaluation

- participants' subjective impressions

- a detailed analysis of their actual verbal and nonverbal communication behavior

Video : GoPro
Audio : lapel microphones or through the HMD microphone
VR : POV
Transformations of each object

Evaluating Communications Tools

- quality of task solutions achieved and,
- their times-to-completion

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.

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.

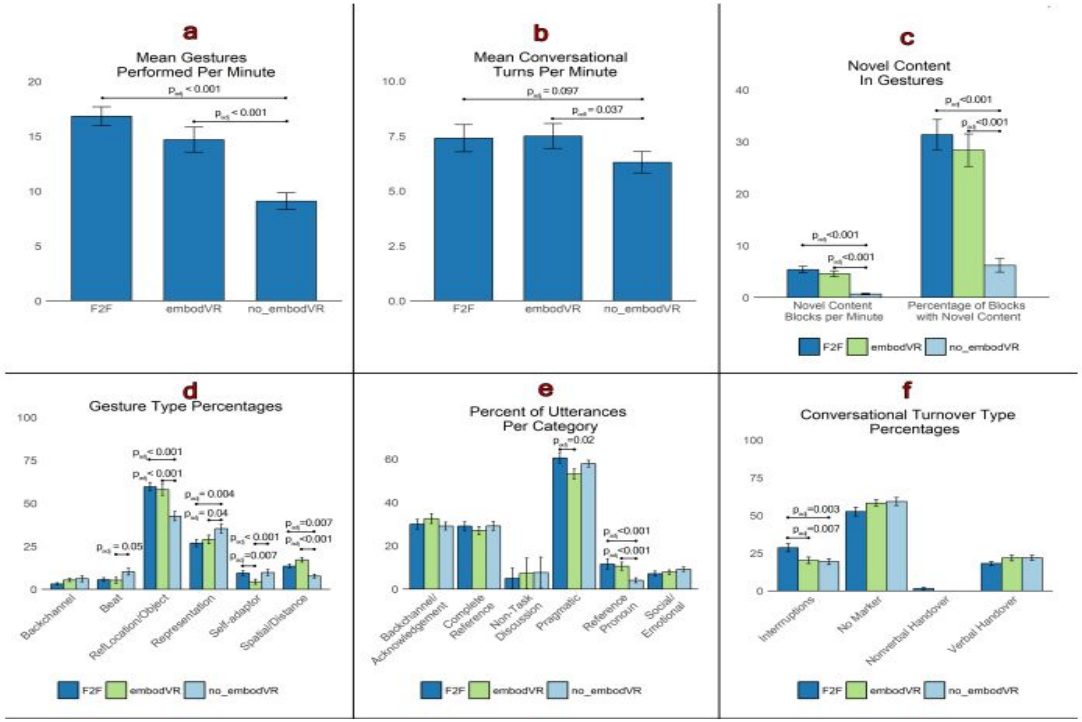
This suggests that participants employed similar communication patterns in F2F and embodied virtual reality.

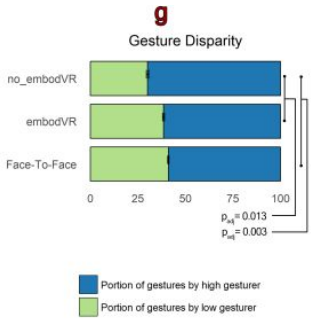
ANNOTATED PARTICIPANT BEHAVIOR

- Condition had a significant effect on the frequency of interruptions, with **interruptions occurring more frequently in the F2F condition than in either embodVR or no_embodVR** (2.c)
- In no_embodVR, the less frequent gesturer made about 30 percent of the gestures, compared to 40 percent for F2F and embodVR (2.a et 2.g)
- A higher proportion of gestures in **no_embodVR** were **representational or beats** and a lower proportion were object/location references and spatial/distance gestures. Self-adaptors were a higher proportion of gestures in F2F and no_embodVR, compared with embodVR, with no significant difference between these two categories. (2.d)
- no statistical differences between F2F and embodVR, but **gesture with novel content was significantly lower in no_embodVR**. (2.c)

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.

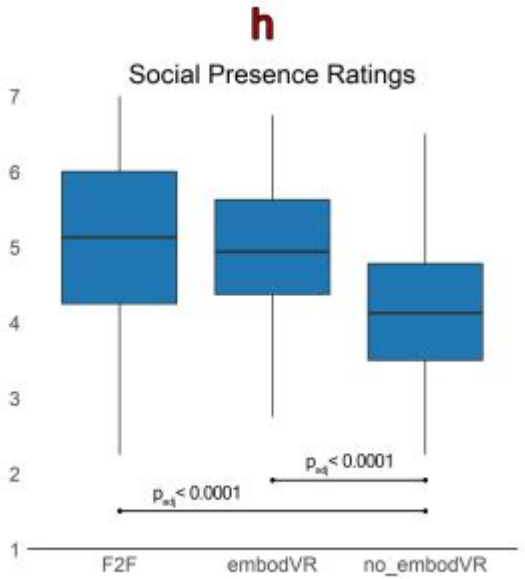
Table 2. Speech is tagged in the two levels specified, with individual tags listed below each level.


a
Mean Gestures Performed Per Minute
F2F
embodVR
no_embodVR
b
Mean Conversational Turns Per Minute
F2F
embodVR
no_embodVR
c
Novel Content In Gestures
Novel Content Blocks per Minute
Percentage of Blocks with Novel Content
F2F
embodVR
no_embodVR
d
Gesture Type Percentages
Backchannel
Beat
Ref/Location/Object
Representation
Self-adaptor
Spatial/Distance
F2F
embodVR
no_embodVR
e
Percent of Utterances Per Category
Backchannel/ Acknowledgement
Complete Reference
Non-Task Discussion
Pragmatic
Reference Pronoun
Social/ Emotional
F2F
embodVR
no_embodVR
f
Conversational Turnover Type Percentages
Interruptions
No Marker
Nonverbal Handover
Verbal Handover
F2F
embodVR
no_embodVR


g
Gesture Disparity
no_embodVR
embodVR
Face-To-Face
Portion of gestures by high gesturer
Portion of gestures by low gesturer

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 :

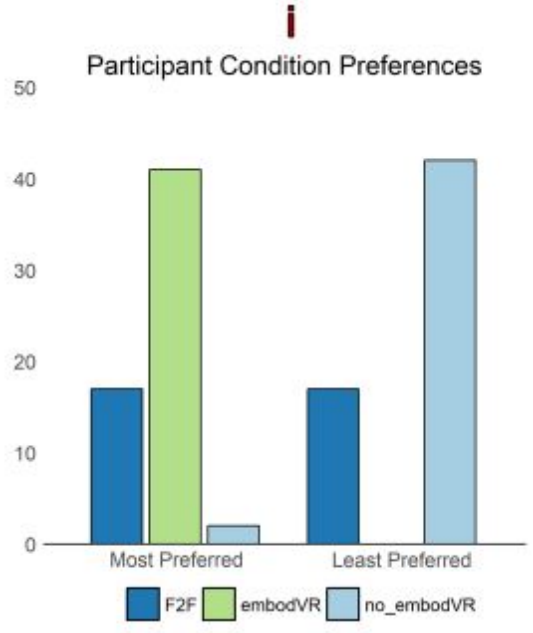
- Clarity of Communication,
- Social Awareness,
- Conversation Management,
- Disconnection to Partner.

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.



General Conclusion

😃 Tracked body

😔 No tracked body

- Better results
- Except for social awareness
  + people preferred it

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.