A Communication-Focused Framework for Understanding Immersive Collaboration Experiences

Jerald Thomas* Sang Won Lee† Alexander Giovannelli‡ Logan Lane§ Doug Bowman¶

Department of Computer Science & Center for Human-Computer Interaction, Virginia Tech, Blacksburg, VA, USA

Abstract

The ability to collaborate with other people across barriers created by time and/or space is one of the greatest features of modern communication. Immersive technologies are positioned to enhance this ability to collaborate even further. However, we do not have a firm understanding of how specific immersive technologies, or components thereof, alter the ability for two or more people to communicate, and hence collaborate. In this work-in-progress position paper, we propose a new framework for immersive collaboration experiences and provide an example of how it could be used to understand a hybrid collaboration among two co-located users and one remote user. We are seeking feedback from the community before conducting a formal evaluation of the framework. We also present some future work that this framework could facilitate.

Index Terms: Human-centered computing—Human computer interaction (HCI)—HCI theory, concepts and models; Human-centered computing—Collaborative and social computing theory—Theory, concepts and paradigms; Human-centered computing—Collaborative and social computing theory—Systems and tools;

1 Introduction

Immersive technologies, such as virtual reality (VR) and augmented reality (AR), are impacting how users work, play, and interact with one another. One area that is particularly well studied is immersive collaboration. Also sometimes termed collaborative virtual environments [18], immersive collaboration is generally defined as any form of collaboration between two or more people facilitated by immersive technologies. There has been extensive research on immersive collaboration in areas such as training [8,13], learning [3,5], and groupwork [11,16]. We are interested in studying how hybrid-location collaboration affects the group's ability to collaborate when using various immersive technologies. We define hybrid-location collaboration as any collaborative activity where at least two participants are located in the same physical environment (co-located) and at least one participant is in a different physical environment (remote). Research in immersive collaboration has focused on co-located [7,9] or remote collaboration [1,14], but there is relatively little work on immersive collaboration with hybrid locality (with some notable exceptions [2,15]).

We have not been able to find a formal framework for describing how different immersive technologies affect users' ability to communicate, which motivated us to develop such a framework in order to understand how technology choices might influence communication among users, and hence the entire collaboration. While collaboration with hybrid locality is our primary interest, this framework should work for all collaboration scenarios using immersive technologies including fully co-located and remote, asynchronous and synchronous, as well as symmetric and asymmetric collaborative processes.

To obtain an analytical understanding of how various immersive technologies can impact the ability of two or more people to collaborate, we must first have a formal method for describing and classifying the abilities for users to communicate with each other when using said technologies. This is a standard practice in the Human-Computer Interaction (HCI) and Computer Supported Collaborative Work (CSCW) fields, with notable examples being Johansen's time-space matrix [6], the classification of collaborative territoriality by Scott et al. [17], and the framework for describing workplace awareness in collaborative settings [4] by Gutwin and Greenberg. Each of these frameworks provides researchers with the ability to categorize, and in some cases quantify, aspects of collaboration scenarios in order to understand and compare them [7].

We are developing the Pairwise Immersive Collaborative Communication (PICC) framework to address the need for a method to quantitatively describe the ability for immersive technology users undergoing a collaborative task. The framework can be used in much the same way as the other aforementioned frameworks, and works by rating pairs of users' ability to communicate with each other through their self-representation (e.g. non-verbal avatar expressions) and their environment (e.g. writing on a shared virtual whiteboard). The goal of this position paper is to present the PICC framework, and we are presenting at this workshop with the intention of garnering feedback from the research community before designing and executing formal evaluations to test and validate it.

2 The PICC Framework

The PICC framework is intended to describe two individuals' ability to communicate in an immersive collaboration setting. We intentionally reduced the problem to pairs of users, instead of the group as a whole, to provide the ability to examine where communication deficiencies might exist. For example, imagine a (non-immersive) collaboration scenario with hybrid location, where co-located users are in a physical meeting room and remote users are participating via videoconference on a large television at one end of the room. In this scenario, it might be the case that overall collaboration is not very effective, but only by examining individual pairs of users can we understand why (e.g., communication among co-located users is high quality, communication among remote users is moderately effective, but communication between one co-located and one remote user is poor).

2.1 Framework Scales

The underlying premise of the PICC framework is that in order for User X to communicate with User Y, User X has to somehow affect User Y's perception. There is a myriad of conceivable ways for User X to affect User Y's perception, but for scenarios using immersive technologies, we have reduced the set of objects that User X can alter to: 1) User X's self-representation and 2) User Y's environment.

*e-mail: jeraldlt@vt.edu
†e-mail: sangwonlee@vt.edu
‡e-mail: agiovannelli@vt.edu
§e-mail: logantl@vt.edu
¶e-mail: dbowman@vt.edu

We define self-representation broadly as a user's representation as the other user perceives it. If the two users are co-located and using see-through AR, then the appearance of their physical bodies is the visual component of their self-representation (assuming there is no further augmentation). We also consider other non-visual sensations to be part of a user's self-representation, including their voice and their touching of the other user. A user's environment is defined as everything that a user can sense, apart from the other user's self-representation. This includes a user's perception of 1) physical objects in their own physical environment and 2) purely virtual objects displayed by the system. For example, User X could manipulate User Y's environment by remotely controlling a robot in User Y's physical environment, or by writing on a virtual whiteboard that User Y can observe.

Thus, User X's ability to communicate an idea to User Y can be described using the following four scales, each of which gets rated on a scale between zero and one:

- **Avatar Control:** User X's ability to modify their self-representation. This scale represents the affordances that a technology provides for a user to alter their self-representation. A score of 0 could mean that there is no self-representation to control, or that the self-representation is static and does not change in response to User X's actions. Conversely, a score of 1 could denote a fully-controllable self-representation, which typically would only occur when their physical being is their self-representation. Some factors that could affect this score include gestural expressiveness, facial expressiveness, and tracking fidelity.

- **Avatar Fidelity:** User Y's ability to perceive modifications made to User S's self-representation. This scale represents the limitations that a technology imposes on a user that limits their ability to perceive the changes made to the other user's self-representation. A score of 0 could mean that they are not able to perceive the other user's self-representation at all. A score of 1 represents a lack of limitations imposed by the technology, for example because the two users are co-located and using see-through AR designed not to occlude the users' faces. Some factors that could affect this score include field of view, display processing power, and opacity.

- **Environment Control:** User X's ability to modify User Y's environment. This scale represents the affordances that a technology provides for a user to alter the other user's environment. A score of 0 would mean that User X is not able to modify User Y's environment, and a score of 1 could mean that User X has complete control over User Y's environment. As we defined an environment as everything a user can sense, it should be impossible for one user to have complete control over another user's environment. Some factors that could affect this score include available interaction techniques and the placement of the environment on the reality-virtuality continuum [10].

- **Environment Fidelity:** User Y's ability to perceive modifications made to their environment. This scale represents the limitations that a technology imposes on a user that limits their ability to perceive the changes made to their environment. A score of 0 should not be possible, as it would mean that the user has no perception of their environment. A score of 1 would mean that the user can perfectly perceive any change made to their environment. Some factors that could affect this score include field of view, resolution, and display processing power.

2.2 Calculating the Communication Score

Communication among a pair of users has a directionality as well. To fully describe User A and B's ability to communicate with each other, we have to determine ratings for the four scales separately for A-to-B communication and B-to-A communication. For a given pair of participants, each of the scales is rated from 0 to 1, and a mean is calculated to represent the "communication score" for the pair. The values for each scale are impacted by the technologies that each user is using, the environment that each user is immersed in, and any time-space boundaries between the users, hence a rating of pairwise communication using this framework is heavily dependent on the scenario.

To calculate a communication score, each scale is rated for both the User X to User Y direction and the User Y to User X direction. Then two super-scales, Avatar and Environment, are calculated for each direction. The Avatar super-scale is calculated as the minimum of the Avatar Control and Avatar Fidelity scales. Similarly, the Environment super-scale is calculated as the minimum of the Environment Control and Environment Fidelity scales. These super-scales represent the ability for a pair of users to communicate via self-representations or the environment. The ability to communicate will be restricted by the lower score representing the ability of a user to control their self-representation/environment, and the ability of the other user to perceive the changes. By taking the minimum of these values, we can see how the self-representation and environment components contribute to the pair's ability to communicate. The final communication score is then calculated by simply taking the arithmetic mean of the four super-scales.

3 Example

To understand how the PICC framework describes communication in immersive collaboration, we present an example of how the framework scales could be rated for a given technology setup and the resulting communication scores. Scores for each of the scales, for each pair of participants, and for each direction are shown in Table 2.1, along with the Avatar and Environment super-scales and the final pairwise communication scores. These ratings were determined by consensus among the authors during a joint discussion.

3.1 Scenario

This scenario involves three users (A, B, and C) trying to complete a collaborative task. User A and User B are co-located, and User C is participating remotely. All users are participating synchronously. User A is using a Microsoft HoloLens 2 with hand tracking. There are no other trackers on their body, so the system only has the position and orientation of their head, and their hands if they are within view of the on-board tracking cameras. User B is not using any immersive technology, and as such cannot sense any of the virtual information shared by User A and User C. However, User B is able to communicate with User C via voice communication. User C is participating from a remote location, using a Valve Index head-worn display. The system takes the physical environment mapping data from User A's Hololens 2 and uses it to recreate the physical environment shared by User A and User B as a virtual environment that User C inhabits. User C's system has accurate full-body tracking to control a realistic virtual representation, but there is no facial tracking.

3.2 Grading Explanation

User A and User B: Both users have complete control over their self-representations since they are communicating to each other through their physical bodies, thus their avatar control scores are 1.0. User B can perceive User A's self-representation perfectly as well, since User B is not limited by any technology. However, User A's perception of User B is slightly hindered by the HoloLens 2's semi-transparent lens and the possibility that virtual objects could occlude their view of User B. Since User B is only perceiving the physical environment, User A has almost complete control over that environment (with the exception of User B's body and other private

	Avatar Control	Avatar Fidelity	Environment Control	Environment Fidelity	Avatar Super-Scale	Environment Super-Scale	Communication Score
A → B	1.0	1.0	0.9	1.0	1.0	0.9	0.85
B → A	1.0	0.8	0.7	1.0	0.8	0.7	
A → C	0.3	1.0	0.9	1.0	0.3	0.9	0.5
C → A	0.5	0.8	0.3	0.6	0.5	0.3	
B → C	0.2	1.0	0.7	1.0	0.2	0.7	0.275
C → B	0.5	0.2	0.0	1.0	0.2	0.0	

Table 1: Scores for the individual scales, the combined Avatar and Environment super-scales, and final communication scores for the example described in Section 3.

objects), so we rate this as 0.9. However, User B can't control any of the virtual objects perceived by User A, so this rating is lower at 0.7. Both User A and User B have perfect perception of the physical environment around them, since this is not mediated by the technology.

User A and User C: User A has very little control over their self-representation as viewed by User C, only being able to control their voice, the avatar head position and orientation, and the avatar hands when the system is able to track them, thus we rated User A's avatar control at 0.3; User B can perceive any change that User A makes to their self-representation. User C has more control over their self-representation as they have full-body tracking, but because they are lacking facial tracking, we still only rate User C's Avatar Control as viewed by User A at 0.5. User A's ability to perceive changes made to User C's self-representation is diminished by the Hololens 2's small field of view, so the Avatar Fidelity in this direction is rated at 0.8. User A has nearly full control of User C's environment, but User C can only control the virtual aspects of User A's environment, hence a lower score of 0.3.

User B and User C: User B has very little control over their self-representation, as they are only able to communicate with User C via voice, and User C has the same control as with User A, hence Avatar Control scores of 0.2 and 0.5 respectively. User C can perfectly observe User B's changes to their self-representation, but User B can only perceive voice-related changes that User C makes to their self-representation, so we rated Avatar Fidelity at 1.0 and 0.2. User B is able to somewhat control User C's environment as it is a recreation of User B's physical environment, but User C has no control over User B's environment as User C can only manipulate the virtual components and User B is not using technology that allows them to sense virtual objects.

The scores on the scales result in communications scores of 0.85, 0.5, and 0.275 for User A and User B, User A and User C, and User B and User C respectively. The purpose of this example is not necessarily to demonstrate how to analyze these values, but there are interesting surface-level observations to make. The major takeaway is that User C is going to have a hard time communicating in this scenario, especially with User B. We could hypothesize that this will result in poorer overall collaboration, and potentially alter the technologies used until the scores become more even, potentially by having all three users use a completely virtual environment for example.

4 Potential Research Applications

There are several potential applications for the PICC framework. Here, we describe two: exploring hybrid location collaboration and quantifying the effects of a user moving from one immersive technology to another in cross-reality scenarios.

Hybrid location collaboration is well known to be a difficult problem, sometimes resulting in less effective collaboration than if all participants were remote [12]. Immersive technologies have several benefits that could improve the ability for all participants to effectively collaborate together. Through experience with hybrid location scenarios, we believe that a major hurdle is that all the co-located individuals can communicate much easier with each other than the remote participants can, thus creating a rift among the participants. This often results in the co-located participants easily collaborating, and the remote participants not being able to contribute as readily. However, we would first need to confirm this hypothesis with a formal evaluation, and this framework would allow us to quantify each pair's ability to communicate and use that as an independent variable in our experiment. A follow-up evaluation using the framework could explore how to best improve hybrid location collaboration by altering the technologies used.

Evaluating cross-reality scenarios provides another great example of how the PICC framework can be used. In cross-reality scenarios, one or more users might be changing what immersive technology they are using in a given scenario. To understand the effects of these changes, we must first have the ability to somehow quantify the system before and after the change. An evaluation to explore these effects could use this framework to quantify a particular scenario and then obtain a desired dependent variable, such as performance on a collaborative task. Then one or more users could change the immersive technologies they use, which in turn will change the score provided by this framework. The same dependent variable would be measured, and then a pre- and post-change comparison can be made.

5 Limitations and Future Work

While this framework will undoubtedly be useful for research in our field, there are some limitations with the framework as it is. A goal of presenting this work, in this state, at this workshop is to get community feedback about what limitations should be addressed, as well as some ideas on the best way to improve the framework. Probably the largest limitation of this framework is that rating each scale is an inherently subjective process. The effect of this subjectivity could be reduced by having multiple raters, but individual factors such as culture or experience might make inter-rater reliability low. To address this, one of our ideas to evaluate this framework is to present several scenarios to several raters and determine what the inter-rater reliability is for various factors, such as geographic location or amount of experience in the field. Further future research might explore the creation of guidelines for more repeatable ratings. Another major limitation is that raters would need to possess a deep understanding of each scenario, the technologies involved, and how they would likely interact with each other in order to determine an accurate score. Finally, it is probable that the scales and super-scales selected might not be sufficient to accurately represent all scenarios, and that using the mean to calculate the communication score is too simplistic. As we move forward, we want to determine a way to change the scales and/or super-scales used, or how they are weighted, to more accurately represent any specific scenario.

6 Conclusion

In this position paper, we introduced the concept for the PICC framework, which quantifies the ability for a pair of users to communicate in a specified scenario. The framework was specifically developed with immersive technologies and collaborative scenarios in mind and provides a new way to examine immersive collaboration. The goal of presenting this work at this workshop is to obtain feedback from the community on the structure and usage of the framework before conducting evaluations to validate it.

References

[1] J. Amores, X. Benavides, and P. Maes. Showme: A remote collaboration system that supports immersive gestural communication. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1343–1348, 2015.

[2] S. Beck, A. Kunert, A. Kulik, and B. Froehlich. Immersive group-to-group telepresence. *IEEE transactions on visualization and computer graphics*, 19(4):616–625, 2013.

[3] W. Bricken. Learning in virtual reality. 1990.

[4] C. Gutwin and S. Greenberg. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW)*, 11(3):411–446, 2002.

[5] R. L. Jackson and E. Fagan. Collaboration and learning within immersive virtual reality. In *Proceedings of the third international conference on Collaborative virtual environments*, pp. 83–92, 2000.

[6] R. Johansen. *Groupware: Computer support for business teams*. The Free Press, 1988.

[7] K. Kiyokawa, M. Billinghurst, S. E. Hayes, A. Gupta, Y. Sannohe, and H. Kato. Communication behaviors of co-located users in collaborative ar interfaces. In *Proceedings. International Symposium on Mixed and Augmented Reality*, pp. 139–148. IEEE, 2002.

[8] G. Kurillo, R. Bajcsy, K. Nahrsted, and O. Kreylos. Immersive 3d environment for remote collaboration and training of physical activities. In *2008 IEEE Virtual Reality Conference*, pp. 269–270. IEEE, 2008.

[9] B. Lee, X. Hu, M. Cordeil, A. Prouzeau, B. Jenny, and T. Dwyer. Shared surfaces and spaces: Collaborative data visualisation in a co-located immersive environment. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1171–1181, 2020.

[10] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, 77:1321–1329, 1994.

[11] J. Mortensen, V. Vinayagamoorthy, M. Slater, A. Steed, B. Lok, and M. Whitton. Collaboration in tele-immersive environments. In *EGVE*, vol. 2, pp. 93–101, 2002.

[12] T. Neumayr, B. Saatci, S. Rintel, C. N. Klokmose, and M. Augstein. What was hybrid? a systematic review of hybrid collaboration and meetings research. *arXiv preprint arXiv:2111.06172*, 2021.

[13] K. Patel, J. N. Bailenson, S. Hack-Jung, R. Diankov, and R. Bajcsy. The effects of fully immersive virtual reality on the learning of physical tasks. In *Proceedings of the 9th Annual International Workshop on Presence, Ohio, USA*, pp. 87–94, 2006.

[14] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–13, 2018.

[15] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs. The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pp. 179–188, 1998.

[16] D. Roberts, R. Wolff, O. Otto, and A. Steed. Constructing a gazebo: supporting teamwork in a tightly coupled, distributed task in virtual reality. *Presence*, 12(6):644–657, 2003.

[17] S. D. Scott, M. S. T. Carpendale, and K. Inkpen. Territoriality in collaborative tabletop workspaces. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pp. 294–303, 2004.

[18] A. Steed and R. Schroeder. Collaboration in immersive and non-immersive virtual environments. In *Immersed in Media*, pp. 263–282. Springer, 2015.