Human Interaction and Emerging Technologies (IHIET-AI 2024), Vol. 120, 2024, 260–269
https://doi.org/10.54941/ahfe1004576

A Framework for Assessing and Enhancing Social and Spatial Presence in Mediated Communication to Support Remote Collaboration

Liv Ziegfeld[1], Maarten Michel[1], and Ivo Stuldreher[2]

[1]Human Machine Teaming, Netherlands Organisation for Applied Scientific Research (TNO), Soesterberg, The Netherlands
[2]Human Performance, Netherlands Organisation for Applied Scientific Research (TNO), Soesterberg, The Netherlands

ABSTRACT

Although extended reality (XR) technologies offer promising new ways to interact and collaborate, adoption of XR in real world applications comes with significant challenges. This is especially true for remote collaboration settings, which are often characterized by heterogeneity. Knowledge is transferred from an expert to a novice, users are located in different physical locations with different systems and often only a limited number of users perform activities in the physical world. In this setting an especial challenge is a lack of guidance on how to measure and improve the quality of mixed reality interaction. Presence, specifically social presence and spatial presence, are two key aspects to measure the quality of experience in XR and computer-mediated communication in general. A multitude of social presence measures is available, but few studies attempt to provide a comprehensive way to measure all underlying factors of presence that affect the quality of experience, much less offer support on how to improve on these factors. Moreover, most measures of presence focus solely on XR applications, while a mix of communication systems can be used. In this paper, we aim to address these issues and propose a comprehensive framework for the assessment and enhancement of social and spatial presence based on existing literature, regardless of the actual technology used. The framework provides a breakdown of relevant subscales of spatial and social presence in remote collaboration, easy-to-use means to measure these subscales and offers different degrees of system recommendations based on social cues that have been found to evoke presence on the respective subscales. Lastly, our framework includes practical suggestions for accommodating different types of users within a use case to ensure quality of experience for all users involved. Our framework should enable XR-developers and end-users to optimize their system for their specific application.

Keywords: Social presence, Spatial presence, Remote collaboration, Mediated communication

INTRODUCTION TO THE FRAMEWORK

Providing one's expertise and knowledge at a distance can be of great benefit, as the global job market experiences a severe lack of specialized and skilled workers in various domains, such as the energy sector, high-tech industry and manufacturing, health care and education. Leveraging eXtended Reality (XR) technology to offer remote assistance and expertise allows teams across the globe to collaborate during tasks and operations, or to have effective virtual meetings that can enhance remote education. XR-based remote assistance applications have been explored in settings of remote professional assistance and training when performing skilled physical tasks, such as field maintenance (Aschenbrenner et al., 2018; Cachada et al., 2019) or surgery (Tang et al., 2019; Gasques et al., 2021).

Though promising, adoption of XR in real world applications comes with significant challenges. This is especially true for remote collaboration settings, which are inherently non-symmetrical (Wang et al., 2021). Users are located in different physical locations, with access to different systems to communicate. Moreover, users will often have different levels of expertise on the collaborative task at hand. As a consequence, it is vital for participants to understand their counterpart(s), and the virtual or physical task space, in order to perform tasks successfully.

However, to the best of the writers' knowledge, there is no comprehensive overview of measurements and pragmatic ways to improve the quality of experience (QoE) in mixed reality collaboration. A multitude of questionnaires exists that measure either interaction aspects, like the Networked Minds Measure (Harms & Biocca, 2004) or immersive aspects (e.g. Hartmann et al., 2016; Vorderer et al., 2004). Most measures, however, focus solely on XR applications, while a mix of communication systems is often the case for remote collaboration settings. Although Toet et al. (2022) do provide a measure of the overall QoE in their Holistic Mediated Social Communication Questionnaire (H-MSC-Q), the resulting score is based on five high level cognitive processing levels, providing little practical guidance on where to improve upon to increase the QoE. Studies on improvement of QoE that do exist focus on one aspect of the solution, not considering other possible actions that could improve QoE.

To facilitate the development and evaluation of – mostly – XR-based tools in currently unexplored settings, we here describe a framework that can assist developers and users of XR-technology in measuring the QoE and in defining the functional and system requirements to ensure effective remote collaboration. We hope to assist both use cases where a new system has to be developed, and use cases where an existing technological solution needs to be measured and improved upon. In order to do so, we deconstruct two pillars of QoE: social presence and spatial presence into six subscales, provide measures and ways to achieve a baseline level of each subscale, and describe recommendations to further enhance on each subscale based on existing literature. We lastly demonstrate how to apply the framework to an example use case and show that by taking the asymmetric interaction in account, different users of the same system may require optimization for different factors underlying presence. The framework will be outlined in the following, while the full framework can be found online: https://osf.io/3yab9/.

SOCIAL PRESENCE AND SPATIAL PRESENCE

A key factor in determining the QoE in XR and computer-mediated communication in general is presence (Skowronek et al., 2022; Toet et al., 2022). Effective mediated social communication should provide for a sense of both social presence and spatial presence (Toet et al., 2022). The difference between these two concepts is that social presence primarily deals with human-human relations (Lombard & Jones, 2015), whereas spatial presence deals with human-object relations (Toet et al., 2022). Enhancing the sense of social and spatial presence is important, as it allows for more natural interaction with the remote environment and the people in that environment. XR provides ample opportunity to improve the level of presence, as traditional methods for remote collaboration (video/audio) lose important non-verbal cues, like gaze direction, gestures, depth perception, and facial expressions (Wang et al., 2021).

There are some general system recommendations to enhance the sense of social and spatial presence. Firstly, latency should be prevented as much as possible, as synchronous communication is central to many remote collaboration use cases. Be it in emergency circumstances where a remote expert must explain to a local worker how to fix a problem as quickly as possible, or in a remote work meeting where real-time behavioral feedback responses such as nodding and blushing are important to allow for an affective interaction: Delays can strongly limit the interactivity of many mediated communication systems (Yim et al., 2017; Oh et al., 2018; Rosenthal-von der Pütten et al., 2018).

Further, evoking realism is important. Behavioral realism is especially vital for achieving a high quality of experience in terms of social presence in mediated communication (Oh et al., 2018). This means that the representations and the behaviors (e.g. eye gaze, body language) of the users in the virtual shared environment should reflect real humans as much as possible to allow for natural interactions, while also being cautious of uncanny valley effects (Sharan et al., 2022). A system should ideally be able to transfer social cues of high quality, for instance high resolution video transmission via high bandwidths (Oh et al., 2018; Voinea et al., 2022). For spatial presence, sensory realism is of greater importance (Slater, 2003). This relates to the fact that virtual environments should reflect real environments as much as possible, for instance in terms of dimensions and in the movement through the environment.

Questionnaires have been developed to measure spatial and social presence on a general level. Toet et al. (2022) developed the Holistic Mediated Social Communication Questionnaire (H-MSC-Q) to measure the quality of mediated social communication experiences. This questionnaire applies a comprehensive, general, and holistic multi-scale approach, based on an established conceptual framework for multisensory perception. Thereby it is very suited for obtaining an overall indication of the levels of spatial and social presence.

SIX SUBSCALES OF PRESENCE

Although the H-MSC-Q provides an overall score on spatial presence and social presence, we argue that for optimizing a mediated-communication system to the users' specific needs, a more in-depth framework is needed. Therefore, we have consulted existing literature on presence, and identify six subscales of presence. Expanding upon the work of Toet et al. (2022), we identified three components of social presence, namely co-presence – the sense of being physically together with one's communication partner in the same environment – and the sense of having an affective and intellectual interaction with one's communication partner. Spatial presence also consists of three components, telepresence – the feeling of being located in the mediated shared environment by a sense of self-location and by obtaining a spatial situation model, ownership – the feeling of self-attribution of the virtual self, and agency – the feeling of being able to act within that environment. Figure 1 depicts the relation between the described concepts.

Affective Communication

The concept affective interaction refers to the ability to which mediated communication allows the development of an emotional connection. Affective interactions are of importance in every social interaction, but especially for close relationships. The *'perceived affective understanding'* factor from the Networked Minds Social Presence Measure (Harms & Biocca, 2004) has been adopted as a measure for affective interaction. Users can judge the importance of evoking an affective interaction in their specific use case by reflecting on the following statement: "It is essential that emotions and non-verbal communication can be shared between users." As a baseline requirement, users must have a means of communicating their (intensity of) sentiments, for instance via text-based communication through chats, emails or emoticons (Harris & Paradice, 2007; Derks et al., 2008). Since social presence is built up more slowly when restricted to text-based expressions of emotions, users who wish to increase the social presence evoked beyond the baseline should increase the number and quality of verbal and nonverbal social cues. For instance, adding audio and video can increase social presence (Gunawardena & Zittle, 1997; Oh et al., 2018). Especially facial expressions and eye-gaze are important in communicating emotional and mental states (Sharan et al., 2022; Gjoreski et al., 2023).

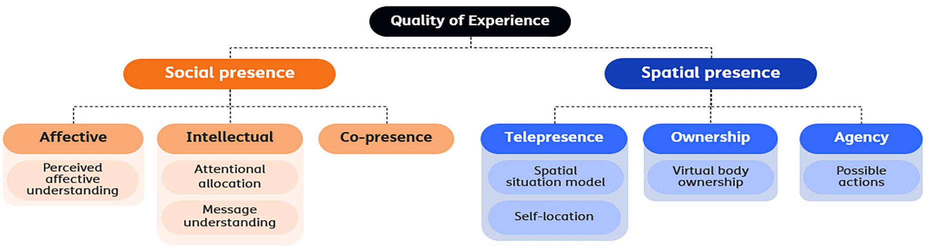

Figure 1: Overview of key factors for in QoE in mediated communication. Top level: social presence and spatial presence, divided into the six subscales. Bottom level: sub-components that are included for each subscale, as labelled by the respective questionnaires they originate from.

Intellectual Connection

Following a review of the literature to identify existing metrics to measure the different components of social presence, it was decided to further divide *intellectual interaction* into the sub-components '*attention allocation*' and '*perceived message understanding*', according to the Networked Minds Social Presence Measure by Harms and Biocca (2004). For 'attention allocation', it is important that the communication media allow for synchronous feedback modules. Users should be able to follow each other's actions and monitor each other's attention in real-time, for instance via mouse tracking. An improvement in attention allocation can be achieved by increasing the signals revealing shared attention, for instance by transmitting facial expressions and gaze through video. Backchannels are also important for the factor 'perceived message understanding', although synchronous backchannels are here seen as the enhancement of the baseline requirement of asynchronous backchannels. For instance, a chat function can be used to convey the understanding of a message, while this can be enhanced by adding auditory (e.g. 'mhm') and visual (e.g. nodding) backchannels.

Co-Presence

The concept of co-presence refers to the degree to which an individual believes they are not alone and secluded in their environment, their level of peripheral or focal awareness of the other, and their sense of the degree to which the other is peripherally or focally aware of them (Harms & Biocca, 2004). Co-presence can be measured using the 'co-presence' items from the Networked Minds Social Presence Measure (Harms & Biocca, 2004). As baseline requirement, there should be a basic representation of each user and their status (Voinea et al., 2022). This can for instance be achieved by displaying profile pictures of all users (Oh et al., 2018) or through the use of personas as is done by Microsoft (Microsoft, n.d.). To increase the feeling of co-presence, a system should allow for an increased sense of sensory proximity. This can be obtained by adding a video representation of other users, enabling non-anonymity and eye-contact (Oh et al., 2018; Sharan et al., 2022), or by adding haptic feedback (Oh et al., 2018; Voinea et al., 2022). Facilitating a shared view between users also enhances co-presence (Piumsomboon et al., 2019).

Telepresence

Telepresence refers to the degree to which users are aware of spatial dimensions and layout in the shared environment and the extent to which users feel as if they are located in the shared environment. Based on the Measurement, Effects, and Conditions Spatial Presence Questionnaire (MEC-SPQ) by Vorderer et al. (2004), telepresence was split up into the subcomponents '*spatial simulation model (SSM)*' and '*self-location*', which also reflect the measurements that can be used to evaluate telepresence. As baseline requirement, the mediated communication platform should provide a visual representation of the remote environment and objects therein to obtain a

SSM – for instance through means of a video stream or virtual environment – and there should be an indication of one's location within the shared environment to obtain a sense of self-location, for instance by seeing the location of one's mouse cursor on the screen or a map of one's position in the environment. To enhance the telepresence, the fidelity and number of depth and spatial cues in the shared environment should be increased. For instance, add stereoscopy, use a third-person view to provide better awareness of the environment (Falcone et al., 2022) or add spatial referencing by adding clear landmarks that can be used as spatial anchors to clarify the location of other objects in the shared environment (Müller et al., 2017).

Ownership

Ownership refers to the degree of self-attribution of one's representation in the shared environment. To evaluate the level of ownership one can apply the Self Presence Questionnaire (Bailey et al., 2016) and Virtual Body Ownership Illusion (VBOI) questionnaire (Argelaguet et al., 2016). A baseline sense of ownership can be achieved by some representation of the self in the shared environment. This can already be achieved by means of a profile picture or user icon. To optimize for ownership, one can increase the realism of one's own representation and actions in the shared environment. One can for instance increase the visual likeness of the surrogate with the human operator (Falcone et al., 2022). This realism is also positively affected by natural visuals, such as a normal human observable area for field of view, visuo-proprioceptive synchronicity and a first-person view (Falcone et al., 2022).

Agency

Agency refers to the feeling of being able to exercise control in the shared environment. It can be evaluated with the Spatial Presence Experience Scale (SPES) (Hartmann et al., 2016). To enable a baseline sense of agency, the user should be provided with some form of control over their own actions in the shared environment. Think for instance of the ability to control the viewpoint. To increase the sense of agency, the level of control over one's own and shared activities in the shared environment should be increased. For instance, ensure high accuracy and control in directing view (Fribourg et al., 2020; Falcone et al., 2022) and use a wide field of view (Falcone et al., 2022). Additionally, haptic feedback can be included to allow for exploring important information about the environment, but if included ensure visuo-tactile synchronicity (Falcone et al., 2022).

APPLYING THE FRAMEWORK TO AN EXAMPLE USE CASE

The six subscales of presence – affective interaction, intellectual interaction, co-presence, telepresence, agency and ownership – are all required to some extent for effective mediated social communication. Depending on the specific mediated-communication use case and role that a user has in that use case, optimization of certain factors underlying social- and spatial presence may however be of more importance than others. One can, for instance, argue that enhancing the affective interaction is more important when communicating with close friends than when communicating with a mechanic. Our framework should encourage developers and end-users to optimize their mediated communication system to their envisioned use case. We here demonstrate how the framework could be applied to an example use case, namely a remote maintenance scenario.

The scenario consists of an on-site user with a maintenance issue of a machine on-site for which a remote expert has to be contacted. The maintenance task requires physical interaction with the machine, such that the remote expert needs to guide the on-site user through the task.

By default we consider that the mediated-communication system should meet the baseline functional requirements for all six social- and spatial presence subscales outlined above. Regarding social presence, a strong affective interaction is not of too high importance in this task-oriented scenario. In terms of system requirements, seeing the facial expressions of each other (and thereby including video of each other's face) is therefore of less importance here, while audio should be included in the system so that salient information can be transmitted verbally, allowing the on-site user to have their hands free to focus on the task. The factor intellectual interaction is of more importance to both users. The on-site user should be able to explain the machinery issue to the remote expert and be able to understand the instructions from the remote-expert, who on the other hand needs to understand the problem at hand and be able to convey instructions to the on-site user. Our framework highlights shared attention and mutual understanding as relevant concepts here, which could also be covered by including audio in the system, which would for instance allow synchronous auditory backchannels as a means to evaluate comprehension and attention. While seeing the users' facial expressions is less important, it would be beneficial to allow the remote expert to have sight of the on-site user's actions through video. For the factor copresence the baseline level is sufficient, as the feeling of being in the shared environment together is less important in this scenario.

In terms of spatial presence factors, the importance of the three factors is more asymmetric. For the remote expert, telepresence is very important. The expert needs to be aware of the remote environment, its dimensions and the objects in it to be able to understand the problem at hand and give effective instructions. The user on-site however, does not need to feel strong telepresence with the environment of the remote expert. Similar reasoning holds for agency, as it is important that the remote expert experiences a sense of control of the remote environment. A resulting functional requirement could be that the system should allow the remote expert to autonomously change the camera angle in the on-site worker's environment. For the user on-site, it is important to have a sense of control in their own environment to conduct the manipulations recommended by the expert, but they do not need to have insight into, nor feel in control of, the remote expert's direct environment. For ownership the baseline requirements may be sufficient to both users.

As seen from this example use case, the proposed framework can stimulate users and developers to analyse what functional requirements exist for their system in order to meet a certain level of the social- and spatial presence subscales. Asymmetries in the requirements between the different users can also be identified with this approach, using which decisions on the technologies to be used by each user can be made.

CONCLUSION

In this paper we proposed a framework for the improvement of the quality of experience of mixed mediated communication technologies for remote collaboration settings via enhancements of social- and spatial presence. Developers and users of such technologies can employ the current framework to identify which subscale(s) of social and spatial presence are of importance to their specific use case and measure these presence levels using the proposed scales. It further supports analysing which functional requirements exist for their system and offers pragmatic suggestions for further enhancement of the experience based on previous domain literature. Hereby this paper addresses the lack of practical guidance on improving and measuring the quality of mixed reality collaboration and will allow for more effective development and usage of these technologies in the future.

REFERENCES

Argelaguet, F., Hoyet, L., Trico, M., and Lécuyer, A. (2016). The role of interaction in virtual embodiment: Effects of the virtual hand representation, in *Proceedings - IEEE Virtual Reality* doi: 10.1109/VR.2016.7504682.

Aschenbrenner, D., Rojkov, M., Leutert, F., Verlinden, J., Lukosch, S., Latoschik, M. E., et al. (2018). Comparing Different Augmented Reality Support Applications for Cooperative Repair of an Industrial Robot, in *Adjunct Proceedings - 2018 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2018* doi: 10.1109/ISMAR-Adjunct.2018.00036.

Bailey, J. O., Bailenson, J. N., and Casasanto, D. (2016). When does virtual embodiment change our minds? *Presence Teleoperators Virtual Environ.* 25. doi: 10.1162/PRES_a_00263.

Cachada, A., Romero, L., Costa, D., Badikyan, H., Barbosa, J., Leitao, P., et al. (2019). Using AR Interfaces to Support Industrial Maintenance Procedures, in *IECON Proceedings (Industrial Electronics Conference)* doi: 10.1109/IECON.2019.8927815.

Derks, D., Fischer, A. H., and Bos, A. E. R. (2008). The role of emotion in computer-mediated communication: A review. *Comput. Human Behav.* 24. doi: 10.1016/j.chb.2007.04.004.

Falcone, S., Englebienne, G., Van Erp, J., and Heylen, D. (2022). Toward Standard Guidelines to Design the Sense of Embodiment in Teleoperation Applications: A Review and Toolbox. *Human-Computer Interact.* doi: 10.1080/07370024.2022.2039147.

Fribourg, R., Argelaguet, F., Lécuyer, A., and Hoyet, L. (2020). Avatar and Sense of Embodiment: Studying the Relative Preference between Appearance, Control and Point of View. *IEEE Trans. Vis. Comput. Graph.* 26. doi: 10.1109/TVCG.2020.2973077.

Gasques, D., Johnson, J. G., Sharkey, T., Feng, Y., Wang, R., Robin, Z., et al. (2021). Artemis: A collaborative mixed-reality system for immersive surgical telementoring, in *Conference on Human Factors in Computing Systems - Proceedings* doi: 10.1145/3411764.3445576.

Gjoreski, H., Mavridou, I., Archer, J. A. W., Cleal, A., Stankoski, S., Kiprijanovska, I., et al. (2023). OCOsense Glasses - Monitoring Facial Gestures and Expressions for Augmented Human-Computer Interaction: OCOsense Glasses for Monitoring Facial Gestures and Expressions. *Conf. Hum. Factors Comput. Syst. - Proc.* doi: 10.1145/3544549.3583918.

Gunawardena, C. N., and Zittle, F. J. (1997). Social presence as a predictor of satisfaction within a computer–mediated conferencing environment. *Int. J. Phytoremediation* 21. doi: 10.1080/08923649709526970.

Harms, C., and Biocca, F. (2004). Internal Consistency and Reliability of the Networked Minds Measure of Social Presence. *Seventh Annu. Int. Work. Presence 2004*.

Harris, R. B., and Paradice, D. (2007). An Investigation of the Computer-mediated Communication of Emotions. *J. Appl. Sci. Res.* 3.

Hartmann, T., Wirth, W., Schramm, H., Klimmt, C., Vorderer, P., Gysbers, A., et al. (2016). The Spatial Presence Experience Scale (SPES). *J. Media Psychol.* 28. doi: 10.1027/1864-1105/a000137.

Lombard, M., and Jones, M. T. (2015). "Defining presence," in *Immersed in Media: Telepresence Theory, Measurement and Technology* doi: 10.1007/978-3-319-10190-3_2.

Microsoft (n.d.). Persona. Available at: https://fluent2.microsoft.design/components/web/react/persona/usage [Accessed May 7, 2023].

Müller, J., Rädle, R., and Reiterer, H. (2017). Remote collaboration with Mixed Reality displays: How shared virtual landmarks facilitate spatial referencing. in *Conference on Human Factors in Computing Systems - Proceedings* doi: 10.1145/3025453.3025717.

Oh, C. S., Bailenson, J. N., and Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Front. Robot. AI* 5. doi: 10.3389/frobt.2018.00114.

Piumsomboon, T., Dey, A., Ens, B., Lee, G., and Billinghurst, M. (2019). The effects of sharing awareness cues in collaborative mixed reality. *Front. Robot. AI* 6. doi: 10.3389/frobt.2019.00005.

Rosenthal-von der Pütten, A. M., Krämer, N. C., and Herrmann, J. (2018). The Effects of Humanlike and Robot-Specific Affective Nonverbal Behavior on Perception, Emotion, and Behavior. *Int. J. Soc. Robot.* 10. doi: 10.1007/s12369-018-0466-7.

Sharan, N. N., Toet, A., Mioch, T., Niamut, O., and van Erp, J. B. F. (2022). The Relative Importance of Social Cues in Immersive Mediated Communication, in *Lecture Notes in Networks and Systems* doi: 10.1007/978-3-030-85540-6_62.

Skowronek, J., Raake, A., Berndtsson, G. H., Rummukainen, O. S., Usai, P., Gunkel, S. N. B., et al. (2022). Quality of Experience in Telemeetings and Videoconferencing: A Comprehensive Survey. *IEEE Access* 10. doi: 10.1109/ACCESS.2022.3176369.

Slater, M. (2003). A note on presence terminology. Presence connect. *Presence Connect* 3, 1–5.

Tang, K. S., Cheng, D. L., Mi, E., and Greenberg, P. B. (2019). Augmented reality in medical education: a systematic review. *Can. Med. Educ. J.* doi: 10.36834/cmej.61705.

Toet, A., Mioch, T., Gunkel, S. N. B., Niamut, O., and van Erp, J. B. F. (2022). Towards a Multiscale QoE Assessment of Mediated Social Communication. *Qual. User Exp.* 7, 4.

Voinea, G. D., Gîrbacia, F., Postelnicu, C. C., Duguleana, M., Antonya, C., Soica, A., et al. (2022). Study of Social Presence While Interacting in Metaverse with an Augmented Avatar during Autonomous Driving. *Appl. Sci.* 12. doi: 10.3390/app122211804.

Vorderer, P, Wirth, W., Gouveia, F. R., Biocca, F., Saari, T., Jäncke, F., Böcking, S., Schramm, H., Gysbers, A., Hartmann, T., Klimmt, C., Laarni, J., Ravaja, N., Sacau, A., Baumgartner, T. & Jäncke, P. (2004). MEC Spatial Presence Questionnaire (MEC-SPQ): Short Documentation and Instructions for Application. Report to the European Community, Project Presence: MEC (IST-2001-37661). Online. Available from https://www.ijk.hmt-hannover.de/presence.

Wang P, Bai X, Billinghurst M, Zhang S, Zhang X, Wang S, He W, Yan Y, Ji H (2021) AR/MR remote collaboration on physical tasks: a review. Robot Comput Integr Manuf 72:1–32.

Yim, M. Y. C., Chu, S. C., and Sauer, P. L. (2017). Is Augmented Reality Technology an Effective Tool for E-commerce? An Interactivity and Vividness Perspective. *J. Interact. Mark*. 39. doi: 10.1016/j.intmar.2017.04.001.