

# A Classification of Human-to-Human Communication during the Use of Immersive Teleoperation Interfaces

Martin Kraus\*, Martin Kibsgaard<sup>‡</sup>

\*Department of Architecture, Design, and Media Technology  
Aalborg University

Rendsburggade 14, 9000 Aalborg, Denmark

phone: + 45 99 40 87 85, email: martin@create.aau.dk

www: <http://personprofil.aau.dk/121021>

<sup>‡</sup>Department of Architecture, Design, and Media Technology  
Aalborg University

Rendsburggade 14, 9000 Aalborg, Denmark

phone: + 45 28 83 40 54, email: kibsgaard@create.aau.dk

www: <http://personprofil.aau.dk/132170>

## Abstract

We propose a classification of human-to-human communication during the use of immersive teleoperation interfaces based on real-life examples. While a large body of research is concerned with communication in collaborative virtual environments (CVEs), less research focuses on cases where only one of two communicating users is immersed in a virtual or remote environment. Furthermore, we identify the unmediated communication between co-located users of an immersive teleoperation interface as another conceptually important — but usually neglected — case. To cover these scenarios, one of the dimensions of the proposed classification is the level of copresence of the communicating users. Further dimensions are the virtuality of the immersive environment, the virtual

transport of the immersed user(s), the point of view of the user(s), the asynchronicity of the users' communication, the communication channel, and the mediation of the communication. We find that an extension of the proposed classification to real environments can offer useful reference cases. Using this extended classification not only allows us to discuss and understand differences and similarities of various forms of communication in a more systematic way, but it also provides guidelines and reference cases for the design of immersive teleoperation interfaces to better support human-to-human communication.

**Keywords:** Telepresence, teleoperation, presence, copresence, immersion, collaboration, virtual reality, augmented reality, collaborative virtual environment, shared virtual space, computer-mediated communication, human-to-human communication.

### Digital Peer Publishing Licence

Any party may pass on this Work by electronic means and make it available for download under the terms and conditions of the current version of the Digital Peer Publishing Licence (DPPL). The text of the licence may be accessed and retrieved via Internet at <http://www.dipp.nrw.de/>.

*First presented at the Virtual Reality International Conference 2015, extended and revised for JVRB*

## 1 Introduction

Teleoperation allows human users to operate machines, in particular robots, at a distance [Min80]. Today, teleoperation is used routinely for operating surgical robots, handling dangerous materials, defusing bombs, remotely piloting aircrafts, working in outer space, the deep sea, and other hazardous environments. Furthermore, teleoperation in simulated, virtual environments is routinely used for the training of

teleoperators and many other professionals, e.g., aircraft pilots and medical personnel.

Many teleoperation interfaces attempt to immerse the human operator in a remote or virtual environment in order to achieve “telepresence,” i.e., the sense of being in the remote or virtual environment. A prime example is the use of head-mounted displays for immersive teleoperation interfaces. A disadvantage of these interfaces is the isolation of the human operators from their immediate environment. This isolation usually compromises communication with co-located humans, for example, co-workers, expert advisors, trainers, apprentices, etc., and, therefore, hinders collaboration with them.

While this isolation results in many challenging problems, research on collaboration using immersive teleoperation interfaces tends to focus on collaborative virtual environments (CVEs) [BGRP01], in which multiple users are immersed in a shared, virtual environment. However, research on the training in robot-assisted surgery [MTG<sup>+</sup>14] convinced us that communication (and therefore collaboration) in CVEs is fundamentally different from the communication between an immersed user and a non-immersed user. Yet another very different scenario is the unmediated communication between co-located users of an immersive teleoperation interface, e.g., two pilots in a flight simulator.

To gain better understanding of these differences, this work presents a systematic classification of common forms of human-to-human communication during the use of immersive teleoperation interfaces. Such a classification can provide better understanding of the scope of technical solutions for computer-mediated communication, and it can inspire new solutions by revealing similarities between different scenarios — in particular if the classification is extended to cover operation in an unmediated, real environment. Furthermore, the classification can help to design support for communication in immersive teleoperation interfaces more systematically.

After reviewing previous work in Section 2, we present our new classification in Section 3. Section 4 discusses how the proposed classification can help to design support for human-to-human communication while conclusions and future work are presented in Sections 5 and 6.

## 2 Previous Work

Minsky was one of the first to discuss the concepts of teleoperation and telepresence [Min80], where teleoperation focuses on the ability to manipulate a remote or virtual environment, i.e., a mediated environment; and telepresence focuses on the sense of being in a mediated environment. We use the term “presence” to include also the sense of being in an unmediated environment. In addition to this conceptualization of (tele)presence as a sense of transportation, Lombard and Ditton [LD97] discuss further conceptualizations.

In this work, the term “copresence” denotes the (unmediated) sense of being together in the real world as well as the (possibly mediated) sense of “being there together” as discussed by Schroeder [Sch02]. Lombard and Ditton refer to this concept as the impression of a shared space [LD97]. However, we distinguish two possibilities of “being there together”: the mediated sense of being together in a mediated environment and the unmediated(!) sense of being together in a mediated environment, as discussed in more detail in Section 3.2.1.

Copresence is known to be an important part of the experience of mediated communication situations. For example, Aaltonen et al. [ATH<sup>+</sup>09] found in an experimental study that copresence provided the clearest difference between various mediated communication situations.

Milgram et al. [MTUK95] have classified various technologies that can provide telepresence on a reality-virtuality continuum. Benford et al. [BGR<sup>+</sup>98] have generalized this classification to include collaborative virtual environments (CVEs) as well as computer-supported collaborative work (CSCW). Our classification is different in two aspects. First, it focuses on the classification of different kinds of communication. Second, our classification of communication situations attempts to include CVEs and other real-life collaborative uses of immersive teleoperation interfaces while we do not attempt to include typical CSCW systems.

One common typology of CSCW systems is represented by the CSCW matrix [BGBG95], which classifies CSCW systems based on whether users are at the same place or at different places, as well as whether the communication between users is synchronous or asynchronous. While the concept of copresence covers the spatial relation between users, asynchronous communication in CVEs has received less attention. Wu et al. [WMW<sup>+</sup>15] described space travel to Mars as one sce-

nario which can lead to asynchronous communication in virtual worlds. Therefore, we include asynchronicity as one dimension of our classification.

Our classification also covers nonverbal communication in CVEs, which was discussed by Guye-Vuillème et al. [GVCP<sup>+</sup>99]. However, we decided not to use the distinction between verbal and nonverbal communication in our classification since this distinction is often not relevant for the mediation of the communication. For example, an audio signal may or may not include nonverbal communication regardless of its mediation.

Similarly, our classification covers communication with the purpose of establishing awareness [CCS12, Gre96, GG02] even though awareness is not related to any of its dimensions.

### 3 Proposed Classification

#### 3.1 Scope and Exemplary Situations

The proposed classification is supposed to provide guidance when designing support for human-to-human communication during the use of immersive teleoperation interfaces. Therefore, we are mainly concerned with immersive interfaces, e.g., Virtual Reality (VR) simulators (see Figure 1) or remote teleoperation systems (see Figure 2).

Some of these interfaces are designed for collaborating users and, therefore, communication between them, e.g., multi-user flight simulators (see Figure 1) or multi-user telesurgery systems (see Figure 3). However, in practice, communication also occurs for single-user teleoperation interfaces: Figure 2 shows a surgeon using a robot-assisted surgical system and an assistant who communicates with the surgeon by drawing on a touchscreen that displays the endoscopic camera view of the operating field. In this case, only the surgeon uses an immersive teleoperation interface while the assistant uses a non-immersive interface.

In our experience, this kind of communication between immersed users and non-immersed humans in various roles (e.g., co-workers, expert advisors, trainers, apprentices, etc.) is very common and occurs regardless of whether a teleoperation interface is designed to support it or not. In fact, this kind of communication is probably the most common form of communication during the use of teleoperation interfaces.

Some teleoperation interfaces work well with a lower level of immersion, e.g., the interface for re-



Figure 1: Multi-user flight simulator. © NASA.



Figure 2: Telesurgery system for one surgeon (left). An assistant (right) is able to see the endoscopic view and draw lines on a touchscreen to visually communicate with the surgeon. © 2014 Intuitive Surgical, Inc.



Figure 3: Multi-user telesurgery system: both surgeons are able to control the surgical system at the same time and over large distances. © 2014 Intuitive Surgical, Inc.



Figure 4: Multi-user drone control. © Gerald Nino, U.S. Department of Homeland Security.



Figure 5: Single-user drone control. © Senior Airman Elliott Sprehe.



Figure 6: Two spacewalkers. © NASA.

mote drone control in Figure 4 in comparison to the interface in Figure 5. Thus, less immersive interfaces should also be covered by our classification. While we do not focus on collaborative Augmented Reality (AR) systems, our classification extends naturally along a “transport” dimension [BGR<sup>+</sup>98] (see Section 3.2.3) and, therefore, covers many such systems.

Extending our classification to unmediated, real environments can provide useful reference cases for communication without teleoperation interfaces. Particularly interesting examples are spacewalks (see Figure 6) since they involve unmediated visual communication but mediated auditory communication.

## 3.2 Classification of Communication Situations

We classify human-to-human communication during the use of immersive teleoperation interfaces in seven dimensions. The first five (copresence, virtuality, transport, point of view, and asynchronicity) classify the communication situation and are discussed in this section. Section 3.3 discusses the remaining two dimensions (communication channel and mediation of communication). Classifying the communication situation is important since it has a strong effect on the mediation and its technical implementation.

### 3.2.1 Copresence

Aaltonen et al. [ATH<sup>+</sup>09] identified copresence as an important dimension to characterize mediated communication situations. While copresence is usually considered an emerging effect that most collaborative teleoperation interfaces try to achieve, we interpret the broad level of copresence as a design decision. Consider the example of an immersed surgeon communicating with a surgical assistant (Figure 2): while both users see a shared workspace (the operating field), the interface for the assistant is not designed to immerse the assistant, nor is the assistant represented in the shared workspace, nor is the assistant (in the depicted situation) able to manipulate it. Thus, full copresence of the surgeon and assistant was clearly not a design goal for this system.

On the other hand, two pilots in a multi-user flight simulator (Figure 1) will usually experience each other as copresent since they are actually co-located in the same physical environment. Technically, the level of copresence is extremely high in this situation because these users can see, hear, touch, and smell each other

without any mediation, and they are usually both manipulating the controls in their shared, immediate environment, which was designed for two co-located users.

Achieving this level of copresence in a CVE is impossible today and probably for several decades to come due to the required display resolution and frame rate [BJK<sup>+</sup>13]. On the other hand, CVEs usually achieve a higher level of copresence than systems that are not designed to support it, such as the single-user telesurgery system in Figure 2.

Based on these examples, we distinguish three broad levels of copresence (see also the horizontal axes in Figures 7 and 8). Analogously to Lombard and Ditton [LD97], we provide short verbal descriptions of the situations in quotation marks (but from the point of view of the immersed user):

- copresence is not a goal: single-user teleoperation by an immersed user communicating with a non-immersed user — “Only I am there.”
- mediated copresence: collaborative teleoperation by two connected, immersed users — “We are both there.”
- unmediated copresence: joint teleoperation by two co-located, immersed users — “We are there together.”

We do not provide precise definitions of these cases; instead they should be considered exemplary cases on the gradual dimension of copresence. It should also be noted that the proposed classification only describes pairwise relations between communicating users: if more than two users are involved, multiple communication relations in different categories can occur at the same time.

While the third case of unmediated copresence includes collaboration, we label it as “joint teleoperation” to distinguish it from the second case. Also note that the co-location of the third case is a necessary but not sufficient requirement for unmediated communication: co-located users can use mediated communication in one or more channels, for example, in order to increase the realism of a simulation; e.g., the mediated auditory communication in Figure 1. Furthermore, users might also be co-located in the first two cases and then use unmediated communication in some channels (in particular the auditory channel). Note that a mixture of mediated and unmediated communication channels is not specific to teleoperation interfaces: the co-located spacewalkers in Figure 6 re-

quire mediation of auditory communication and they would require mediation of visual communication to see each other’s facial expressions or gaze direction. Due to the required mediation, we consider the situation of the spacewalkers in Figure 6 an example of mediated copresence. The actual level of copresence depends on many factors [Sch02]. Nonetheless, the presented broad levels of copresence appear to be crucial when designing support for communication in teleoperation interfaces as discussed in more detail in Section 4.

### 3.2.2 Virtuality

Our classification includes the dimension of virtuality which was proposed by Milgram et al. [MTUK95]. Extreme cases are completely virtual environments (e.g., in flight simulators) and completely real (but mediated) environments (e.g., for telesurgery). This corresponds to the artificiality dimension proposed by Benford et al. [BGR<sup>+</sup>98].

In many cases, the technology for mediation of environments also allows users to record these environments. These recordings can be used, for example, to review teleoperation sessions for training or to test trainees by pausing the recording and asking trainees about the appropriate next step in the teleoperation session. Furthermore, advances in the technologies for capturing photo spheres have led to an increasingly popular use of immersive teleoperation interfaces to experience telepresence in recorded environments without the possibility to manipulate these environments. For these reasons, we include mediated, recorded environments in the virtuality dimension. Recorded environments can be thought of as remote environments with extremely large delay or — alternatively — as virtual environments with a specific form of image-based rendering. Thus, we place recorded environments between remote and virtual environments.

Including unmediated, real environments in the classification provides additional reference cases that can lead to a deeper understanding of specific communication situations. Therefore, we extend our classification by further distinguishing between mediated, remote environments (e.g., in the case of telesurgery) and unmediated, real environments (e.g., in the case of spacewalks), where the latter is considered to be less virtual than the former. The rationale is that the immersive mediation of a real environment limits the

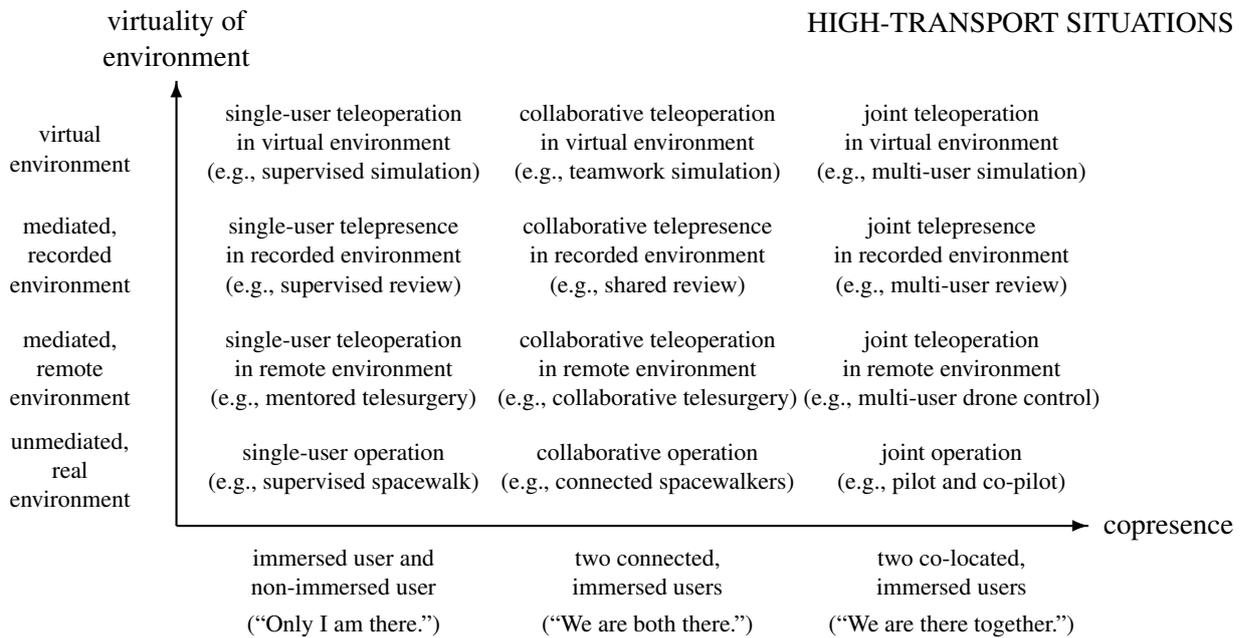


Figure 7: The dimensions of copresence (horizontal) and virtuality (vertical) of our classification for high-transport situations. The row labeled “unmediated, real environment” is an extension for real environments.

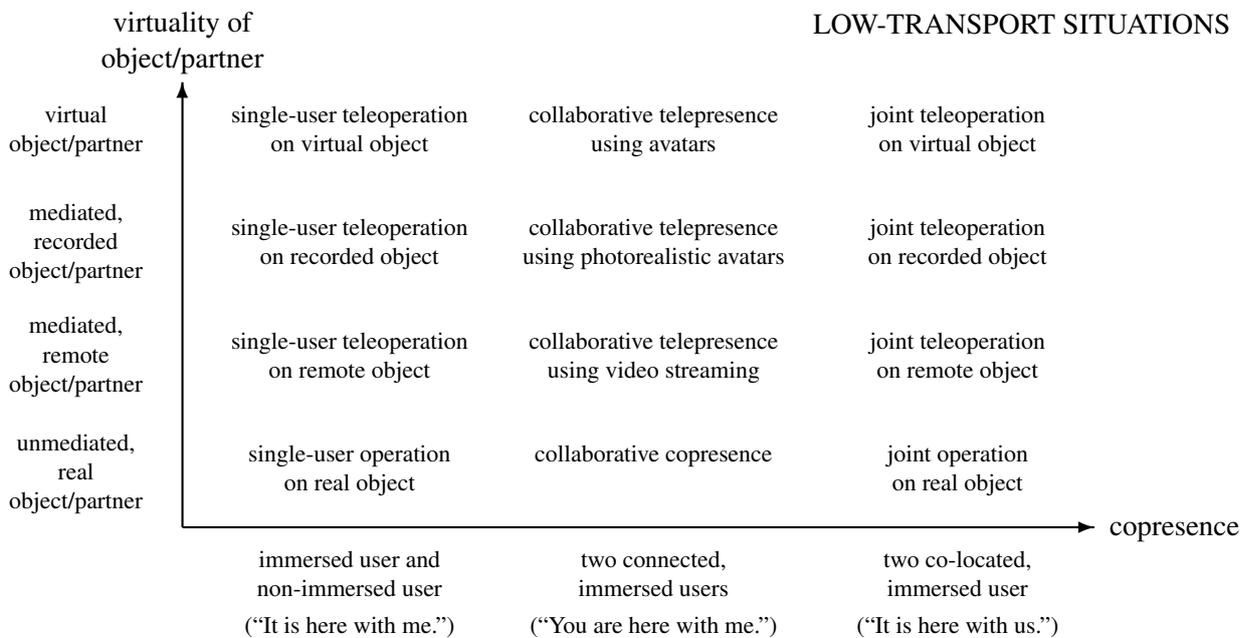


Figure 8: The dimensions of copresence (horizontal) and virtuality (vertical) of our classification for low-transport situations. The row labeled “unmediated, real object/partner” is an extension for real objects/partners.

ways in which it can be experienced and in this sense makes it more similar to the immersion in a virtual environment.

Figure 7 shows the proposed classification of communication situations according to the dimensions of copresence and virtuality — including the extension for unmediated, real environments.

### 3.2.3 Transport

The transport dimension is based on the work by Benford et al. [BGR<sup>+</sup>98]. It is related to the extent-of-presence-metaphor by Milgram et al. [MTUK95] and the conceptualization of presence as transportation by Lombard and Ditton [LD97]. Specifically, we distinguish between high-transport situations in a virtual or remote environment as discussed so far (see Figure 7) and low-transport situations, where virtual or remote objects or people appear in the immediate environment (see Figure 8). Examples for the latter are teleconference systems and collaborative augmented reality systems.

For low-transport communication situations, we distinguish the same three broad levels of copresence as for high-transport situations. However, their descriptions have to be adapted:

- copresence is not a goal: single-user (tele)operation on a virtual or remote object by an immersed user communicating with a non-immersed user — “It is here with me”
- mediated copresence: collaborative telepresence of two connected, immersed users — “You are here with me.”
- unmediated copresence: joint (tele)operation on a virtual or remote object by two co-located, immersed users — “It is here with us.”

### 3.2.4 Point of View

Otto et al. [ORW06] distinguish between CVEs that use a “look-into” metaphor and, therefore, a third-person view, and CVEs that use a “step-into” metaphor and, therefore, the first-person view of the avatar or robot that the user is controlling. Instead of using these metaphors, we prefer to classify systems by the employed point of view because there is an increasing number of immersive systems — in particular VR games for head-mounted displays — that mix the two metaphors by letting players “step into” the virtual

world and “look into” the world including their avatar using a third-person view. In these cases, it is easier to determine the point of view of the camera: either a first-person view or a third-person view. We propose the following descriptions of these two cases:

- first-person view of avatar or robot — “My avatar represents my body.”
- third-person view onto avatar or robot — “My avatar represents me.”

While the point of view of the camera is easily determined, the effect of different points of view on users depends on many factors, for example, the level of the user’s engagement and the specific movements that the avatar is capable of. In some cases, the difference between a first-person view and a third-person view might be very subtle for a specific user — similarly to the difference between “my body” and “me” in our descriptions.

Figure 9 illustrates the point-of-view dimension for interfaces for single users in high-transport situations. In the case of the unmediated environment, the first-person view onto oneself is usually the most common way of observing one’s own interaction with the environment. However, sometimes a third-person view is used, for example, when operating a nearby robot. In other cases, the third-person view arises from the use of optical devices, e.g., mirrors. Operating on an object that is looked at in a mirror is a very common scenario for anyone who is used to “operating” on his or her own face while looking at it in a mirror. The fact that this often is possible without any conscious effort might help to make it plausible that it can be relatively easy for many users to identify with an avatar that is displayed from a third-person point of view.

It should be noted that the third-person view requires avatars for co-located users, who might otherwise have no need for avatars. Thus, co-located users might see their avatars in exactly the same way that connected users see their avatars, and co-location of the users becomes the crucial — and possibly only — difference between the two situations.

### 3.2.5 Asynchronicity

Asynchronicity of communication is a common dimension for the classification of CSCW systems in general [BGBG95] and, therefore, also of CVE systems. In the case of virtual environments that are not specifically designed for collaboration, asynchronous

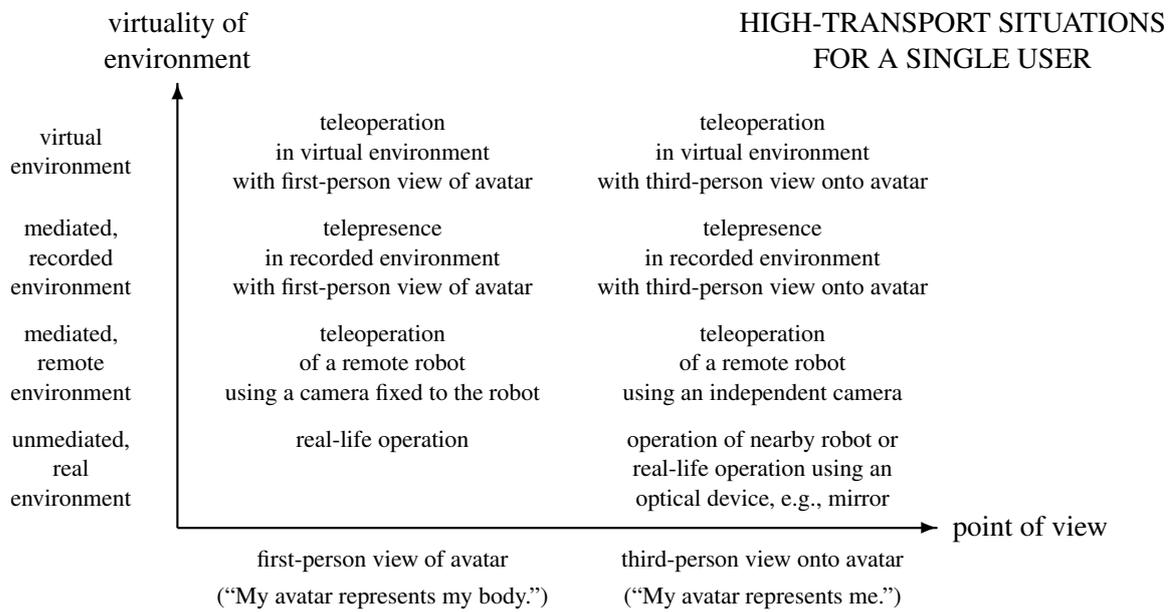


Figure 9: The dimensions of point of view (horizontal) and virtuality (vertical) of our classification for high-transport situations for a single user. The row labeled “unmediated, real environment” is an extension for real environments.

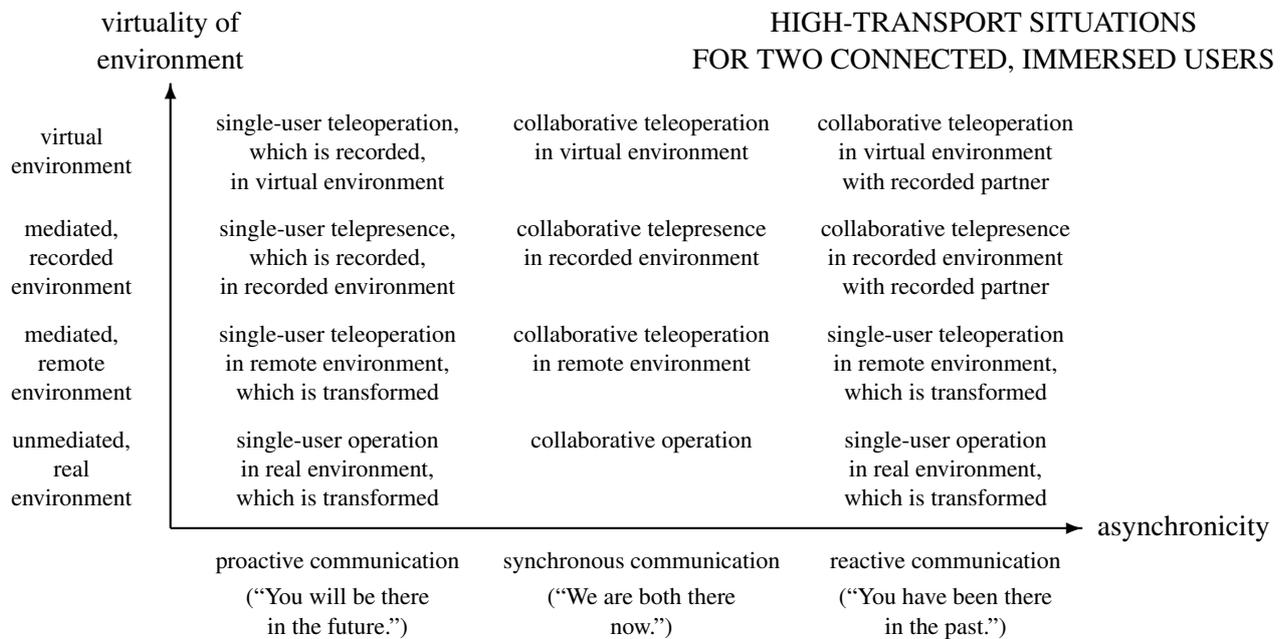


Figure 10: The dimensions of asynchronicity (horizontal) and virtuality (vertical) of our classification for high-transport situations for two connected, immersed users. The row labeled “unmediated, real environment” is an extension for real environments.

communication is often made possible in a limited form by persistent changes to the virtual environment. Richer forms of asynchronous communication are made possible by recording and replaying actions of users. While asynchronous communication is important in special scenarios that make synchronous communication impossible [WMW<sup>+</sup>15], it is also very important for social media in virtual reality.

As shown in Figure 10, we distinguish between three cases of asynchronicity:

- proactive communication — “You will be there in the future.”
- synchronous communication — “We are both there now.”
- reactive communication — “You have been there in the past.”

Since co-location implies synchronous communication, asynchronous communication is only relevant for connected, immersed users and immersed users communicating with non-immersed users.

### 3.3 Classification of the Communication Processes

The proposed classification is intended to inform the design of support for human-to-human communication, which often includes a form of mediation. Therefore, we classify the communication process by the level of mediation. Since the mediation usually depends strongly on the communication channel, the latter is another dimension of our classification.

#### 3.3.1 Communication Channel

We distinguish the following communication channels:

- auditory without using media, e.g., speech or nonverbal utterances
- visual without using media, e.g., facial expressions, gaze direction, hand gestures, or full-body gestures
- using media, e.g., using written text, using visuals, using audiovisual recordings
- others, e.g., haptic, olfactory, etc.

The use of media for communication in shared workspaces includes drawing and writing, in particular, writing lists [Tan91]. The auditory and visual communication without media is considered different from the use of media since the latter is always a form of mediated communication, which usually requires input and/or output devices, while the former can also occur without mediation.

#### 3.3.2 Level of Mediation of Communication

Analogously to the broad level of copresence, we distinguish between three broad levels of mediation:

- explicit mediation, i.e., in general, users are fully aware of the mediation of the communication
- transparent mediation, i.e., users are not (or less) aware of the mediation of the communication
- no mediation, i.e., unmediated communication

Note that the mediation of the communication is different from the mediation of the environment or an object. However, all forms of mediation are likely to influence the resulting level of copresence.

## 4 Supporting Communication

To demonstrate the usefulness of the proposed classification for the design of support for human-to-human communication in immersive teleoperation interfaces, this section sketches how the dimensions of the classification and the provided examples can guide and inspire the design of support for communication in such interfaces.

The first question might be whether a single-user, immersive teleoperation interface needs to support human-to-human communication at all. In our experience, the communication between an immersed user and a non-immersed person (as depicted in Figure 2) is not only useful but often also critical for training, supervision, expert advice, etc. Even consumer head-mounted displays benefit from features such as a (video) see-through function, which supports visual communication with a co-located, non-immersed person. It is also important to realize that even without support for communication, many users will nonetheless try to communicate with non-immersed persons, which is likely to lead to a frustrating user experience if there is no support for it.

When assessing the need for support of communication, non-verbal communication is easily overlooked — in particular if it is used to establish awareness, which is important in shared workspaces [GG02]. It should also be noted that awareness cues are not limited to visual communication channels but are also common in auditory communication; e.g., in multiplayer games [CCS12].

#### 4.1 Choosing a Level of Virtuality

In most cases, the level of virtuality cannot be chosen to support human-to-human communication in the best way possible since it is determined by other constraints. There are, however, some exceptions, in particular, regarding the virtuality of the communication signal. We provide two examples, which illustrate that less virtuality usually allows for communication of better quality and more expressiveness.

**Example 1:** Pointing in a small workspace by a remote user can be supported by recording the user's hand and overlaying the view of the workspace for another user with the recorded hand [SDS<sup>+</sup>11]. Alternatively, a virtual hand or even a line drawing can be employed (as in the system for surgical assistants depicted in Figure 2). In terms of the quality of communication, the recording of a user's hand allows the user to employ the full expressive power of natural hand gestures while the expressiveness of a virtual hand or a line drawing is in most cases significantly lower.

**Example 2:** In some systems, users can specify facial expressions, which are then applied to the faces of avatars [GVCP<sup>+</sup>99]. Alternatively, a webcam can record a user's face, which is then displayed to other users. It should be noted that a similar display would be necessary for spacewalkers to see each other's facial expressions. Thus, it is not always clear which approach is closer to reality and, therefore, more immersive. The quality of facial expressions of avatars does not only suffer from limitations of the specification of the expressions but also from limitations of their rendering. Thus, a display of recorded facial expressions usually provides a considerably higher quality of communication.

#### 4.2 Choosing a Level of Transportation

Similarly to the case virtuality, it is usually not possible to choose the level of transport to support human-

to-human communication in the best way possible. If it is feasible, the decision affects large parts of the design of the teleoperation interface and the communication between co-located users as illustrated by three examples:

**Example 1:** The teleoperation interface for the surgeon in Figure 2 is designed to immerse surgeons (and to let them rest their forehead and forearms on the console); however, this interface isolates the surgeon — at least visually — from the rest of the surgical team and, therefore, requires support for mediated communication. In traditional laparoscopic surgery, surgeons would often stand upright and watch the endoscopic camera view on a screen — similarly to the surgical assistant in Figure 2. In this case, the level of transport is lower but the interface would allow for easier, unmediated visual communication between the surgeon and the rest of the team.

**Example 2:** Augmented reality glasses and see-through head-mounted displays have become mass-market products. Thus, they offer very affordable platforms to implement low-transport collaborative systems that support unmediated visual communication between co-located users at a level that today cannot be achieved in VR environments.

**Example 3:** A virtual window that provides two different views for two co-located users depending on their positions can be implemented with mass-market 3D TVs using passive stereo glasses and mass-market body trackers. This allows for joint operation on virtual objects with easier, unmediated communication than in the case of VR environments.

#### 4.3 Choosing a Broad Level of Copresence

Often, the broad level of copresence is determined by other considerations than human-to-human communication. If there is a choice, however, it should be noticed that advances in mass-market products (augmented reality glasses, see-through head-mounted displays, 3D TVs with passive stereo glasses, etc.) have made it much more affordable to immerse co-located users while supporting unmediated communication (see Examples 2 and 3 in the previous section). This is important because co-located users can experience stronger copresence at a fraction of the costs for connected teleoperation interfaces. For the other extreme

of very low copresence, we stress again that communication between an immersed user and a non-immersed user is often unavoidable. Therefore, users are likely to appreciate support for this kind of communication. Additionally, it might be worth considering whether all users have to be immersed. Not immersing certain users removes many constraints on the design of their user interfaces and might not hinder communication if sufficient support for this communication situation is provided.

While we believe that the proposed broad levels of copresence are relevant concepts, there are no perfectly clear boundaries between them. For example, the co-located users in Figure 1 use unmediated communication in all but the auditory channel. The following example shows the reverse case: mediated (or no) communication between co-located users in all but the auditory channel. Therefore, we consider this example closer to the case of two connected users even though the users clearly benefit from the unmediated communication in the auditory channel.

**Example 1:** Imagine two, co-located users wearing head-mounted displays whose body movements are motion tracked and translated one-to-one into a shared virtual space. Furthermore, assume that these users have no other way of changing their relative positions and orientations to each other than physical body movements such that their virtual and physical relative positions and orientations are always consistent. In this case, unmediated auditory communication can provide them accurate three-dimensional information about the position of each other. Achieving this effect for mediated auditory communication using 3D audio techniques requires considerable effort and computational costs.

The next example shows that the distinction between an immersed user and a non-immersed user is not always completely clear. That is, even though users are not fully immersed, they still can perform some functions that are usually only available to immersed users.

**Example 2:** Imagine that the surgical assistant in Figure 2 is equipped with two controllers that simulate the controllers of the surgical system. Furthermore, assume that the assistant can control a virtual robotic instrument that is shown to the immersed surgeon. From the point of view of the immersed surgeon, the assis-

tant can play a role that is very close to the role of another immersed surgeon who controls one of the actual robotic instruments. This kind of visual communication was proposed for the training of robotic surgery as a cost-efficient alternative to the multi-user telesurgery system depicted in Figure 3 [MTG<sup>+</sup>14].

#### 4.4 Choosing a Point of View

There are many factors that influence the choice of a first-person view or a third-person view. However, since there are basically only these two options, it is usually worthwhile to consider both alternatives. For example, the use of a third-person view for low-transport situations might be counterintuitive, but “augmented mirrors” [UTNH02] provide a compelling example that there are useful applications of third-person views even in low-transport situations.

With respect to support for communication, a first-person view is likely to improve the awareness of non-verbal communication by avatars of other users, e.g., their gaze direction, while a third-person view is likely to improve the awareness of and control over the non-verbal communication of one’s own avatar as well as the awareness of nearby avatars and objects that would be outside the field-of-view of a first-person view.

While third-person views are quite common in screen-based interfaces to (collaborative) virtual environments, it remains to be seen whether they will achieve a similar popularity in interfaces based on head-mounted displays.

#### 4.5 Choosing between Synchronous and Asynchronous Communication

Many application scenarios require support for synchronous communication while not excluding the possibility of support for asynchronous communication. Thus, the choice in these cases is not between supporting either synchronous or asynchronous communication but whether or not to support asynchronous communication. In this context, it is important to be aware of the many forms of asynchronous communication. These include, among others:

- changes to the environment
- recording and replaying of screen captures
- recording and replaying of (tele)operations
- text, audio, and video messages

- messages using other media

Whether or not to support these features has to be decided for each specific form of asynchronous communication and the specific application.

#### 4.6 Choosing a Communication Channel

Ideally, the specific application should determine the supported communication channels; however, there is usually a cost associated with the support of each channel. Fortunately, humans are often quite flexible and effortlessly switch between different channels of communication if specific communication channels are unavailable or ineffective. For example, one might try to get someone's attention by calling, then — if unsuccessful — by waving, and finally by tapping on the other person's shoulder. While this might suggest that one channel of communication could be sufficient, it should be noted that natural human-to-human communication often combines and relies on multiple channels, for example, by referencing visual gestures in auditory communication: "Take the one over there.", "Do it like this.", etc. Supporting multiple channels of communication is therefore often beneficial or even necessary for efficient communication.

In practice, most auditory communication (possibly including 3D information and multiple audio tracks) can be classified as one channel except communication using audio media, which is different since it requires additional support for selecting and playing audio data. Visual communication requires more channels since facial expressions, gaze direction, hand gestures, full-body gestures, and the use of visual media often have to be supported separately. Haptic communication and other forms of communication usually are separated into at least as many channels as there are output devices available to support them.

#### 4.7 Choosing a Level of Mediation

For a given communication situation and a given communication channel, the level of mediation depends on cost considerations, the targeted level of copresence, consistency with the rest of the experience, etc.

In general, unmediated communication provides the best quality and the highest expressiveness at the lowest costs. Thus, it also provides the strongest level of copresence. However, unmediated communication is impossible in many applications. Limitations of unmediated communication that are specific to commu-

nication channels are discussed in the following sections.

Mediated communication usually provides lower quality and often less expressiveness. Thus, the level of copresence is reduced compared to unmediated communication. Increasing the quality of the mediation to the point that users are no longer aware of it (i.e., that the mediation becomes transparent) often requires excessively high costs. Therefore, support for high copresence is often associated with high costs, even though the best copresence is usually achieved with unmediated communication at low costs.

Transparent mediation, however, is often unnecessary. In fact, explicit mediation of communication can often provide better quality of communication at the same costs and, therefore, improve collaboration or task performance more than a stronger level of copresence would do. Thus, unless copresence is a goal in itself (which it usually is not in real workplaces), the possibility of explicit mediation should not be ruled out.

##### 4.7.1 Mediating Auditory Communication

As mentioned, unmediated auditory communication usually provides the best quality including information about the relative 3D position of the communication partner. However, this 3D information might be inconsistent with the relative position in a shared virtual environment. To keep the information consistent, the relative movement in the virtual environment has to be constrained as mentioned in the first example of Section 4.3.

As illustrated by Figure 1, auditory communication is sometimes mediated even though unmediated communication is possible. In other cases, unmediated communication is not possible even in real environments as illustrated by the spacewalk example in Figure 6. For mediated auditory communication, sound quality, latency, and the communication of 3D positional information are often limiting factors.

Adapting the auditory communication to the shared environment (e.g., by adding sound reflections from virtual walls) can improve the level of presence and, therefore, the level of copresence. While this adaptation to the shared environment can be costly, it is often easier in a virtual environment since more information about the environment is known to the system in this case.

#### 4.7.2 Mediating Visual Communication

Unmediated visual communication provides a higher quality than mediated visual communication in terms of latency, resolution, field of view, dynamic range of colors, depth perception (in particular due to optical accommodation), vergence (in particular consistent vergence and accommodation), frame rate, etc. Achieving transparent mediation is therefore practically impossible in most cases and attempting to achieve it can lead to excessive costs [BJK<sup>+</sup>13].

However, this limited quality is usually more important for mediation of the environment than for mediation of communication signals. Moreover, unmediated communication can suffer from other limitations, for example, facial expressions are easily missed when the other user focuses on a shared workspace. Another example is the limited precision of pointing at a distance with hand gestures or gaze direction.

Mediation of visual communication can take very different forms. Video streams of the face, body, or hands are just one possibility and often lead to quite explicit mediation, which sometimes reflects mediation in real situations, e.g., spacewalkers could use video streams of each other's face to communicate with facial expressions. Other forms of mediation often rely on tracking and recognition techniques (e.g., gaze direction, hand and full-body motion, etc.), in particular, when mediating through a virtual avatar.

In analogy to the mediation of auditory communication, the mediation of visual communication is sometimes easier in virtual environments since more information about those environments is known to the system. For example, pointing with a virtual laser pointer in a virtual environment is straightforward while augmenting a video stream of a remote environment with the illumination by a virtual laser pointer is considerably more difficult.

#### 4.7.3 Mediating Communication Using Media

This channel is special since unmediated communication is no option as the communication is already mediated by the use of a medium. However, the mediation can be more or less consistent with the virtual environment. For example, a sheet of paper could be represented as a textured 3D polygon in a virtual environment. An alternative would be to overlay the view of the virtual environment with a 2D representation of the contents of the paper, which is less consistent with the 3D virtual environment but is also likely to

provide better readability without requiring potentially complex handling of a 3D representation of the paper.

#### 4.7.4 Mediating Other Communication Channels

For other communication channels, e.g., haptic or olfactory, unmediated communication is usually the best option if available. Mediated communication in other channels than auditory and visual tends to be even more restricted and more expensive than the mediation of audiovisual communication. Therefore, a useful approach is often to transform the communication from its original channel to the auditory or visual channel, i.e., to use sensors (e.g., touch sensors) to record the communication and then communicating the recorded information as audio or visuals.

## 5 Conclusion

The proposed classification of human-to-human communication during the use of immersive teleoperation interfaces is based on examples of real-life usage of teleoperation interfaces. These examples suggest that the broad level of the experienced copresence mainly depends on the relation of the communicating users with respect to immersion and co-location. We have identified three main cases (immersed user and non-immersed user; connected, immersed users; and co-located, immersed users) and showed that these cases are relevant for different levels of virtuality, transport, asynchronicity, and different points of view. We conclude that the broad level of copresence is a suitable dimension for the proposed classification, which is one of the main contributions of this work along with the actual classification.

We demonstrate the usefulness of this classification for designing support for human-to-human communication by discussing design decisions in terms of the dimensions of the proposed classification. This provides a structured way of identifying challenges as well as alternatives and reference cases.

## 6 Future Work

In this work, we focus on communication between users, which often is a prerequisite for collaboration. Shifting the focus to collaboration is, therefore, a natural next step.

While we are mainly concerned with professional applications of teleoperation in this work, we acknowl-

edge that these interfaces could also be used for entertainment. Applying the proposed classification in this context is another avenue for future work.

One part of the proposed classification is the classification of the communication situation. Whether this part is a useful classification of teleoperation scenarios in its own right, is yet another interesting question that has to be left for future work.

## References

- [ATH<sup>+</sup>09] Viljakaisa Aaltonen, Jari Takatalo, Jukka Hakkinen, Miikka Lehtonen, Gote Nyman, and Martin Schrader, *Measuring Mediated Communication Experience*, Proceedings of the International Workshop on Quality of Multimedia Experience (QoMEX 2009), 2009, DOI 10.1109/QOMEX.2009.5246967, pp. 104–109, ISBN 978-1-4244-4370-3.
- [BGBG95] Ronald M. Baecker, Jonathan Grudin, William A. S. Buxton, and Saul Greenberg, *Readings in human-computer interaction: Toward the year 2000*, 2nd ed., ch. Groupware and Computer-Supported Cooperative Work, pp. 741–782, Morgan Kaufmann Publishers, San Francisco, CA, USA, 1995.
- [BGR<sup>+</sup>98] Steve Benford, Chris Greenhalgh, Gail Reynard, Chris Brown, and Boriana Koleva, *Understanding and Constructing Shared Spaces with Mixed-Reality Boundaries*, ACM Transactions on Computer-Human Interaction (TOCHI) **5** (1998), no. 3, 185–223, ISSN 1073-0516, DOI 10.1145/292834.292836.
- [BGRP01] Steve Benford, Chris Greenhalgh, Tom Rodden, and James Pycock, *Collaborative Virtual Environments*, Communications of the ACM **44** (2001), no. 7, 79–85, ISSN 0001-0782, DOI 10.1145/379300.379322.
- [BJK<sup>+</sup>13] Mathias Borg, Stine Schmiege Johansen, Kim Srirat Krog, Dennis Lundgaard Thomsen, and Martin Kraus, *Using a Graphics Turing Test to Evaluate the Effect of Frame Rate and Motion Blur on Telepresence of Animated Objects*, Proceedings of the International Conference on Computer Graphics Theory and Applications and International Conference on Information Visualization Theory and Applications (Portugal) (S. Coquillart, C. Andujar, R. S. Laramée, A. Kerren, and J. Braz, eds.), Institute for Systems and Technologies of Information, Control and Communication, 2013, pp. 283–287, ISBN 978-989-8565-46-4.
- [CCS12] Victor Cheung, Y.-L. Betty Chang, and Stacey D. Scott, *Communication Channels and Awareness Cues in Collocated Collaborative Time-Critical Gaming*, Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12) (New York, NY, USA), ACM Press, 2012, DOI 10.1145/2145204.2145291, pp. 569–578, ISBN 978-1-4503-1086-4.
- [GG02] Carl Gutwin and Saul Greenberg, *A Descriptive Framework of Workspace Awareness for Real-Time Groupware*, Computer Supported Cooperative Work **11** (2002), no. 3-4, 411–446, ISSN 0925-9724, DOI 10.1023/A:1021271517844.
- [Gre96] Saul Greenberg, *Peepholes: Low Cost Awareness of One's Community*, Conference Companion on Human Factors in Computing Systems (CHI '96) (New York, NY, USA) (Michael J. Tauber, ed.), ACM Press, 1996, DOI 10.1145/257089.257283, pp. 206–207, ISBN 0-89791-832-0.
- [GVCP<sup>+</sup>99] A. Guye-Vuillème, T. K. Capin, I. S. Pandzic, N. Magnenat-Thalmann, and D. Thalmann, *Nonverbal Communication Interface for Collaborative Virtual Environments*, Virtual Reality **4** (1999), no. 1, 49–59, ISSN 1359-4338, DOI 10.1007/BF01434994.
- [LD97] Matthew Lombard and Theresa Ditton, *At the Heart of It All: The Concept of Presence*, Journal of Computer-Mediated Communication

- 3 (1997), no. 2, ISSN 1083-6101, DOI 10.1111/j.1083-6101.1997.tb00072.x.
- [Min80] Marvin Minsky, *Telepresence*, *Omni* 2 (1980), no. 9, 45–51.
- [MTG<sup>+</sup>14] Florin Octavian Matu, Mikkel Thøgersen, Bo Galsgaard, Martin Møller Jensen, and Martin Kraus, *Stereoscopic Augmented Reality System for Supervised Training on Minimal Invasive Surgery Robots*, Proceedings of the Virtual Reality International Conference: Laval Virtual (VRIC '14) (New York, NY, USA), ACM Press, 2014, article no. 33, ISBN 978-1-4503-2626-1.
- [MTUK95] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino, *Augmented Reality: A Class of Displays on the Reality-Virtuality Continuum*, Proceedings SPIE, Vol. 2351: Telem manipulator and Telepresence Technologies, 1995, DOI 10.1117/12.197321, pp. 282–292.
- [ORW06] Oliver Otto, Dave Roberts, and Robin Wolff, *A Review on Effective Closely-Coupled Collaboration Using Immersive CVE's*, Proceedings of the 2006 ACM International Conference on Virtual Reality Continuum and its Applications (VRCAI '06) (New York, NY, USA), ACM Press, 2006, DOI 10.1145/1128923.1128947, pp. 145–154, ISBN 1-59593-324-7.
- [Sch02] Ralph Schroeder, *Social Interaction in Virtual Environments: Key Issues, Common Themes, and a Framework for Research*, *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments* (Ralph Schroeder, ed.), Springer, London, UK, 2002, DOI 10.1007/978-1-4471-0277-9\_1, pp. 1–18, ISBN 9781852334611.
- [SDS<sup>+</sup>11] Mahesh B. Shenai, Marcus Dillavou, Corey Shum, Douglas Ross, Richard S. Tubbs, Alan Shih, and Barton L. Guthrie, *Virtual interactive presence and augmented reality (VIPAR) for remote surgical assistance*, *Operative Neurosurgery* 68 (2011), no. Suppl. 1, 200–207, ISSN 2332-4252, DOI 10.1227/NEU.0b013e3182077efd.
- [Tan91] John C. Tang, *Findings from observational studies of collaborative work*, *International Journal of Man-Machine Studies* 34 (1991), no. 2, 143–160, ISSN 0020-7373, DOI 10.1016/0020-7373(91)90039-A.
- [UTNH02] Keita Ushida, Yu Tanaka, Takeshi Naemura, and Hiroshi Harashima, *i-mirror: An Interaction/Information Environment Based on a Mirror Metaphor Aiming to Install into Our Life Space*, Proceedings of the 12th International Conference on Artificial Reality and Telexistence (ICAT2002) (Tokyo, Japan), The Virtual Reality Society of Japan, 2002.
- [WMW<sup>+</sup>15] Peggy Wu, Jacquelyn Morie, Peter Wall, Eric Chance, Kip Haynes, Jack Ladwig, Bryan Bell, Tammy Ott, and Christopher Miller, *Maintaining Psycho-Social Health on the Way to Mars and Back*, Proceedings of the Virtual Reality International Conference (VRIC '15) (New York, NY, USA), ACM Press, 2015, Article No. 3, ISBN 978-1-4503-3313-9.

Citation
Martin Kraus and Martin Kibsgaard, <i>A Classification of Human-to-Human Communication during the Use of Immersive Teleoperation Interfaces</i> , <i>Journal of Virtual Reality and Broadcasting</i> , 14(2017), no. 1, August 2017, urn:nbn:de:0009-6-45476, DOI 10.20385/1860-2037/14.2017.1, ISSN 1860-2037.