How Social Cues shape Remote Collaborative Work Processes: Establishing an Abstraction Hierarchy for Evaluating Computer-supported Collaborative Work Tools

Maikel ter Riet

Faculty of Behavioral, Management and Social Sciences, University of Twente

Master Thesis

First supervisor: Prof. dr. J.M.C. Schraagen

Second supervisor: dr. S. Borsci

External supervisor: T. Hueting, MSc

15th of April 2022

Abstract

How do social cues influence the work processes of remote collaboration? By conducting a systematic literature review, this study tries to establish the fundamental relations between social cues and the work processes in remote collaboration. Additional literature was examined to establish a preliminary abstraction hierarchy describing teamwork and technologies in remote teams. Based on evidence of relationships between social cues and work processes found in the literature review, the preliminary abstraction hierarchy was modified to visualize which social cues facilitate certain remote collaborative work processes. On the basis of the established abstraction hierarchy, this study provides an ‘in-principle’ evaluation of ‘VRComm', an end-to-end web system that enables remote conferencing in virtual reality to test whether the abstraction hierarchy provides a framework for identifying the strengths and points for improvement for computer-supported remote collaborative work tools.

According to the evidence found in the literature review, shared gaze and gesture cues mostly facilitate the processes of conversational grounding and co-presence awareness. Furthermore, gaze cues also play a role in distributed performance monitoring. Overall, gaze and gesture cues facilitate the work processes in similar ways. Namely, to re-direct a co-worker’s attention and enable the use of deictic references to make communication more efficient. Furthermore, gesture cues enable co-workers to perceive actions in a direction relative to target objects. Moreover, non-oriented action gestures cues like head-nodding enable co-workers to perceive that their partner is understanding the dialogue. This allows collaborators to better express themselves and, consequently, being able to understand each other better. Moreover, shared gaze and gesture cues also benefit the feeling of togetherness, as these cues display an interaction. These findings are similar to the findings of the ‘in-principle’ evaluation of VRComm. Gesture and gaze directions cues benefited the processes of conversational grounding and co-presence awareness. Furthermore, the projected user’s representations in virtual reality (VR) allowed collaborators to intervene when incorrect actions were being performed. Moreover, these representations allowed to perceive each other’s posture or user’s intensity in performing actions. Therefore, the projected user’s representations promoted the processes of distributed performance- and workload monitoring. Overall, the abstraction hierarchy demonstrated to be a useful framework for identifying the strengths and points of improvement of VRComm. By using this framework,

recommendations could be made to develop VRComm for broader remote collaboration purposes.

Contents

1. Introduction

1.1. Background

Over the past decades, several computer-supported collaborative work (CSCW) tools have been developed for supporting collaborative work. These technologies provide tools aimed to realize remote collaboration with an experience similar to co-located collaboration. For instance, co-located collaborators are able to perceive where their partner is looking at. In the field of CSCW, eye-tracking technologies have been used to visualize the gaze of remote collaborators in a shared workspace (Higuch, Yonetani & Sato, 2016). Moreover, co-located collaborators can use pointing gestures to refer to task objects and locations. In the field of CSCW, visual augmentation techniques have been used to provide users with a shared virtual space that includes embodiments of remote gestures (Yamashita, Kaji, Kuzuoka & Hirata, 2011). Nowadays, a wide range of these technologies is used to allow distributed work teams to engage in real-time collaboration including video conferencing, audio conferencing, application sharing, chatting and sharing visual media spaces in 2D and 3D format (Bradner & Mark, 2002).

In a co-located collaborative setting, individuals can pay attention to the social cues of others. Social cues are implicit behaviours that convey social and contextual information by verbal or non-verbal signals. For example, facial expressions, gestures, posture, and eye gaze. These signals are required to predict the intentions of partners and to modify one’s own behaviour as a response to that prediction (Adams, Albohn & Kveraga, 2017; Sauppe & Mutlu, 2014). Therefore, social cues facilitate collaboration, as successful collaboration requires effective communication and coordination of actions and intentions. For example, eye gaze is used to determine where someone is looking at in the environment. When you see a person looking at you, you may conclude that his or her thoughts and intentions are directed at you (Adams, Albohn & Kveraga, 2017). However, this does not always have to be the case, as the person could also be day dreaming or thinking about something else while looking at you.

Moreover, gestures can be used for enhancing or clarifying verbalizations (Fussell, Setlock, Yang, Ou, Mauer & Kramer, 2004). For example, pointing gestures are used to refer to task objects and locations while representation gestures, like hand shapes and movements can be used to represent the nature of actions of those task objects.

While remote collaboration draws on many of the same social cues as face-to-face collaboration, the technologies that mediate such collaboration do not always provide the necessary mediums to perceive these cues (Sauppe & Mutlu, 2014). For example, audio conferencing shows limitations in providing social cues, because people cannot see each other. Therefore, interactions and social awareness are limited to what can be communicated verbally. Non-verbal cues like gestures, eye gaze, and proxemics are not transmitted in audio conferencing (Wolf, Steed & Otto, 2007).

The term ‘proxemics’ describes the interpersonal distances individuals maintain between each other. The maintained distance cues convey different social and cultural meanings, and interpersonal relationships, which form a set of implicit rules of maintaining distance (Llobera, Spanlang, Ruffini & Slater, 2010; Williamson et al., 2021). While video conferencing can provide social cues like gestures and facial expressions, it shows limitations in providing proxemics, because the collaborators are spatially separated from each other.

As mentioned, there is a myriad of research into CSCW tools that have been developed to enable the perception of social cues for supporting remote collaboration. However, in previous research there is a variety in the implementation of social cues in CSCW tools and which collaborative tasks were performed with these tools. Therefore, on the basis of only one study, no generalized statements can be made to what extent social cues support remote collaboration, since both the characteristics of the collaboration task and the CSCW tools could influence the quality of remote collaboration. For this reason, this study aims to combine all the findings of the current research of CSCW tools that enable the perception of social cues. Additionally, supplementary literature will be examined to identify the general processes, values, and functions of remote collaboration. Furthermore, an abstraction hierarchy is applied to provide a visualization how these concepts are related to one other, and which collaborative work processes are proven to be promoted by certain social cues

An abstraction hierarchy is a widely used representation framework proposed by Rasmussen (1986) for describing complex work environments (Lind, 2003). It follows a multileveled representation that conceptualizes a system for each level in a set of attributes. The higher levels of an abstraction hierarchy represent the system in terms of functional purposes, values and priority measures. While the lower levels represent the system in terms of purpose-related functions, object-related processes, and physical objects (Lind, 2003; Reising, 2000). The functional purposes are the main purposes of the system as a whole. In other words, the reasons why the system has been designed. Moreover, the level of the functional purposes relates to the purposes across the work domain in which the stakeholders are particularly interested in. The purpose-related functions level relates to how the functional purpose could be achieved. This level could be thought of as the level in which managers of the work domain would be concerned. The managers need to ensure that the general functions are carried out in order to achieve the functional purpose. Furthermore, the values and priority measures indicate the measures of performance or goal attainment, or in which aspects the purpose-related functions achieve the functional purposes. Domain experts examine these measures to ascertain that the system is functioning properly. For example, some measurements that convey performance or goal attainments are percent efficiency, percent capacity or error ratio's. Lasty, the lower levels describe the system in its physical form. So, of what physical objects the system consists of and how these resources are utilized and shaped into processes for achieving the purpose-related functions.

An abstraction hierarchy was established by examining literature. Afterwards, a systematic literature review was conducted to provide a clear synopsis and assessment of the research that has been carried out in the scientific field of co-located and distributed collaboration. The literature review illustrates the main findings of the research that has been conducted so far, while the abstraction hierarchy visualizes the object-related work processes, values, and functions of remote collaboration. Furthermore, the abstraction hierarchy visualizes the evidence of which social cues support certain object-related work processes and how the perception of these social cues are enabled trough CSCW tools. Thus, the previously established abstraction hierarchy was modified to visualize the relations between the work processes of remote collaboration, and the social cues.

1.2. Research questions

This study will attempt to address the following research questions:

- What are the object-related processes, values and general functions of remote collaboration?
- How are the object-related processes, values and general functions of remote collaboration related to each other?
- How do the implementations of social cues in the current CSCW tools support the object-related processes for specific remote collaborative tasks?

2. Methodology

2.1. Current Study

To determine which social cues are enabled to be perceived during synchronous remote collaboration, and how those social cues support the work-related processes for specific tasks during remote collaboration, we first establish the preliminary abstraction hierarchy for describing teamwork and technologies in remote teams. Next, a systematic literature study is conducted to verify relations based on evidence provided by the included literature between the social cues and the object-related processes. The selection of the articles that were included in the final sample of the literature research followed the guidelines set out by PRISMA (Moher, Liberati, Tetzlaff, & Altman, 2009).

To already gain an impression of which concepts are involved in remote collaboration, all the object-related processes and functions of remote collaboration have been documented during the final selection of articles. Furthermore, additional literature (literature excluded from the qualitive synthesis) was examined to define and link these concepts to establish a preliminary abstraction hierarchy. After examining the evidence provided by the included literature of the systematic literature review, the preliminary abstraction hierarchy was modified to visualize the relations between the work processes and social cues in the object-related processes level to establish the final abstraction hierarchy that describes and visualizes: (1) how social cues are enabled through technology in CSCW tools, (2) how those social cues support work-related processes, and (3) how those processes further support the values, and function of remote collaboration to achieve the functional purpose of remote collaboration. In the paragraphs below we describe how we conducted the systematic literature review.

2.2. Eligibility criteria

A set of criteria was determined before starting the systematic literature review. All the articles that met these criteria were included in the final sample of the literature review. The eligibility criteria focused on including empirical studies that (1) had a sufficient quality of evidence (see section 2.6.) and (2) were in the scientific field of co-located and distributed collaboration. A publication was only included if:

1. It discusses synchronous collaboration. Therefore, the team members had to be working at the same task at the same time.

2. It was empirical, meaning that the evidence that was found in the study is derived from empirical research.
3. It discussed co-located collaboration, distributed collaboration, or a comparison of the two.
4. It was assessed to be of sufficient quality, as determined by the “Standard Quality Assessment Criteria” proposed by Kmet, Cook and Lee (2004).

2.3. Information source

Scopus was used as the sole source of information for this study. The reason for this is that (1) Scopus provided a sample that was large enough as a starting point and (2) Scopus is suitable for replicating the literature search in terms of search reproducibility. In contrast to Google Scholar, which may fail to deliver identical results for identical queries used between certain time periods (Gusenbauer & Haddaway, 2020).

2.4. Search Term

The initial sample was determined on the 27th of January 2021. The following search string was used:

(TITLE-ABS-KEY (gaze OR gesture OR touch OR haptic OR tactile) AND TITLE-ABS-KEY ((collaboration OR teamwork OR cooperation OR cooperative) W/5 (remote OR distributed OR face-to-face OR collocated OR co-located)))

There were three elements that were important for finding the initial sample of articles. Firstly, the articles had to incorporate social cues. Therefore, the first part of the search term specified different social cues: TITLE-ABS-KEY (gaze OR gesture OR touch OR haptic OR tactile). Secondly, there had to be a form of collaboration. This was specified in the second part of the search term string: TITLE-ABS-KEY (collaboration OR teamwork OR cooperation OR cooperative). Finally, the articles had to describe work that was done in a co-located setting, distributed setting or in both settings. This was specified in the last chunk of the search string: (remote OR distributed OR face-to-face OR collocated OR co-located). We chose to use the proximity operator “W/5”, because the work that was done in the different settings had to require collaboration and therefore the second and third elements of the search string had to be within five words of each other.

2.5. Study selection process

The selection of articles was executed by three researchers. Studies that did not meet the eligibility requirements were removed from the sample. Figure 1 visualizes the study selection process. The following steps were executed:

1. **First selection** – Screening the title, when this did not provide sufficient information, the abstract was read.
2. **Second selection** – Reading the abstract and skimming through the introduction and discussion. If did this does not provide sufficient information, the methods section was read to determine the type of study that was conducted.
3. **Final selection** – Reading the full text to determine the final selection of all the articles.

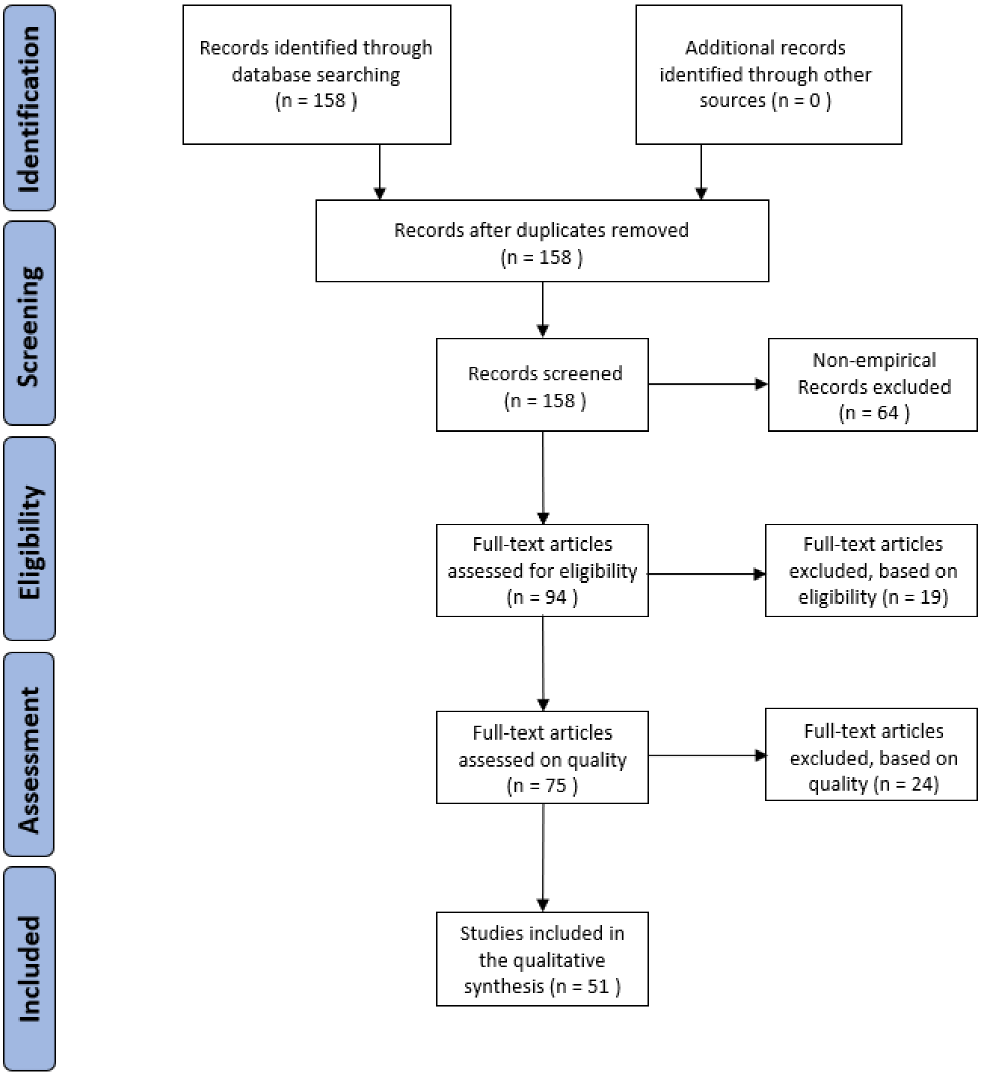


Figure 1. PRISMA flowchart of the study selection process

2.6. Data collection process

The selection included both qualitative and quantitative research in the scientific field of co-located and distributed collaboration. Eligibility criteria to ensure the selection of articles with sufficient quality of evidence was determined by the "Standard Quality Assessment Criteria" as proposed by Kmet, Cook & Lee (2004). This assessment is a method to evaluate the quality of qualitative and quantitative medical research. "Quality" is defined in terms of internal validity of the studies or the extent to which the design, treatment and analyses minimize errors and biases (Kmet, Cook & Lee, 2004).

In their study, they implemented two checklists with scoring systems: one for quantitative research, and one for qualitative research (see Appendix A, Appendix B respectively for the used questionnaires). Two questions in the quantitative checklist concerned the blinding of researchers and subjects. While blinding is important to minimize bias and maximize the validity of results, it is not feasible in the domain of co-located and distributed collaboration. Therefore, both questions are not included for assessing the papers in the current study. Because the checklists differ in the number of questions, the mean question score for each article was calculated (see Appendix C). Figure 2 shows the distribution of the articles by mean question score. The distribution is skewed to the right, and at the score of 1.25, the moving average line becomes more gentle. This represents the asymmetric tail of a skewed distribution. Articles with a mean question score higher than 1.25 are included in the study. This resulted in the inclusion of 51 articles in the qualitative synthesis and the exclusion of 24 articles.

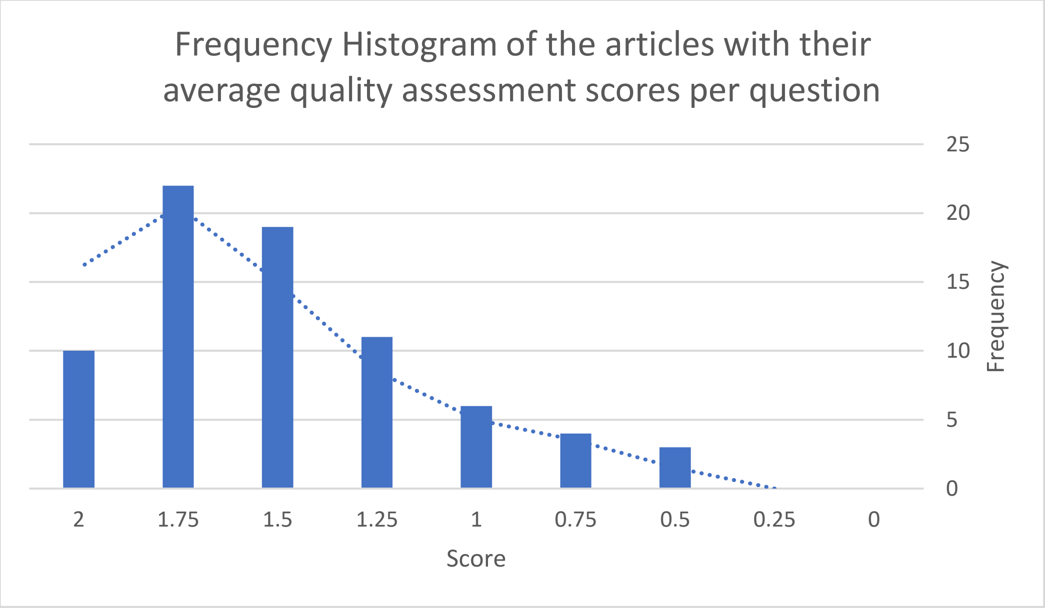


Figure 2. Frequency of articles by mean quality assessment question scores with intervals of 0.25. The dotted line represents the moving average line.

3. Abstraction Hierarchy

3.1. Establishing the pre-liminary Abstraction hierarchy

As a starting point for establishing the abstraction hierarchy, the object-related processes and functions of remote collaboration featured in the final selection of articles were documented (see Figure 3). The relative frequencies encompass the percentage of included articles that feature one or more particular value, function or process. For example, Figure 3 illustrates that the attribute "co-referencing" is featured in 53% of the included literature.

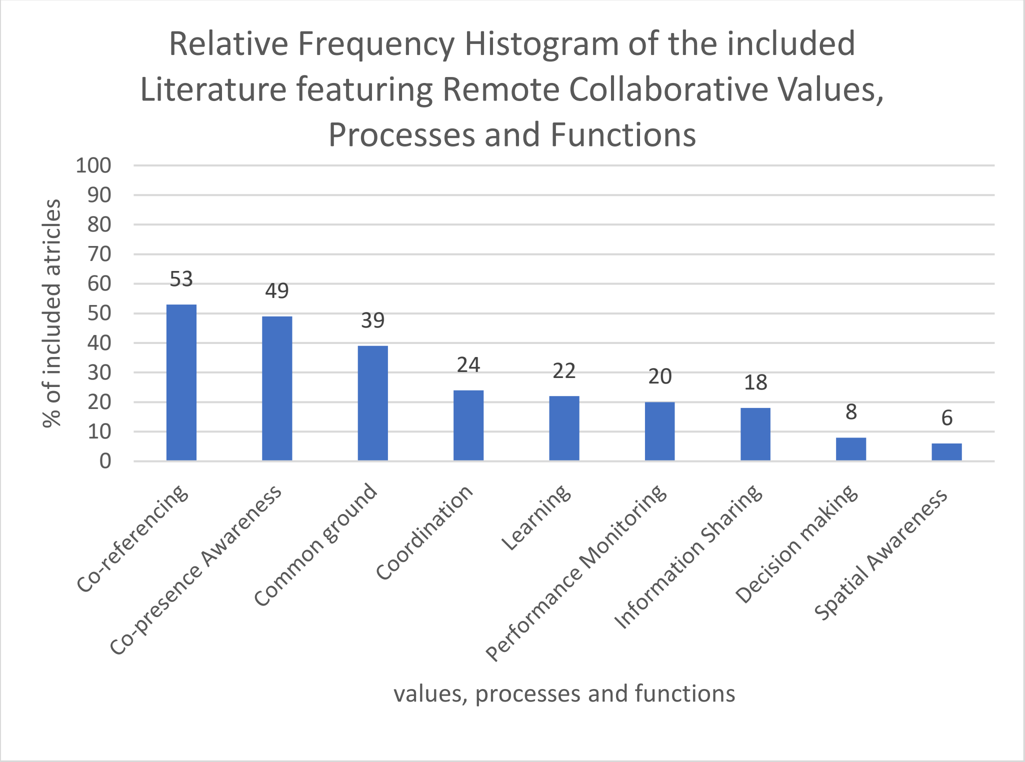


Figure 3. Relative frequencies of the included literature featuring values, processes and functions of remote collaboration.

As mentioned above, an abstraction hierarchy follows a multileveled representation that conceptualizes a system for each level in a set of attributes. The higher levels of an abstraction hierarchy represent the system in terms of the functional purposes, values and priority measures, while the lower levels represent the system in terms of purpose-related functions, object-related processes, and physical objects (Lind, 2003; Reising, 2000).

In the following paragraphs the levels mentioned above and their contents will be presented to describe teamwork and technologies in remote teams.

More attributes may be added throughout the assessment of additional literature, as the attributes in Figure 3 only serve as a starting point. Moreover, the level under which certain level attributes will be subdivided is elaborated. An italics font is used to create overview as to the relations between the attributes of adjacent levels.

Lastly, these relations will be visualized in the pre-liminary abstraction hierarchy trough dotted lines which connect the attributes of adjacent levels (see Figure 4).

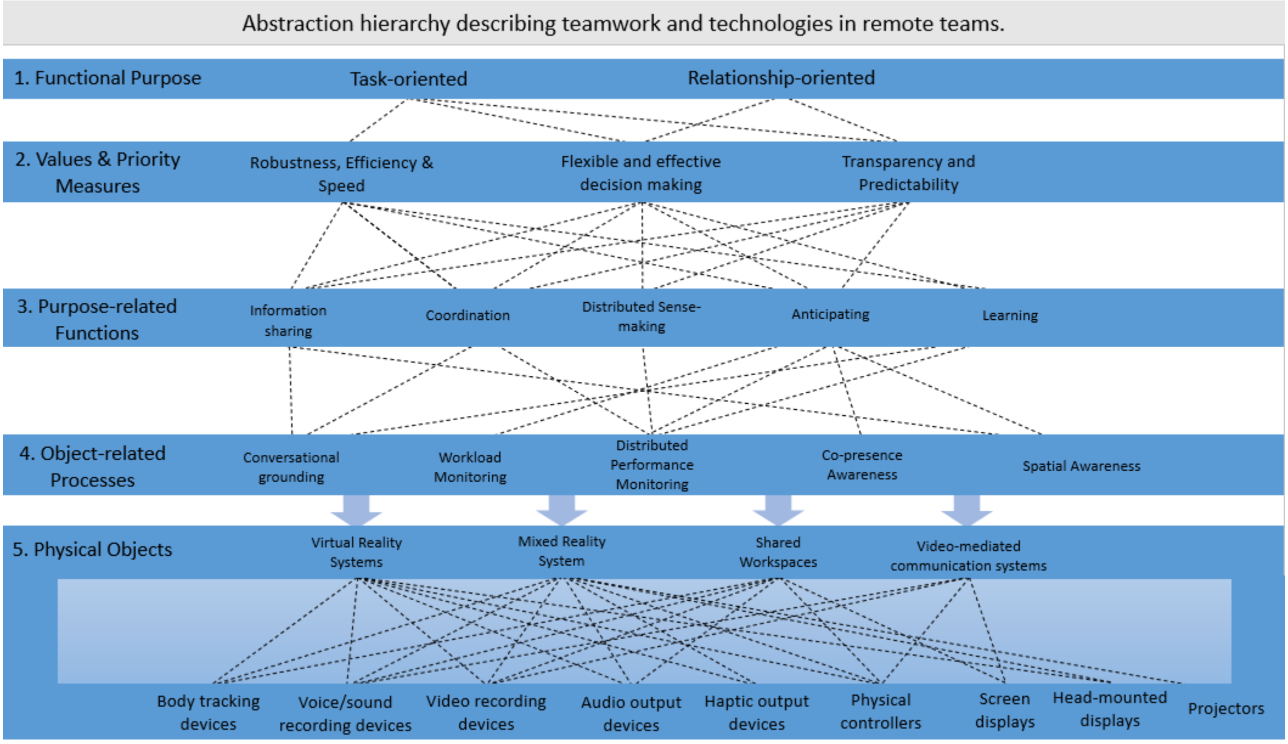


Figure 4. Preliminary abstraction hierarchy consisting of the functional purpose, values and priority measures, purpose-related functions, object-related processes, and the physical objects needed to realize the collaboration.

3.1.1. Functional purpose

3.1.1.1. Task-oriented

Teams can be defined as a group of individuals where members have different specializations in roles and work together to accomplish common goals (Neale, Carroll & Rosson, 2004). The interdependence of tasks and the coordination to accomplish these goals are the defining characteristics of a team. Therefore, teams differ from groups in the sense that groups have limited role differences in their task structures. Consequently, group performances mainly depend on individual efforts rather than joint efforts, because group members are generally not dependent on each other for accomplishing their task goals. On the contrary, the task structure of team members consists of task-driven interactions in which team members must work closely with each other to accomplish the team's goals (Hertal, Konradt & Orlikowski, 2004). In turn, the accomplishments of one member on their task has strong implications on the work processes of other team members. Therefore, the team member's own goals and the team's shared common goals are related to one other. This refers to goal interdependence, that is: "the degree in which teams have clear goals or a clearly defined mission, and the extent to which member's goals are linked to these team goals" (Campion et al., 1993, as cited in Hertal, Konradt & Orlikowski, 2004, p.4).

3.1.1.2. Relationship-oriented

Given the role differences in teams, individuals are generally specialized for specific tasks within the task structure. The different specializations within the team ensure diversity. Diversity is defined as "the distribution of differences among team members of a unit concerning a common attribute such as tenure, ethnicity". (Harrison & Klein, 2007, as cited in Shin, Kim, Lee & Bian, 2012, p.198). The value of team diversity resides in the increased range of overall team skills, perspectives, and knowledge. For this reason, it is important that team members are aware of each other's specializations, so that team members can complement each other on task activities that require a variety of subject knowledge and skills. Therefore, teams should also focus on relationship-oriented activities that focus on improving social relationships and clarifying roles (Klein et al., 2009). In turn, relationship-oriented activities can also tackle interpersonal problems between team members, which can affect the whole team's performance.

3.1.2. Values & priority measures

As mentioned, the task interdependencies among team members require interactions across multiple specialized team components to accomplish the *functional purpose*. The values and priority measures indicate how well the purpose-related functions are achieving the function purpose of the system. Therefore, the values and priority measures covey some measure of performance or goal attainment. These include (1) robustness and efficiency, (2) flexible and effective decision making, and (3) transparency and predictivity.

3.1.2.1. Robustness, Efficiency, and speed

Some terms related to efficiency in the working domain are: faster, better, cheaper and pressures (Woods & Hollnagel, 2006). As mentioned, the characteristics of the task and the CSCW tools place demands on the task activities of collaborators. Consequently, collaborators are required to keep up with the pace of task activities. Therefore, task activities need to be carried out as efficiently as possible. To increase efficiency, it is important to design more robust CSCW tools. This means that CSCW tools should support teams in identifying potential errors or a triggering event at an early stage. Triggering events can create a subsequent cascade of effects, which require collaborators to keep up with the fast pace of changing situations. When CSCW tools are more robustly designed, collaborators might notice signals or critical cues. These signals can distract the current focus on the activities of collaborators, but in other cases, critical cues can shift the focus of collaborators to switch to other work processes, strategies and coordination which may fit the situation at hand better.

3.1.2.2. Flexible and effective decision making

When signals or critical cues are noticed, it needs to be decided whether the attention of team members needs to be re-oriented to other work processes, even when they are busy with other tasks (Woods & Hollnagel, 2006). Generally, decision-making addresses communication across multiple team members who have different perspectives, knowledge, and information of the current situation. Furthermore, decision-making addresses the criteria of how to extract irrelevant data from that information and how to avoid the consequences of false alerts. Additionally, there are specific situations in which goals often conflict. For example, achieving one goal necessarily hinders achieving the other one. Trade-offs and dilemmas are formed when multiple goals interact with one other (Woods & Hollnagel, 2006). It is required to identify how these trade-offs and dilemmas contribute to particular situations so that collaborators can decide how to handle them. However, in practice “human decision making always occurs in a context of expectations in which one may be called to give accounts for those decisions to different parties” (Tetlock, 1999; Lerner & Tetlock, 1999, as cited in Woods & Hollnagel, 2006, p.153). These different systems of accountability can either aid or impair the ability to resolve the goal conflicts.

3.1.2.3. Transparency and Predictability

Predictability refers to one’s actions being predictable enough that others can rely on them when considering and carrying out their actions (Johnson et al., 2014). Mutual predictability is also important, meaning that collaborators should not only focus on making their actions predictable for others, but should also consider the actions of others while planning their own actions (Klein et al., 2004). Team members can become mutually predictable through shared knowledge and CSCW tools that have been developed through extended experience in remote collaboration. Additionally, predictability is essential for synchronizing actions between team members and achieving higher efficiency in team performance (Johnson et al., 2014).

3.1.3. Purpose-related functions

In order to achieve the functional purpose of the system, teams should be able to share information, coordinate, make sense, learn, and anticipate remotely. The following sub-sections describe these purpose-related functions and how they are related to above-mentioned values and priority measures of remote collaboration.

3.1.3.1. Information sharing

According to Fidel et al. (2004), information sharing is about direct information exchanges between individuals to solve a problem. As mentioned before, teams are diverse; they consist of different specialized team members with different task knowledge and domain knowledge. It is important that team members understand the activities of other team members to provide context for one's activity.

Furthermore, teams have a common goal and work together to accomplish this. Because their tasks activities are divided between them, their information needs are different from each other (Talja & Hansen, 2006). This requires team members to communicate and exchange information to satisfy their information needs to keep task activities on track. As such, information sharing is important for promoting the team's performance or *efficiency*, but also for being *transparent* and *predictable* for each other.

Finally, when critical cues or signals are noticed that could result in dilemma, teams need to decide whether the attention of team members needs to be re-oriented. In this case, different perspectives and different knowledge about the involved task activities need to be communicated with each other, to come up with a routine that fits the situation at hand best. Thus, information sharing is necessary for *flexible* and *effective decision making*.

3.1.3.2. Coordination

Woods & Hollnagel (2006), defined coordination as:" the flow of information across distributed parties even with a central actor; the multiple roles played by a single agent at centre stage in the story; the synchronization demands in terms of timing across agents and tasks" (p.66). In agent theory, coordination has been defined as "the process by which an agent reasons about its local actions and the (anticipated) actions of others to try to ensure the community acts coherently" (Jennings 1996, as cited in Boella & van Der Torre, 2006, p. 5) or “the activity that involves the selection ordering and communication of the results of agent activities as that an agent works effectively in a group setting” (Lesser 1998 as cited in Boella & van Der Torre, 2006, p.6). Combining these three definitions, it can be implied that (1) coordination always occurs in the context of multiple agents with different roles, (2) agents are involved in carrying out their own task activities, and these tasks must be carried out in an organized or planned manner, and (3) communication between agents or parties is necessary to ensure the tasks are effectively carried out so that the team goal can be achieved. In other words, when individual tasks-activities are carried out according to plan in terms of timing, then agents work effectively in a group setting. Thus, coordination benefits the team performance or *efficiency*. Furthermore, when there is no flow or smoothness in task activities and no communication between parties, then there is no sequential or explicit analysis of data to assess situations (Woods & Hollnagel, 2006). As such, coordination also plays a role in making *effective* and *flexible decisions*, because correct assessments of situations are important for making the right decisions. Lastly, when task activities between agents are synchronized or performed according to a time schedule, they can predict or anticipate the actions of others. Therefore, coordination also benefits *transparency* and *predictability*.

3.1.3.3. Distributed Sense-making

According to Attfield et al. (2015), sensemaking concerns how individuals use information to construct interpretations of the world around us. Furthermore, Klein et al. (2007) describe sense-making as the process of comprehension. They proposed the data-frame model to offer a detailed account of the cognitive processes involved in sensemaking. The model presents sensemaking as a process of two kinds of entities: data and frame, which interact dynamically. Data are the aspects of the environment that individuals experience as they interact with it, while a frame is the explanation and interpretation of the data. Thus, sensemaking is the process of framing and re-framing considering new data. When an individual encounters a situation, one or two key data are experienced for creating an understanding. These data elements are the anchors to bring out an initial frame. Furthermore, this frame is used to search for more data elements to determine whether the frame needs to be modified or can be maintained. The particular frame that is activated depends on several factors including available cues, motivation, workload, and the individual’s repertoire of frames (Attfield et al., 2015). Individuals’ different repertoires of frames are based on experience training, which explains why experts use different approaches to tackle problems compared to novices. Considering distributed sense-making in teams, cognitive processes transcend the boundaries of individuals and consider the interplay between team members' representations, and the use of artefacts that form a wider cognitive system. Team members have different representations due to different domain knowledge and experienced data. Distributed sense-making requires team members to communicate their experiences and representations to successfully interpret situations as a distributed cognitive system. In turn, with better interpretations of situations, better decisions can be made to implement a certain routine. Therefore, distributed sensemaking affects *effective decision-making*. Furthermore, in distributed cognitive systems, team members that become more *transparent* about their representations are considered to make sense of a situation.

3.1.3.4. Learning

According to Lave and Wenger (1991), learning takes place by participating in the execution of concrete work tasks and through observing their co-workers. Learners see how executors perform actions and interact with information, target objects, tools, and technologies. In other words, seeing is learning. This view is one example that expands the learning process from an individual level to a collective level. Additionally, Ellis et al. (2003) defined team learning as "a relatively permanent change in the team's collective level of knowledge and skill produced by the shared experience of team members" (Ellis et al., 2003, p.822). Therefore, each team member's ability to individually acquire knowledge and skills as well as their ability to collectively share their information with teammates is essential for the *efficiency* and *effectiveness* of the team's collective learning process.

Furthermore, when teams attempt to learn or solve problems, a critical discussion of the available data and ideas is required. Therefore, discussing divergent perspectives is crucial for group problem solving and *decision-making* accuracy (Ellis et al., 2003).

In addition, when a problem is encountered during a specific task activity, adaptation is necessary to work around this gap to continue the plan and meet the relevant goals (Woods & Hollnagel, 2006). Agents adapt the situation they experience and act with their knowledge depending on their mindset to pursue their goals. However, agents may respond narrowly the situation in front of them and might miss the side-effects of their performed actions and decisions. In contrast, learning agents develop broader abstraction models that provide a larger perspective on effective strategies. Without such effort, agents tend to fall back to strategies dominated by local factors and contexts. Therefore, the gap leads to learning about employing specific strategies or work- arounds to bridge the gap. Thus, this gap is an important source of information for learning. Monitoring where and how this gap arises is necessary to achieve better coordination between different levels of control (Woods & Hollnagel, 2006).

3.1.3.5. Anticipation

Anticipation is a fundamental aspect to stay ahead or keep pace with changing situations, and refers to the preparation to cope with unknown future events (Woods and Hollnagel, 2006). As such, anticipation is linked to resilience or *robustness* as resilience refers the ability to anticipate and adapt to potential surprises. In other words, anticipation is about how individuals use the available data of the situation to understand what could happen next. With this in mind, it is necessary to combine multiple perspectives of team members to predict upcoming situations more accurately, so that *effective decisions* can be made to employ a routine that fits future situations. Furthermore (as mentioned in paragraph 3.1.2.3), it is essential that each team member has a responsibility to make their actions sufficiently *predictable* to others, so that other team members can anticipate each other's actions to synchronize tasks and increase the team's *efficiency*.

3.1.4. Object-related processes in remote collaboration

Moving further down the abstraction hierarchy, one arrives at the object-related processes of remote collaboration that underlie the purpose-related functions. These processes consist of: (1) conversational grounding, (2) workload monitoring, (3) distributed performance monitoring, (4) co-presence awareness, and (5) spatial awareness. To make the relationships explicit between remote collaboration artifacts and how they support the collaborator's activities and goals, it is important to identify these processes and how non-verbal behaviours influence these processes. First, in the following section, the work processes of remote collaboration will be defined, after which the relations of these work processes with non-verbal cues in the object-related processes level will be elaborated.

3.1.4.1. Conversational grounding and co-referencing

It is essential for *coordination* to know how a partner is acting with task-relevant objects. In co-located cooperation settings, this can be achieved by verbal communication and observing the partner's actions in the environment. Consequently, conversational grounding arises when visual information is used during verbal communication to establish common ground. However, the lack of visual cues in remote collaboration complicates this process. (Müller, Helmert, Pannasch & Velichkovsky, 2013).

In a shared space, visual cues such as the *gaze direction* and *gestures* of partners can provide *information* that can improve mutual understanding and contribute to a more efficient dialogue. For example, for identifying target objects in a shared space, ambiguity can be avoided by looking at or pointing at the target object during verbal communication, making it possible to use co-references.

Evans, Feenstra, Ryon & McNeill (2011) defined a reference as “an object or other meaning entity nominated in speech and/or indicated in gesture or action” (p. 259). A co-reference is a set of speech and gestures indicating the same objects or referents. There is no specific sequence in the set, and it expands and links interaction with different collaborators. In co-located settings, collaborators are able to use gestures and gaze cues that their partner can perceive. Therefore, collaborators can make use of references to redirect their partner’s attention As a consequence of using co-references, verbal communication can become more efficient because the characteristics of the target object do not need to be described because the collaborators partner has perceived the target object themselves. When it is not possible to redirect your partner’s attention by making use of references, an increase in verbal feedback and descriptive is needed to tackle ambiguities.

Furthermore, being aware of how a partner is interacting with task-relevant objects is necessary to *coordinate* the collaboration process. Additionally, observing how co-workers perform actions and interact with information, target objects, tools and technologies is necessary to *learn* task activities. Conversational grounding is also essential for *distributed sense-making* because team members must communicate their experiences and representations as clearly as possible to successfully interpret situations as a team. Therefore, it is important that CSCW tools enable the perception of the partner’s gaze directions, gestures, and actions in a shared space.

3.1.4.2. Workload Monitoring

There are numerous definitions of workload. For example, Gopher (1986, as cited in Miller, 2001) defined mental workload as: “a hypothetical construct that describes the extent to which the cognitive resources required to perform a task have been actively engaged by the operator” (p.5). This corresponds with the definition proposed by Hoedemaeker (2002, as cited in da Silva, 2014), who argue that workload is defined: “by the amount of resources required by a set of concurrent tasks, as well as by the use of resources needed to perform them" (p.311). Furthermore, Verwey (2000, as cited in Miller, 2001) proposed that "mental workload is related the amount of attention required for making decisions" (p.5).

High levels of cognitive workload can negatively affect human performance (Dell'Agnolo, Momeni, Arza & Atienza, 2020). When agents do not have the cognitive resources required to perform certain tasks, task-errors or wrong decisions can be made. Therefore, workload monitoring is key to *anticipating* and preventing such matters thereby improving the performance of the team. Workload monitoring can be described as the process by which collaborators can perceive or monitor their partner's workload with the task at hand. To prevent task failures, it is important to intervene when a team member is experiencing a high workload. Therefore, collaborators must be able to detect signals of the high workload from their team members. For example, certain actions can be taken to relieve stress when these physical signals are observed in actions visualized as gestures, just like fatigue can be observed in a partners gaze behaviour. For example by giving task-oriented instructions, adjusting the task or by putting extra people on the task. As high workload is often accompanied by making mistakes in subsequent activities of the task, it is therefore important to intervene when a team member is experiencing a high workload.

3.1.4.3. Distributed performance monitoring

Distributed performance monitoring is similar to workload monitoring. However, it is not about perceiving the workload of the other collaborators, but about observing their task performance and monitoring the actions or sub-tasks to assess the progress towards the goal objective. Additionally, being able to monitor each other also benefits *learning*, because as mentioned earlier: seeing is learning.

Coordination is key to ensuring that all the processes are carried out according to plan. When it is likely that a sub-task will not completed in time, when a sub-task is performed incorrectly or the order of sub-tasks is performed incorrectly, certain actions need to be performed to keep task activities synchronized. Therefore, it is important for task managers to maintain a good overview of the entire work process so that task errors can be *anticipated* and diagnosed at an early stage. In addition, keeping an overview of the processes in which the collaborators are involved provides continuous data and the experience necessary to interpret and *make sense* of situations. Perceiving gesture- and gaze cues plays an important role here. *Gestures cues* could provide information whether actions in a relative direction to target objects are performed correctly, while *gaze cues* could provide information whether collaborators are focussing on the right target objects.

3.1.4.4. Co-presence awareness

When you interact with someone face-to-face, you have the sense of being together and communicating with each other (Yoon, Kim, Lee, Billinghurst & Woo, 2019). Co-presence refers to the feeling of being together with another person through a medium (Bula, 2012; Jo, Kim & Kim, 2016; Sallnäs, Rassmus-Gröhn & Sjöström, 2000). As such, it refers not only the sense of being in the same environment as other individuals but also being mutually aware of one other, which is determined by the sensory properties of the virtual environment. Furthermore, being aware of each other is necessary to *anticipate* each other. Overall, co-presence is a single dimension that represents a cognitive synthesis of different factors to transmit information like *gestures*, facial expression, tone of voice, posture, touch, and *gaze direction*. These factors affect the level of presence that is perceived by users interacting remotely in a certain medium. Co-presence differs from social presence in the sense that social presence relates to the quality and user perception of the medium, while co-presence addresses the interaction between individuals within the medium. However, co-presence and social presence are somewhat related to one other, as users who perceive the quality of the medium to create more intimate interactions also tend to feel more togetherness with one another in the virtual world.

Several studies have found that shared *gaze awareness* improves the sense of co-presence (Bai, Sasikumar, Yang & Billinghurst, 2020; Gupta, Lee & Billinghurst, 2016; Lee et al., 2017b; Wang et al., 2020b). Moreover, visualized gestures in remote collaboration also increase the feeling of co-presence (Bai, Sasikumar, Yang & Billinghurst, 2020; Piumsomboon et al., 2018; Yang et al., 2020; Wang et al., 2019; Wang et al, 2020c). Additionally, visualized *shared gestures* of collaborative partner's display an interaction. For example, individuals can put their thumbs up to communicate that everything is well, or they can nod their heads during conversations to indicate that they understand the dialogue. This could reinforce the feeling of sense of being together. Therefore, the sense of co-presence is not influenced by performing gestures but rather by perceiving the gestures of others.

3.1.4.5. Spatial awareness

It is of importance that users are aware of and understand their environment over time, especially when the workspace is shared with other collaborators (Irlitti, Piumsomboon, Jackson & Thomas, 2019). Without such awareness, it is difficult to gain an understanding of every other collaborator's actions within the shared workspace. Therefore, *gaze*- and *gesture* cues play a similar role compared to distributed performance monitoring here, as both these cues provide understanding of their collaborators actions within the shared environment. Consequently, this promotes *sense-making* and *anticipation*. For example: in a shared face-to-face environment, collaborators' positioning provides socio-contextual information; when someone is positioned with their head towards you while looking at you, you might conclude that they want to talk to you or that they are currently monitoring you to ensure that the work processes stay on track. Furthermore, spatial awareness is necessary to *share* spatially related *information*. For example, by pointing to target objects or describing target objects in terms of their position within the shared environment. Such information provides understanding for collaborators to identify target objects in the shared environment.

Furthermore, interpersonal distances between individuals in a shared space are an integral part of nonverbal communication (Williamson, Vinayagamoorthy, Shamma, & Cesar, 2021). As aforementioned, the proxemic distance between collaborators could provide richness and flexibility during verbal conversations as the distance people keep between each other during conversations can describe the relationship between the two. For example, you may conclude that two persons know each other personally when they are having a conversation while keeping a distance of 0.45m - 1.2m, or you may conclude that two people are engaging in collaborative activities because they are seated 1.2m-3.6m from each other while having interactions. Proximal behaviour is also interconnected with *gaze behaviour*, as individuals balance increased proximity with reduced eye contact.

3.1.5. Physical objects: Technologies and hardware

At the final two levels of the abstraction hierarchy, remote collaboration is described in its physical form. In the following sections, we describe the different systems of remote collaboration that enable the perception of social cues of others. These consist of (1) virtual reality systems, (2) mixed reality systems, (3) shared workspaces, and (4) video-mediated communication systems. Furthermore, a brief overview of which hardware components these systems consist of is given.

3.1.5.1. Virtual reality systems

Virtual reality (VR) is characterized by the illusion of being present in a synthetic environment rather than observing such an environment externally (Earnshaw, 2014). Moreover, VR is a communication medium (Sherman, 2003). A medium refers to something that is between two or more things, through which anything can be passed from one point to the other. In the case of VR, information such as data of the synthetic environment is passed to another. Furthermore, the information can be accessed by others with the user interface (UI). Typically, the hardware components of the UI consist of input devices such as *body tracking devices* (gestures/eye gaze), *video recording devices, voice/sounds recording devices* and *physical controllers* (keyboard/mouse), along with output *visual, audio,* and *haptic output devices* (Earnshaw, 2014).

3.1.5.2. Mixed reality systems

Mixed reality (MR) is like VR and augmented reality, as it is characterized by the illusion that digital objects are in the same space as physical ones (Costanza, Kunz & Field, 2009). Therefore, it is a subclass of VR, as it involves merging real and virtual worlds in real-time. Furthermore, considering the similarities to augmented reality, MR includes systems in which the virtual aspects are as dominant as the physical aspects, in contrast to augmented systems, where physical aspects are more dominant than the virtual aspects. Typically, the hardware components of the UI are similar to that of VR systems, like *body tracking devices, video recording devices, voice/sounds recording devices* and *physical controllers*, along with *output visual, audio,* and *haptic output devices*. Additionally, MR systems could utilize ambient *projectors* or hand-held *projectors* to project visual objects directly into the environment rather than addressing user’s perception trough a head-mounted or hand-held display (Costanza, Kunz & Field, 2009).

3.1.5.2. Shared Workspaces

In face-to-face meetings, collaborators can see, point at, or draw on a whiteboard simultaneously. Furthermore, an overhead projector makes computer-generated or handwritten documents visible for all collaborators (Ishii, Kobayashi & Grudin, 1993). In such cases, shared workspaces are simply a physical space where collaborators can partake in a joint activity (Gutwin & Greenberg, 1999).

In distributed real-time collaboration, such activities can be supported by computer-based groupware. Ellis et al. (1991, as cited in Ishii, Kobayashi & Grudin, 1993) proposed a groupware definition that takes such a work-space oriented view into account, “groupware … the computer-based systems that support groups of people engaged in a common task (or goal) and that provide an interface to a shared environment” (p.352). Therefore, shared workspaces in the context of real-time groupware systems could be defined as an “abounded space where people can see and manipulate artifacts related to their activities” (Gutwin &

Greenberg, 1999, p.5). The hardware components of such UI's consist of input devices like *voice/sounds recording devices, body tracking devices* (gestures/eye gaze) and *physical controllers* (keyboard/mouse), along with *visual* and *audio output devices* (Ishii, Kobayashi & Grudin, 1993).

3.1.5.4. Video-mediated communication systems

Video-mediated communication systems are characterized by using video technology to allow the perception of non-verbal cues of others to enrich conversations across greater distances, which are typically used for videoconferencing (Bohannnon, Herbert, Pelz & Rantanen, 2013). For example, laptop computers and iPhones are both considered video-mediated communication systems, because both devices allow you to use applications or video-conferencing software like Facetime, Microsoft Teams or Skype to have conversations supported by the visual image of the speakers. Considering face-to-face communication as the richest medium that provides immediate feedback and conveys many cues alongside speaking, videoconferencing falls below face-to-face communication, but above telephone communications in term of information richness. The hardware components of video-mediated systems consist of input devices like *voice/sound recording devices, video recording devices* and *physical controllers* along with *visual* and *audio output devices*.

3.2. The preliminary abstraction hierarchy for describing teamwork and technologies in remote teams

In the sections above, we have described the attributes of remote collaboration for the different levels of the abstraction hierarchy and how these attributes are related to each other. Based on this, the preliminary abstraction hierarchy for describing teamwork and technologies in remote teams was modified to visualize the processes of perceiving gesture and gaze cues (see Figure 5). The processes of perceiving gesture and gaze cues relate to the object-related processes of remote collaboration because the physical objects level indicates how these social cues could be perceived by utilizing certain technologies. Furthermore, the perception of these social cues influence the object-related processes of: conversational grounding, workload monitoring, distributed performance monitoring, co-presence awareness, and spatial awareness. Therefore, the social cues are visualized in the object-related processes level, underneath the object-related processes of the pre-liminary abstraction hierarchy. The dotted lines represent the hypothetical relationships of concepts between the different levels. During the systematic literature review, the results of the included studies will be compared, and the processes will be linked to the social cues of the preliminary abstraction hierarchy. When the results provides evidence for a relationship between these concepts, the dotted lines will be adjusted to solid lines in the final abstraction hierarchy. The final abstraction hierarchy provides an overview of which social cues and which work processes in the object-related processes level are proven to be related to each other.

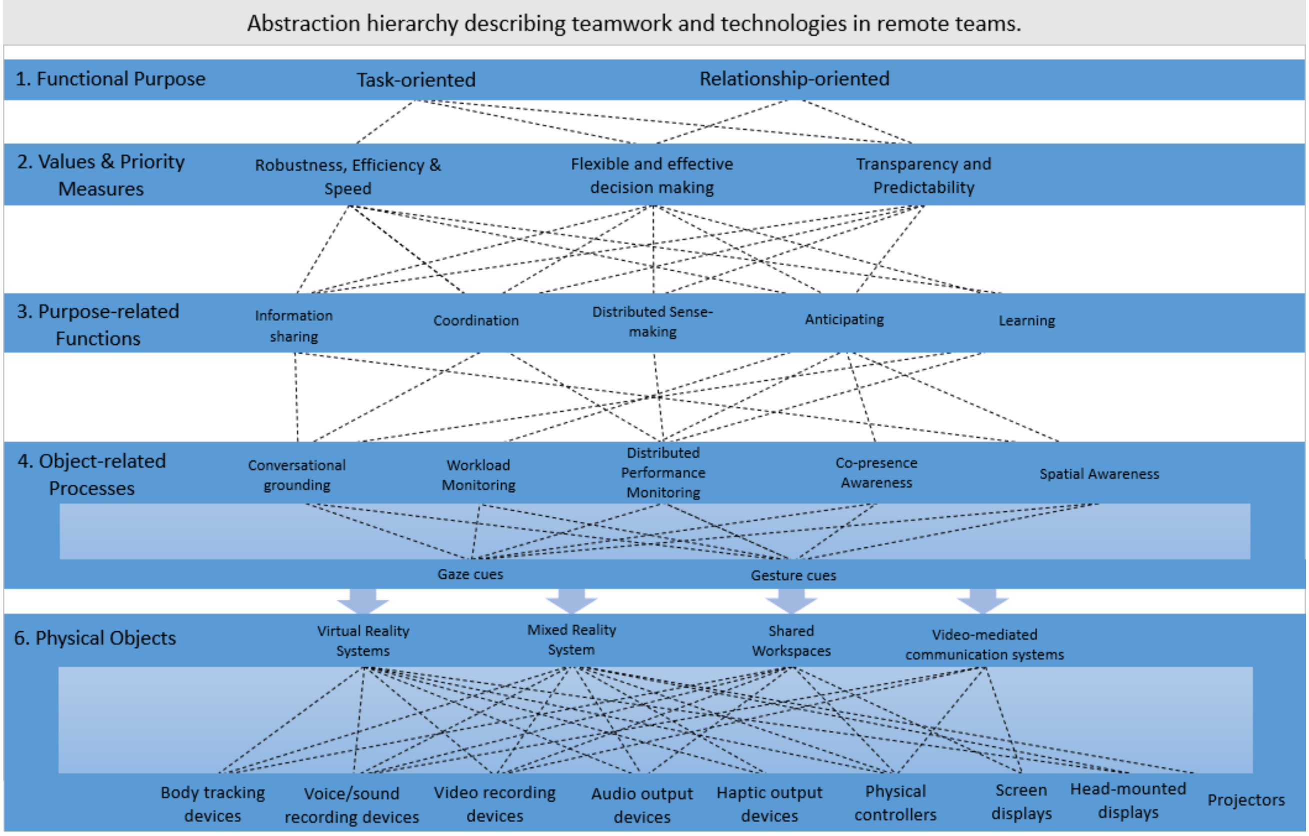


Figure 5. Modified preliminary abstraction hierarchy consisting of the functional purpose, values and priority measures, purpose-related functions, object-related processes including the social cues that could be perceived and the work processes of remote collaboration, and the physical objects needed to realize the collaboration.

4. Literature Review

4.1. Results

To structure findings that emerged from the literature (see Table 1), we will initially focus on issues that arise during remote collaboration compared to during co-located collaboration. Second, we made the relationships explicit between non-verbal behaviours and the underlying object-related work processes, and how these processes influence collaboration at a distance. The relationships between the social cues and object-related work processes have been structured based on evidence in Table 2. Furthermore, the abstraction hierarchy has been modified to illustrate the relationships between the social cues and the work processes in the object-related processes level (see Figure 7).

Two thirds of the studies that compared co-located collaboration to remote collaboration, found that co-located collaboration is more favourable for faster task completion times (D'Angelo & Gergle, 2016; Fussell et al., 2003; Fussell et al., 2004; Kunz, Nescher & Küchler, 2010). According to these studies, co-located collaborators use more pointing gestures, deictic references and acknowledgements of behaviours compared to remote collaborators, even when artifacts have enabled the ability to share gaze and gestures between remote partners. According to Woods and Hollnagel (2006), characteristics of artifacts and tasks interact to place demands on the coordination between collaborators. Furthermore, these demands and task characteristics interact to specify how artifacts should support work. To promote remote collaboration, the artifacts must provide an affordance, in this case, people's natural ability to perform and perceive non-verbal behaviours. Many remote collaboration artifacts have been developed that make it possible to observe gaze behaviours. In the next section, we describe various implementations that enable people to perceive non-verbal gaze cues, and whether such systems indeed offer an affordance. Furthermore, we describe other implementations that enable to convey information during remote collaboration. Afterwards, we describe how gaze cues and gesture cues facilitate the processes of remote collaboration.

4.1.1. Implementations to perceive gaze behaviours with CSCW tools

The majority of the studies developed remote collaboration artifacts that enabled the ability to perceive non-verbal gaze behaviours (Bai et al., 2020; D'Angelo & Begel, 2017; D'Angelo & Gergle, 2018; Higuch, Yonetani, & Sato, 2016; Kütt et al., 2020; Lee et al., 2017a; Lee et al., 2017b; Lee et al., 2018; Müller et al., 2013; Ou et al., 2015a; Piumsomboon et al., 2019; Wang et al., 2020b; Yang et al., 2020; Zhang et al., 2017). For instance, cursor pointers, heat maps, path trajectories and spotlights can all visualize the collaborator’s gaze in a 2D workspace, while head pointers and view frustrums (see Figure 6) can visualize the partner’s gaze in VR environments. However, gaze visualizations like heat maps and path representations can also be distracting, as indicated by participants in the study of D’Angelo and Gergle (2018). Consequently, heat maps had significantly higher performance times compared to a control group with no gaze visualization. Gaze visualizations do not seem to support collaborators activities or goals, meaning that these artefacts do not provide an affordance to maintain focus on the task at hand.

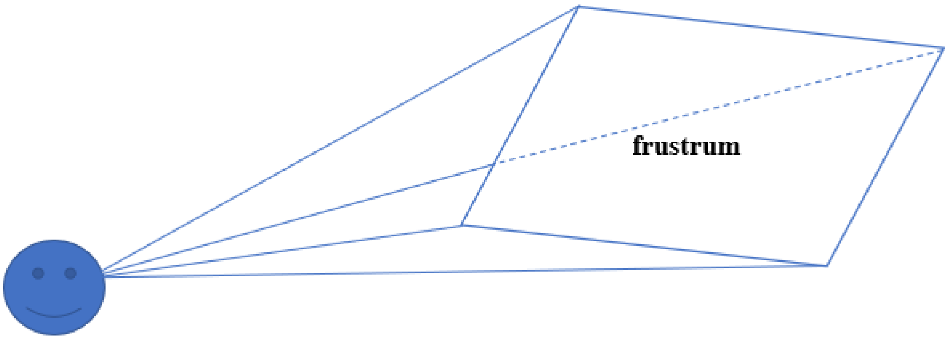


Figure 6. Illustration of a head or view frustrum

4.1.2. Other implementations in CSCW systems that facilitate remote collaborative work processes

Other CSCW systems have been developed that use other approaches to convey information in remote collaboration. For instance, systems have been developed that allow the ability to draw annotations that are visualized in the shared workspace (Gauglitz, Nuernberger, Turk, & Höllerer, 2014; Nuernberger, Lien, Höllerer & Turk, 2016; Wang et al., 2019). According to the study of Nuernberger, Lien, Höllerer and Turk (2016), users tend to draw arrows, circles, and outlines to re-direct attention to target objects. Therefore, instead of providing gaze- and gesture cues in a shared space, non-verbal annotation cues could also provide the ability to reference to objects, locations, and directions. Moreover, annotations could also indicate certain actions needed to be performed on a specific object. Consequently, annotations promote the efficiency of communication like how gaze- and gesture cues disambiguate co-occurring speech. However, performing gestures is a more natural or intuitive way of interacting. The study of Wang et al. (2019) showed that conveying information by performing gestures allowed for a better sense of togetherness compared to conveying information with the use of annotations. Nonetheless, it could be a problem to use deictic gestures when remote collaborators do not have the same viewpoint. Furthermore, 2D annotations can unveil image segmentation problems in a 3D environment. From multiple viewpoints, the annotations are fixed to the object of interest in different orientations. This causes annotations to be occluded by the object or to be inaccurately anchored to the object of interest from certain viewpoints. This may cause remote collaborators to misinterpret spatial information in the task scene. Nuernberger, Lien, Höllerer & Turk (2016) proposed a method to stabilize 2D annotations in a 3D task scene, so that for every viewpoint the annotation conveys the intended information. Their study showed that the proposed method for stabilizing annotations allowed to better convey the user’s attention for both action and referencing tasks compared to a method that did not allow the annotations to be stabilized in the virtual scene.

Another system developed by Yang et al. (2020) provided the virtual task scene with spatial auditory cues. In their study, they investigated the role of the integration of three types of auditory cues: (1) non-spatial voice, this allows communication between the local worker and a remote expert like a typical audio call, (2) spatialized voice, this allows the worker to hear the expert’s voice from a location in the real world that corresponds to the position of the remote expert in the VR scene, (3) spatialized auditory beacons allow the remote expert to virtually attach and play auditory beacons at target objects (thus, spatializing the beacons from the objects to the local worker). Results indicated that there were no differences in terms of performance times and social presence between the spatialized auditory cues and the baseline condition of non-spatial voice. However, there were differences in spatial presence between the cue conditions, indicating that participants had an improved spatial perception of target objects with the spatialized auditory cues.

Table 1. *Findings, study methods and limitations of the included literature classified by collaboration type and technology*

Study	Collaboration	Technology	Study methods	Findings	Methodological limitations
Johnsen-Glenberg et al. (2014)	Co-located	Mixed reality	Comparison of (1) a regular classroom instruction against (2) mixed-reality educational content	Metaphoric gestures may serve as both primes and memory retrieval cues Physical proximity to one other played a critical role for collaboration.	There may have been a small test bias in the regular condition as students already were exposed to certain formulas beforehand.
Terken and Sturm (2010)	Co-located	Projected gaze and speaking time feedback	Comparison of (1) regular group-meetings against (2) group-meetings including gaze and speaking time visualizations	There was a significant effect of the visualization found for under- and over-participators, who changed their speaking behaviour because of the feedback, micro-patterns analysis indicated that feedback on gaze behaviour was not effective.	Some participants knew each other beforehand.
Fussell et al. (2004)	Co-located and remote	Video mediated system	Comparison of (1) a cursor-based pointing device against (2) only a video feed during co-located a remote collaboration	Performance was significantly faster in the co-located condition than in both the distributed conditions with and without the pointing device. The ease of identifying referents and coordination was rated the highest for the co-located setting. The cursor-tool did not improve performance times compared to the use of video alone. However, higher rates of cursor usage were correlated with faster performance times.	The two studies were conducted under different network conditions.
	Remote	Video mediated system	Comparison of (1) a pen-based drawing tool against (2) only a video feed	Participants had faster performance times using the drawing tool than using the video camera alone. Coordination and ease of identifying referents were rated the highest with the drawing tool. Pointing gestures were the most prevalent for indicating objects and locations while providing instructions. In terms of efficiency of communication, helpers used fewer words to complete the task with the drawing tool.	

Table 1. Continued

Study	Collaboration	Technology	Study methods	Findings	Methodological limitations
Piumsomboon et al. (2019)	Remote	Mixed reality	Comparison of three gaze cues: (1) field-of-view, (2) eye-gaze and (3) head-gaze against a baseline condition.	The baseline condition had a significantly higher number of gestures used compared to both head and eye gaze conditions. The head and eye gaze conditions had a significantly higher rate of mutual gaze than the baseline condition. The total travelled distance in both gaze conditions were significantly lower than in the baseline condition.	Limited sample size (N=16)
Evans et al. (2011)	Co-located	Shared Workspace	Observations	Leaners are more likely to discuss their ideas in a computer-supported setting. However, there is a decrease in gestural communication in such a setting.	Unclear why groups were formed based on gender.
Bai et al (2020)	Remote	Mixed reality	Comparison of (1) augmented gaze and (2) gesture cues.	Combined visual cues (gaze and gesture) reduced communication time compared to using only verbal cues. However, there was no significant difference among the conditions of visual cues in performance time (gesture vs gaze vs gesture & gaze). The visual cues significantly increased the feelings of co-presence. In terms of collaboration roles, there was a significant difference in co-presence for both sides	Relatively small sample size (N=12)
Piumsomboon et al. (2018)	Distributed	Mixed reality	Comparison of (1) Mini-me against (2) a baseline condition.	The presence of an adaptive avatar resulted in a significantly higher aggregated social presence score for symmetric and asymmetric remote collaboration, compared to the absence of such avatar. Furthermore, they found that there were significantly lower ratings for task difficulty and mental effort for asymmetric collaboration when the adaptive avatar was used	Some participants had prior experience with VR/AR, which could have influenced the data. There was no practice trial.

Table 1. Continued

Study	Collaboration	Technology	Study methods	Findings	Methodological limitations
Sallnäs, Rassmus-Gröhn & Sjöström (2000)	Remote	Shared Workspace	Comparison of (1) haptic force feedback and (2) no haptic feedback in a virtual desktop shared workspace.	Haptic force feedback significantly improved task performance, increased perceived task performance and increased virtual presence.	The subjective ratings could be influenced by response bias.
Gupta, Lee and Billinghurst (2016)	Remote	Video mediated system	Comparison of (1) gaze sharing and (2) pointing cues.	Task performance was significantly improved by using both pointing and eye tracking cues. Using both cues resulted in the fastest task completion time, while using no cues resulted in the slowest task completion time. Moreover, providing visual cues allowed participants to perform more efficiently and let to better communication with each other.	Limited sample size (N=15). It is unknown whether pairs knew each other beforehand.
D'Angelo and Begel (2017)	Remote	Shared Workspace	Comparison of two tasks including (1) gaze awareness against (2) no gaze awareness.	Gaze awareness allowed collaborators to communicate a higher rate of implicit and explicit references. Consequently, dietic references were more successfully acknowledged	It is unknown whether the participants were randomly paired.
Müller et al. (2013)	Remote	Shared Workspace	Comparison of (1) gaze transfer against (2) a mouse pointer.	The main effect for the communication condition was exclusively based on the longest solution times for speech. Gaze transfer compared to purely verbally interaction led to fewer errors, shorter dialogues, and less specific objects descriptions. According to the overall performance measures, the conditions gaze, gaze & speech, and mouse & speech were similar effective.	It is unknown whether there were demographic variable differences between the groups.
Huang and Alem (2013)	Remote	Video mediated system	User study of a system that allowed remote helpers to perform natural hand gestures, which were displayed onto the task scene.	Helpers gave higher usability ratings for easy of learning, environmental awareness, perception of interaction and pointing gestures, while workers rated higher on task satisfaction, mobility, co-presence, and representational gestures. There were no significant differences between the subjective ratings of workers and helpers.	Absence of a control condition. There is no clear description of the statistical analysis used.

Table 1. Continued

Study	Collaboration	Technology	Study methods	Findings	Methodological limitations
Wang et al. (2020a)	Remote	Mixed reality	Comparison of (1) a tangible drawing interface against (2) a mid-air free drawing interface.	There were no significant differences in performance times and operational errors between the two interfaces. Both interfaces had a similar effect on building common ground.	Measurements of common ground and co-presence were purely based on the feedback from the remote experts.
Kütt et al. (2020)	Remote	Shared Workspace	Comparison of (1) shared gaze in voice-based communication against (2) shared gaze in text-based communication.	Shared gaze reduced the perceived effort during voice-based communication and increased it during text-based communication. However, shared gaze did improve perceived performance for both voice- and text-based communication. Voice-based communication benefits from gaze information in terms of significant shorter completion times and reduced perceived cognitive load compared to text-based communication with and without gaze information.	Limited sample size (N=14)
Cabibihan et al. (2012)	Remote	Video mediated system	Comparison of (1) only robotic speech description against (2) robotic pointing gestures and speech descriptions.	There were no significant differences in term of object recall between these two conditions. In their replicated study, participants recalled significantly more spatial locations in the speech and gesture condition compared to the speech only condition.	There was a 10-minute limit per condition. This may cause time pressure. There is no estimate of variance (e.g., confidence intervals, standard errors).
Ou et al. (2005a)	Remote	Video mediated system	Comparison of 3 puzzles with different piece differentiability and complexity.	The gaze pattern of helpers varies as a function of their instructional task process. When describing a puzzle piece, helpers look at the pieces bay, whereas when there are describing a location, they were more likely to look at the workspace. In terms of worker’s actions, the gaze of the helpers toward the pieces bay was higher when the worker was acting within that area, and vice versa when the working was acting in the workspace	Relatively small sample size (N=12)

Table 1. Continued

Study	Collaboration	Technology	Study methods	Findings	Methodological limitations
Chan, MacLean and McGrenere (2008)	Remote	Shared Workspace	Comparison of (1) haptic feedback for turn-taking against (2) video only turn-taking cues.	There were no significant differences in the distribution of methods of acquiring control (gentle request, urgent request or take control) between the haptic conditions and the visual condition. There was a higher frequency of turnovers in the haptic conditions compared to the visual only condition, this approached a significant difference.	Relatively small sample size (N=4)
Higuch et al. (2015)	Remote	Video mediated system	Comparison of 3 immersive telepresence conditions:(1) Mirror, (2) Tilt, and (3) Hybrid against (4) standard video conferencing.	Standard video conferencing had significantly more horizontal errors than the hybrid and mirror condition. The mirror condition had significantly more vertical errors than the video and hybrid condition. In terms of task completion time, the mirror condition was significantly faster than the video condition. Subjective ratings indicated that hybrid and mirror conditions were ranked significantly better on sense of being together, conveying ideas and perceiving intentions. Moreover, the hybrid condition was also ranked significantly better than standard video conferencing on reading the partner's agreement level.	Relatively small sample size in study 1 (N=6) and study 2 (N=10). Absence of control condition in the second study.
D'Angelo and Gergle (2018)	Remote	Shared Workspace	Comparison of three gaze visualizations: (1) heat map, (2) path representation, and (3) shared area.	Gaze visualisations as path representations and heat maps can be distracting. For the path representations, participants spend more time searching together and were more likely to revisit the same areas, but it did allow pairs to coordinate quickly on object locations. Heat maps had significantly higher performance times compared to the control group and other gaze visualizations.	It is unknown whether the participants were randomly paired.

Table 1. Continued

Study	Collaboration	Technology	Study methods	Findings	Methodological limitations
Lee et al. (2017a)	Remote	Augmented reality	Comparison of (1) sharing eye gaze during augmented video conferencing against (2) a base line condition.	Subjective ratings indicated that local workers felt being able to express themselves significantly more clearly when their eye gaze was shared (p=0.031). Furthermore, remote experts and local workers were significantly more able to understand their partner's message clearly when the local's worker eye gaze was shared. Regarding co-presence, local worker's felt their partner's presence significantly more when the remote helper's eye gaze was shared. sharing of eye gaze resulted in significant better understanding of where partners were focussing on.	Relatively small sample size (N=8)
Wang et al. (2019)	Remote	Mixed reality	Comparison of two interfaces: (1) Pointer against (2) 2.5DHands.	There were no significant differences between the two interfaces on performance time. Subjective ratings showed that the 2.5D hands interface had a significant effect on guidance information (p=0.035 & p=0.011) and sense of co-presence awareness (p=0.015). Moreover, sharing gestures could drive local workers to concentrate better (p=0.015) and be more confident on the task (p=0.010). Furthermore, the 2.5D hands interface had a significant effect on expressing ideas (p=0.016 & p=0.008) and sense of being focussed (p=0.007).	Absence of a control condition. Relatively small sample size (N=12)
Higuch, Yonetani, & Sato (2016)	Remote	Shared Workspace, Video mediated system	Comparison of (1) sharing eye gaze against (2) visualized hand gestures.	Sharing eye gaze show faster (p =0.01, p = 0.05) and more precise pointing capability (p=0.01) than using hand gestures.	Small sample size (N=4) Participants could assign their first role themselves.

Table 1. Continued

Study	Collaboration	Technology	Study methods	Findings	Methodological limitations
Wang et al. (2020b)	Remote	Augmented reality	Comparison of (1) a cursor pointer, (2) a head pointer, and (3) visualized eye gaze in an AR system.	Using a cursor pointer was significantly faster than using eye gaze visualizations ($p < 0.001$). There was no significance difference in performance time between the cursor pointer condition and head pointer condition. There also were no significant differences in perceived workload across the three conditions for local workers and remote experts. Sharing eye-gaze and using a head pointer significantly improved the sense of co-presence ($p < 0.006$) and awareness of the user's attention ($p < 0.001$ & $p< 0.002$).	There was no practice trial, while some participants had experience with VR/AR Limited sample size (N=16)
Zhang et al. (2017)	Co-located	Shared Workspace	Comparison of four gaze indicators: (1) cursor (2) trajectory gaze (over time), (3) highlight points, and (4) spotlight.	The spotlight visualization was significantly faster compared to no gaze visualizations when a target was presented ($p < 0.05$). There were no significant differences between the other visualizations when targets were absent or presented. In the interview study, participants indicated that gaze information was useful when they needed to come to an agreement with their partner. However, they also indicated that the gaze information could be distractive.	Small sample size (N=8) Gaze measure outcomes were based on feedback and observations, which could be biased.
Yamashita et al (2011)	Remote	Shared Workspace	Comparison of (1) a tabletop vision including feedback and feedthrough against (2) a tablet vision that also provided local users with remote lags of the remote tabletop user.	Workers asked questions and confirmations less frequently in the remote lag conditions than the baseline condition ($p < 0.05$). On top of that, helpers in the remote lag conditions talked significantly less than in the baseline condition ($p < 0.05$). Workers also perceived significantly lower workload in the remote lag condition ($p < 0.05$). There was no significant difference in perceived mental effort for the helpers between the two conditions.	Demographic information was not reported.

Table 1. Continued

Study	Collaboration	Technology	Study methods	Findings	Methodological limitations
Fussell et al. (2003)	Co-located and remote	Shared Workspace, Video mediated system	Comparison of co-located collaboration and remote collaboration using (1) a video feed, and (2) video feed including cursor pointer	There was a significant main effect for how easily it was to identify referents on the media conditions ($p < 0.0001$). Both helpers and workers found it the easiest to identify references in the side-to-side condition and easier to identify references in the video with a cursor pointer condition compared to the video only condition. Moreover, performance was the fastest in the side-to-side condition.	Demographic information was not reported. Limited sample size (N=16) There is no estimate of variance (e.g., confidence intervals, standard errors).
Xiao and Ishii (2010)	Remote	Video mediated system	Comparison of three piano learning configurations: (1) shadows, (2) reflection, (3) organ.	The organ configuration helped novices to learn simple melodies more effectively than the hand shadow and reflection configurations	There is no clear description of the statistical analysis used. Relatively small sample size (N=10)
D'Angelo and Gergle (2016)	Co-located and remote	Shared Workspace	Comparison of 3 gaze configurations: (1) co-located with a mouse cursor, (2) remote with shared gaze, and (3) remote without shared gaze.	Pairs efficiently used deictic references and acknowledgements of behaviour when gaze is shared compared to when gaze is not shared. The gaze cursor was used to highlight areas of interest in the shared workspace	Limited sample size (N=18)
Nuernberger et al. (2016)	Remote	Augmented reality	Comparison of (1) an AR system that stabilizes 2D annotation in the task scene against (2) a traditional system	Participants tend to draw arrows (35.5%), circles (36,7%) and outlines (9,7%) more than other types of drawings. Arrows were used the most with action tasks and oblique angles to target objects. Moreover, their results showed that 76% of the Participants rated that the proposed method better conveys the user's intentions both in action and referencing task compared to the median depth plane interpretation.	There is no demographic information given for each of the between groups, only the demographic information of the whole sample. Some of the statistical analysis methods used were unknown.
Lee et al. (2018)	Remote	Mixed reality	Comparison of (1) dependent view against (2) independent view.	The independent view had a significantly higher rating of co-presence than the dependent view. There was no evidence found on for the effects of view independency on task performance.	Relatively small sample size (N=6)

Table 1. Continued

Study	Collaboration	Technology	Study methods	Findings	Methodological limitations
Yang et al. (2020)	Remote	Mixed reality	Comparison of visual cues: (1) head frustrum, and (2) hand gestures in combination with spatialized voice and spatialized auditory beacon cues.	Hand gesture cues did not significantly decrease participants' task completion time, but participants had significantly faster completion times with the remote expert's head frustrum. Both visual cues significantly improved the participants sense of co-presence compared to the control condition. However, the integration of hand gestures did not significantly improved sense of co-presence when the head frustrum was already provided.	Relatively small sample size (N=12)
	Remote	Mixed reality	Comparison of spatial cues: (1) non-spatial voice, (2) spatialized voice and (3) spatialized auditory beacons to a baseline condition.	There were no significant differences in task completion time between all the spatial cues conditions. While there was a generally higher rating of social presence in all the spatial cue conditions compared to non-spatial voice, there were no significant differences of social presence between the conditions. There were significant differences in spatial presence across the cue conditions	Relatively small sample size (N=12)
Ou, Oh, Yang & Fussell (2005b)	Remote	Shared Workspace	Comparison of task properties: (1) shading of puzzle pieces and (2) puzzle complexity.	When the puzzle pieces were harder to discriminate, helpers would spend more time gazing at the piece bay ($p < 0.0001$). When the puzzle was more complex, helpers would look less at the pieces bay when there were fewer and fewer remaining pieces ($p < 0.0001$).	Relatively small sample size (N=12) Demographic information was not reported.
Pauchet et al. (2007a)	Co-located and remote	Shared Workspace, Video mediated system	Comparison of (1) co-located tabletop collaboration against (2) remote tabletop collaboration including visual information.	There were no significant differences between co-located and remote collaboration in completion time, co-presence, and piece manipulation	Precise definitions of the outcome measures were missing. Relatively small sample size (N=12)

Table 1. Continued

Study	Collaboration	Technology	Study methods	Findings	Methodological limitations
Sauppé and Mutlu (2014)	Co-located	Shared Workspace	Comparison of three collaborative settings: (1) negation, (2) instruction, and (3) coordination.	In the negotiation task, gazing toward the partner was a significant predictor of empathy between participants (p = 0.036). In the instruction task, deictic gestures were used to disambiguate locations and objects by the instructor (p = 0.016) and the learner (p = 0.061). Predictors of task success included learners who performed head nods as non-verbal feedback (p = 0.015). In the cooperation task, pointing gestures were used to emphasize particular concepts (p < 0.001)	The regression-based approach used does not offer insight into the temporal characteristics of social cue behaviours and their implications for collaboration.
Argelaguet et al. (2010)	Remote	Mixed reality	Comparison of show-trough techniques (1) cutaway, (2) transparency.	There was no significant effect of technique on retrieval time of the target object. A significant effect of technique was found on the distance between users (p < 0.001).	Lack of demographic information reported (only age range was reported). Small sample size (N=3)
Lee et al. (2017b)	Remote	Mixed reality	Comparison of two view-awareness cues: (1) view frame rectangle, and (2) view frame rectangle with arrow .	User ratings indicated a significant difference in user ratings for the different view awareness cues (p=0.0003). The view frame rectangle with arrow was rated the highest (M =6.75), the view frame rectangle without the arrow was rated moderately easy (M= 4.625). Without a view awareness cue was rated the lowest (M = 2.5).	There was no specific task. The user study is rather a demonstration. Conditions were not counter balanced. Small sample size (N=4)
Gauglitz et al. (2014)	Remote	Shared Workspace	Qualitative User study the interface for drawings annotations.	Observations indicated that users used arrow-like drawings for indicating directions and rotations.	The majority of interview questions are unknown. It lacks a method for analysing the verbal data. No verification procedures were evident.

Table 1. Continued

Study	Collaboration	Technology	Study methods	Findings	Methodological limitations
Lissermann et al. (2014)	Co-located	Shared Workspace	Comparison of (1) the Permulin tabletop system against (2) a split-screen tabletop system.	During tight collaboration, there were no significant differences in the amount of awareness cues participants perceived between the two conditions. However, the awareness cues increased significantly from loose toward tight coupling with the Permulin system ($p < 0.001 – 0.05$). Furthermore, there was a significant difference in perceived awareness cues during loose collaboration between the conditions ($p < 0.001$), whereas Permulin generated the least awareness cues for both tasks.	Conditions were not counter balanced. No practice trials. Demographic information was not reported. Relatively small sample size (N=10)
Pauchet et al. (2007b)	Co-located and remote	Shared workspace	Comparison of (1) co-located, and (2) remote use of Digitable.	While comparing the four remote configurations with the co-located configuration, participants completed the mosaic task significant faster with the remote configurations ($p < 0.046$). Furthermore, the remote side-by-side condition with gesture visualization is significantly faster than the co-located side-by-side condition ($p < 0.033$) and also significantly faster than the co-located face-to-face condition ($p < 0.020$).	Demographic information was not reported. The questionnaire items were unknown. Limited sample size (N=15)
Wang et al. (2020c)	Remote	Mixed reality	Comparison of (1) 3DAM (system that supports sharing 3D CAD models) against (2) 3DGAM (system that supports sharing 3D gesture and CAD models.	For the local workers, sharing 3DGAM cues could improve the sense of co-presence ($p = 0.0026$), working together ($p = 0.008$), my attention to the partner ($p = 0.005$), and the understanding of instructions ($p = 0.020$). Concerning the remote experts, sharing 3DGAM cues could improve the users' focus ($p = 0.009$), confidence ($p = 0.033$), the ability to provide clear instructions ($p = 0.026$) and assistance ($p = 0.020$).	It is unknown whether participants were randomly assigned to a role. Participants did not swap roles between conditions. Limited sample size (N=14)

Table 1. Continued

Study	Collaboration	Technology	Study methods	Findings	Methodological limitations
Olwal et al. (2011)	Remote	Shared Workspace	Interviews, observations, and a qualitative user study.	Precise pointing may (1) save time during discussions, (2) include remote sites, and (3) make references more explicit for less experienced participants. Annotations could be used for highlighting targets. Furthermore, it could help to clarify the discussion further as the augmented information may make the discussed information easier and faster to understand	The analytic methods were not fully described. The questionnaire data was not analysed using a statistical method. No verification procedures were evident.
Ou et al. (2008)	Remote	Shared Workspace	Evaluation of a conditional Markov model to predict the user's focal of attention based on verbal descriptive patterns in a remote collaborative.	The overall accuracy of the model was 65,40% for solid colour puzzles and 74,25% for shaded colour puzzles	Demographic information was not reported. . Relatively small sample size (N=12)
Le Chénéchal et al. (2016)	Remote	Mixed reality	Comparison of (1) Vishnu paradigm interface against (2) a basic desktop screen.	There was a significant difference in task completion time between both interfaces ($p = 0.047$). The Vishnu paradigm interface did not allow for faster task completion times in a simple random target acquisition task. However, it did allow for significant faster task completion times in a more complex layout of regular and dense targets ($p < 10−9$).	The analytic methods were not reported. Relatively small sample size (N=11) The experimenter participated as the expert during the task.
Boucher et al. (2012)	Co-located	Shared Workspace	Comparison of (1) gaze awareness against (2) a baseline condition.	There was a significant effect in the location phase of the target object when the informer was wearing sunglasses ($p < 0.001$).	Relatively small sample size. In total 6 participants participated in the study. Demographic variables were not reported.

Table 1. Continued

Study	Collaboration	Technology	Study methods	Findings	Methodological limitations
Gao et al. (2016)	Remote	Mixed reality	Comparison of a (1) 3D oriented view against (2) a static front view	There was no significant difference found in usability ratings between the oriented view and the static front view.	The analytic methods were not reported. Relatively small sample size (N=10) There is no estimate of variance (e.g., confidence intervals, standard errors).
Huang, Alem and Tecchia (2013)	Remote	Mixed reality	Evaluation of a 3D gesture-based interaction paradigm.	Helpers gave the same ratings for being able to perform both representational and pointing gestures. Furthermore, workers seemed to perceive pointing gestures more easily than representational gestures.	Relatively small sample size (N=7). There is no estimate of variance (e.g., confidence intervals, standard errors). Demographic information was not reported.
Kunz, Nescher and Küchler (2010)	Co-located and remote	Shared Workspace, Video mediated system	Comparison of two interfaces: (1) CollaBoard system, (2) standard video conference software against (3) a baseline condition.	The video conferencing condition performed significantly worse than the co-located condition for all questions and completion time. Furthermore, the video conferencing condition is rated lower on the usability ratings and had slower task completion time compared to the Collaboard condition.	Conditions were not counter balanced. The analytic methods were not reported. Relatively small sample size (N=7)
Novick, Hansen and Ward (1996)	Co-located	Shared Workspace	Observations	Nearly 42% of turn exchanges followed the mutual break pattern where gaze is momentarily mutual, after which the other person breaks the mutual gaze and begin to speak. Furthermore, 29% of the turn exchanges followed a “mutual-hold” pattern where the turn recipient begins speaking without immediately looking away. Moreover, the mutual-hold pattern positively correlated with the number of turns taken.	No verification procedures were evident. Small sample size (N=4)

Table 1. Continued

Study	Collaboration	Technology	Study methods	Findings	Methodological limitations
Kim et al. (2019)	Remote	Mixed reality	Comparisons of 3 visual cues combinations: (1) hand gestures and a pointer, (2) hand gestures, sketching and hand gestures plus pointing, and (3) sketching compared to a gesture only condition.	Adding sketch cues to hand gesture cues helped participants to complete the task significantly faster than hand gestures only. However, adding the pointer cue to the hand gesture cue led not to faster completion times. Participants were significantly faster completing the tangram task with an added sketch cue than with an added pointer cue. Subjective ratings indicated that there was no significant difference in sense of feeling together between the visual cues condition. Adding the pointer cue to the hand gesture cue significantly increased both the local worker's and remote expert's feeling of required mental effort.	Relatively small sample size (N=8)
Jakobsen and Hornbæk (2014)	Co-located	Shared Workspace	Observations	Verbal communication differed by proximity (p < 0.001, Cramer's V = 0.09). Participants were observed more frequently talking and less frequently silent when they were physically close. Moreover, visual attention also differed by proximity (p < 0.001, Cramer's V = 0.32). Observation indicated that participants were looking at the same area more often when they were physically close, and less frequently when they were not physically close. Vice versa, participants were more frequently looking at different areas when they were not physically close, and less frequently when they were physically closely.	Absence of a control condition. There was only one task for one scenario, which is not enough to investigate how groups collaborate around a wall display. Limited sample size (N=15)
Jay, Glencross and Hubbold (2007)	Co-located	Shared Workspace	Comparison of different values of latency.	Latency affected visual feedback from 50 msec and haptic task performance from 25 msec. However, the rates of error slowed considerably after latencies of 100 msec.	Limited sample size (N=15)

4.1.3. How gaze awareness in CSCW systems facilitates remote collaborative work processes

According to the included literature, several CSCW systems have been developed and studied over the past years that enabled the ability to perceive non-verbal gaze cues (Bai et al., 2020; D'Angelo & Begel, 2017; D'Angelo & Gergle, 2018; Higuch, Yonetani, & Sato, 2016; Kütt et al., 2020; Lee et al., 2017a; Lee et al., 2017b; Lee et al., 2018; Müller et al., 2013; Ou et al., 2005a; Piumsomboon et al., 2019; Wang et al., 2020b; Yang et al., 2020; Zhang et al., 2017).

According to the findings of the studies mentioned above and the thirds research question "how do social cues support object-related remote collaborative work processes?", the addition of shared gaze supports object-related remote collaborative work processes in the following ways.

Firstly, gaze awareness facilitates communication, as it enables the use of deictic references. Listeners can comprehend these references in combination with the shared speaker's gaze to identify objects of interest. Consequently, remote collaborators can identify targets faster when gaze of their partner that is attending the target is visualized (D'Angelo & Begel, 2017; Piumsomboon et al., 2019). Without such awareness, it could be difficult to avoid ambiguity in remote collaboration. Therefore, gaze awareness also facilitates conversational grounding (D'Angelo & Gergle, 2016; D'Angelo & Gergle, 2017; Gupta, Lee & Billinghurst, 2016; Lee et al., 2017a; Higuch, Yonetani, & Sato, 2016; Müller et al., 2013; Piumsomboon et al., 2018; Wang et al., 2020b). For instance, collaborators can assure that their instructions were understood in the context of previous interactions and shared vocabulary by seeking additional visual evidence of understanding by monitoring their partner's gaze (Ou et al., 2005a; Ou, Oh, Yang & Fussell, 2005b). Thus, gaze awareness also promotes performance monitoring by observing their partner's attention.

Overall, this corresponds to how joint referencing works in two directions (Woods & Hollnagel, 2006). In one direction a collaborator can signal their partners by "referring to something with the intent of directing another's attention to it" (Bruner, 1986, as cited in Woods & Hollnagel, 2006, p. 91). In the other direction, collaborators could perceive where and to what another is directing their attention, without requiring the need to communicate with each other. To coordinate these processes in CSCW systems, it is important that agents can assess the task through shared representations of the monitored process and interact with it (Woods & Hollnagel, 2006).

4.1.4. How shared gestures in CSCW systems facilitate remote collaborative work processes

According to the included literature several CSCW systems have been developed that enable the ability to share gestures (Bai, Sasikumar, Yang & Billinghurst, 2020; Cabibihan, So, Saj and Zhang, 2012; Fussell et al., 2004; Higuch, Yonetani, & Sato, 2016; Piumsomboon et al., 2018; Wang et al., 2019; Wang et al., 2020c).

According to the findings of these studies and the third research question: “how do social cues support object-related remote collaborative work processes?”, the addition of shared gestures supports object-related remote collaborative work processes in the following ways.

Firstly, sharing pointing gestures enables collaborators to re-direct a partner’s attention for identifying objects of interest in the shared workspace. Similar to sharing gaze cues, this promotes the efficiency of communication because it enables the use of deictic references in speech. Secondly, pointing gestures may also enhance the collaborators’ spatial representation when co-occurring speech is ambiguous (Cabibihan, So, Saj and Zhang, 2012; Piumsomboon et al., 2018; Fussell et al., 2004) Thirdly, sharing representation gestures enables collaborators to convey object-oriented information. This information could indicate specific actions that partners need to perform. For instance, collaborators could use hand gestures to illustrate the exact angle of insertion of a specific target object. Because these actions can be visualized with hand gestures, less verbal feedback is required to describe those actions. Consequently, instructors feel able to better express themselves and provide instructions more clearly and their partners gain a better understanding of the instructions as the visualized actions tackle the ambiguities that arise during verbal feedback (Bai, Sasikumar, Yang & Billinghurst, 2020; Fussell et al., 2004; Higuch, Yonetani, & Sato, 2016; Yamashita, Kaji, Kuzuoka and Hirata, 2011; Wang et al., 2019; Wang et al., 2020c).

4.1.5. Visualizing the relationships between the social cues level and the work processes level in the abstraction hierarchy

The evidence mentioned in section 4.1.3 & 4.1.4 respectively has been documented in Table 2. Based on this, the preliminary abstraction hierarchy has been modified to visualize the relationships between the work processes of remote collaboration with the non-verbal gesture and gaze cues (see Figure 7). Every relationship is made explicit with a red line connecting the processes and social cues. The thicker the line, the more evidence (see Table

2) has been found that implementing the perception of the specific non-verbal cue has a significant effect on the work process ($p < 0.05$).

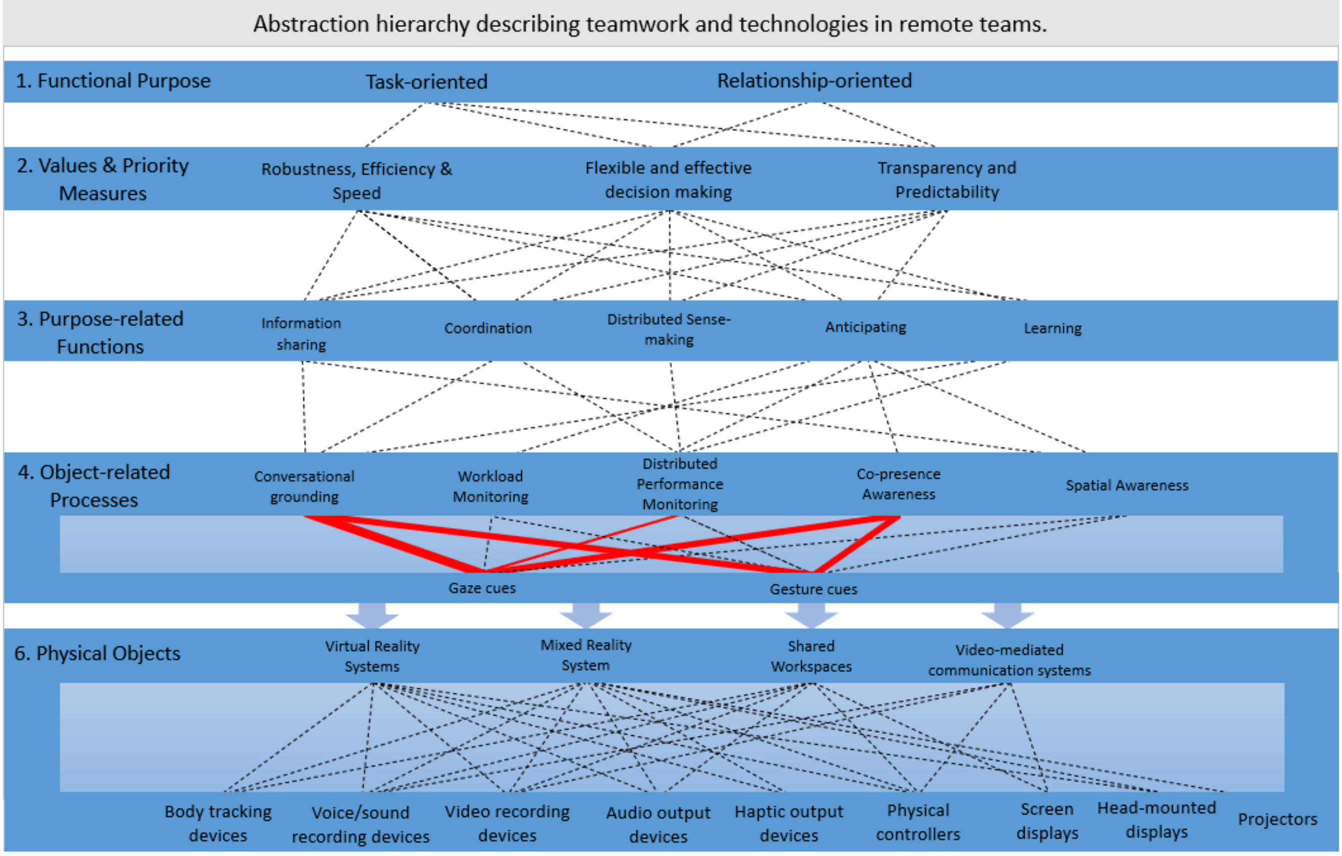


Figure 7. Established abstraction hierarchy consisting of the functional purpose, values and priority measures, purpose-related functions, object-related processes including the social cues that could be perceived and the work processes of remote collaboration and the physical objects needed to realize the collaboration.

Table 2. *Significant results including effect sizes of the included literature categorized by social cue, work process and statistical method*

Study	Social Cue	Work process	Statistical method	Effect sizes
Bai, Sasikumar, Yang & Billinghurst (2020)	Gaze	Co-presence awareness	Pairwise comparisons	0.692
Gupta, Lee & Billinghurst (2016)	Gaze	Co-presence awareness	Two way repeated measures ANOVA	0.217
Wang et al. (2020b)	Gaze	Co-presence awareness	Friedman test	0.301
Lee et al. (2017b)	Gaze	Co-presence awareness	Repeated-measure ANOVA	0.387
Yang et al. (2020)	Gaze	Co-presence awareness	Friedman tests & Kendall’s W tests	0.146, 0.104 & 0.144
Gupta, Lee & Billinghurst (2016)	Gaze	Conversational grounding	Two way repeated measures ANOVA	0.336
Lee et al. (2017a)	Gaze	Conversational grounding	Repeated-measure ANOVA	0.387
D'Angelo and Begel (2017)	Gaze	Conversational grounding	One-way ANOVA	0.308

Table 2. Continued

Study	Social Cue	Work process	Statistical method	Effect sizes
D’Angelo & Gergle (2016)	Gaze	Conversational grounding	Mixed model regression	0.217
Müller, Helmert, Pannasch & Velichkovsky (2013)	Gaze	Conversational grounding	Factorial repeated-measures ANOVA	0.105
Higuch, Yonetani, & Sato (2016)	Gaze	Conversational grounding	Repeated-measures mixed model ANOVA	0.978
Ou et al. (2005a)	Gaze	Performance Monitoring	Repeated measures-mixed model ANOVA	0.132
Ou, Oh, Yang and Fussell (2005b)	Gaze	Performance Monitoring	Factorial repeated-measures ANOVA	0.132
Bai, Sasikumar, Yang & Billinghurst (2020)	Gesture	Co-presence awareness	Pairwise comparisons	0.692
Piumsomboon et al. (2018)	Gesture	Co-presence awareness	Wilcoxon Signed Rank test	0.885
Wang et al. (2019)	Gesture	Co-presence awareness	Wilcoxon Signed Rank test	0.710
Wang et al. (2020c)	Gesture	Co-presence awareness	Wilcoxon Signed Rank test	0.654

Table 2. Continued

Study	Social Cue	Work process	Statistical method	Effect sizes
Yang et al. (2020)	Gesture	Co-presence awareness	Friedman tests & Kendall's W tests	0.146, 0.104 & 0.144
Yamashita, Kaji, Kuzuoka & Hirata (2011)	Gesture	Conversational grounding	Repeated-measures mixed model ANOVA	0.204
Wang et al. (2020c)	Gesture	Conversational grounding	Wilcoxon Signed Rank test	0.654
Wang et al. (2019)	Gesture	Conversational grounding	Factorial repeated-measures ANOVA	0.710
Fussell et al. (2004)	Gesture	Conversational grounding	Wilcoxon Signed Rank test	0.216
Piumsomboon et al. (2018)	Gesture	Conversational grounding	Wilcoxon Signed Rank test	0.885

5. Case Study Social XR Demo ‘VRComm’

Nowadays, computer-supported collaborative work (CSCW) tools such as Microsoft teams and Zoom are used to support remote collaborative work at a distance. However, these video conferencing applications have limitations regarding the perception of social cues. As mentioned in the literature study, many CSCW tools have been developed to enable the perception of various social cues to better support remote collaboration at a distance. Even though these systems make it possible to perceive several social cues of others, it is still questionable to what extent these implementations support remote collaborative work. In this case study, we evaluate VRComm: An end-to-end web system that enables remote communication in virtual environments with real-time photorealistic user representations and imitated eye representations for visualizing the user’s eye gaze (Gunkel et al., 2021). Based on the processes represented in the abstraction hierarchy, VRComm is assessed to what extent the system promotes remote collaboration. Therefore, we try to identify to what extent the technical capabilities of the VR system support the processes of remote collaboration. Furthermore, it will be discussed whether the abstraction hierarchy is an useful framework for identifying the strengths and weaknesses of CSCW tools. In the next section, we present the technical capabilities of VRComm and give an impression of how this manifests itself in the virtual environment. Afterwards, we evaluate how the VR system supports the processes of conversational grounding, workload monitoring, distributed performance monitoring, co-presence awareness and spatial awareness.

5.1. VRComm specifications

VRComm enables real-time remote communications via video conferencing (Gunkel et al., 2021). The system combines video conferencing technology with social VR capabilities that allows to capture, process, transmit and render social cues such as eye gaze, body posture, hand gesturing, and head nodding. Therefore, it differs from traditional videoconferencing in terms of their affordability to transfer social context information. The purpose of the system is to convey remote interactions in virtual environments that allow users to perceive the social cues of others and to understand the environment and objects surrounding the user. In turn, it is theorized that these implementations should enable more natural interactions and communications to increase the feeling of co-presence.

The system utilizes a camera-based capture approach to capture the scene and users in real time. Two cameras are aligned diagonally left and right from the user to capture near 180° (front view) photorealistic representations of the user in real-time. Furthermore, a foreground-background module is used to improve image quality and to extract the background in real-time. The 180° photorealistic representations are rendered in real time into the virtual scene as point clouds. Additionally, eye gaze is captured and projected in real-time as “imitated eyes” into the photorealistic user representations (see Figure 8). Furthermore, spatialised audio has been implemented by capturing and uniquely addressing each of the user’s audio streams and rendering it at the appropriate spatial location.

Figure 8. Googly eyes are visualized onto the head-mounted display of the user represented as a point cloud in the virtual space

Figure 9 shows the layout of the virtual environment that represents a traditional meeting room. In the middle of the room, a large table is visualized. Two chairs are positioned on each long side of the table and one chair is positioned at the head of the table. Furthermore, one side of the room consists of windows that displays a broad natural landscape.

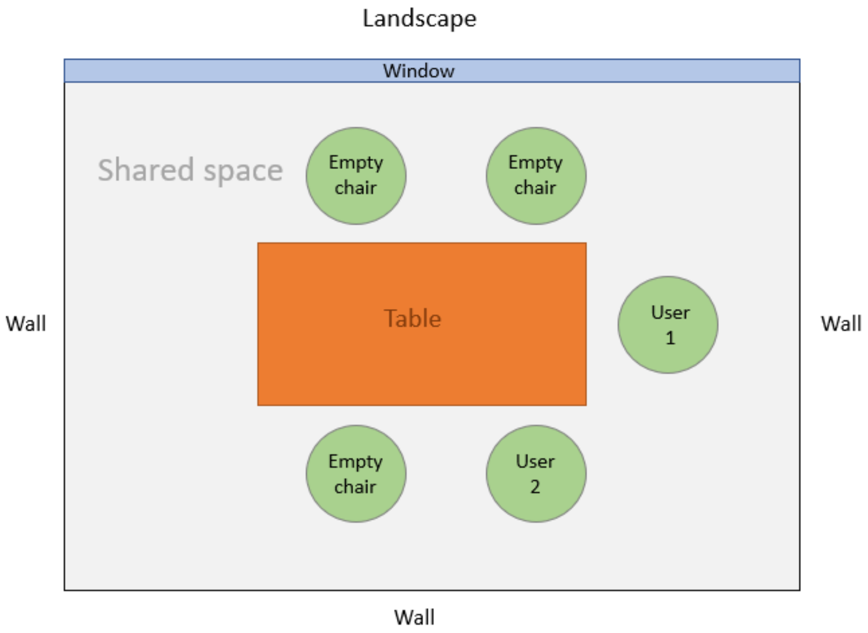


Figure 9. Layout of the shared space of VRComm

5.2. Demo evaluation set-up

To assess VRComm, two users were projected into the virtual scene. One user was positioned at the head of the table, while the other was positioned left to him on the long side of the table. Due to the use of noise-cancelling headsets, the users could only hear each other in VR, despite being in the same room with each other. At the time both users were experiencing VR, they discussed to what extent VRComm supports the processes of conversational grounding, workload monitoring, distributed performance monitoring, co-presence awareness, and spatial awareness. On the basis of the main points of the experiences of the users expressed through the conversation, the question as to whether VRComm promotes certain processes will be answered.

5.3. Conversational grounding

VRComm enables the perception of the partner’s gaze direction and gestures. By means of eye tracking, gaze directions are visualized by imitation eyes that are displayed on the users’ head-mounted display in the VR space. Furthermore, the photo-realistic representations in the shared space allows to perceive gestures of partners. While examining VRComm, we tested whether it was possible for the partner to identify objects by merely pointing or looking at them, which was successful. This implies that VRComm promotes conversational grounding during remote collaboration through the implementation of the social cues mentioned above.

However, there are some limitations considering that it is important that you can also see how partners interact with target objects in a shared space. First, it is not possible to move yourself in the virtual space with the current demo. The boundaries of the spaces are determined by the cameras set up at the edge of the space, meaning that you cannot move past them without virtually leaving the shared space. Consequently, the user does not have much space to move around in VR. Suppose that you want to give a presentation supported by visual information (e.g., PowerPoint) in VR, the presenter has to be projected near the PowerPoint slides if they wants to point to visual information up-close.

Secondly, you are not projected in full detail in VR. Suppose someone has to build a certain object and they are projected in VR with the components of the object at hand. Partners are able to infer that someone is manipulating parts, but they have limited vision regarding the details of the object. This also goes in combination that one cannot get closer to see the parts because they are restricted to the space of the VR set-up. Imagine you're observing someone at a distance who is working on their computer. You can tell that they are typing because you see them pressing the keyboard, but you cannot tell from that distance what they are typing and in which program or browser they are typing. This way of inferring is similar to VRComm. Being able to tell what someone is doing but missing out on the details makes it more difficult for the builder as well as their partners to perceive and refer to specific details of the parts. Moreover, lack of detail also complicates the process of giving and receiving instructions as partners/collaborators.

5.4. Workload Monitoring

VRComm makes it possible to observe the others' postures. This could provide information about the person's workload status. For example, a slumped posture could be an indicator that someone is overtired (Huron, 2018). However, the perception of other non-verbal stress or fatigue indicators like facial expressions is somewhat limited because users are projected with a head-mounted display in the virtual space. For example, you cannot infer if your partner is tired because you can only see their imitated eyes and not their actual eyes. Furthermore, the aforementioned lack of details also limits the number of stress indicators that can be observed. For example, you may be able to determine when someone is shaking, but it is difficult to infer if someone is sweating or not.

5.5. Distributed performance monitoring

Considering distributed performance monitoring in VRComm, it is possible to remotely monitor the work processes in VR. However, this is limited by the lack of details and the distance between each other in VR. Because of this, more dialogue is needed with collaborators to make up for these deficits. Moreover, the ability of VR users to only move within the borders of the VR set-up is a limitation. For example, in face-to-face settings, floor managers are able to move around the entire work floor. They are able to up-close to monitor the work processes carried out by employees. Currently, this not possible within VR. However, it is possible to observe the current view of collaborators outside the VR setting on a desktop monitor. However, it seems logical that this possibly makes collaborators feel uncomfortable while executing the task. Especially when there is communication with a VR-user and work process observer outside of VR. From my own experience this feels like someone is inside my head, who can see what I can see and could communicate with me just like my own thoughts. This could be experienced as someone violating your privacy.

5.6. Co-presence awareness

Considering the sensory properties of VRComm, users are able to perceive each other’s physical appearance, gaze directions, posture, and gestures. As mentioned in the literature study, gaze awareness and visualized gestures improve the sense of co-presence during remote collaboration. This corresponds to my own experiences with VRComm. The intimated eyes made me feel like another person was looking at me during conversations, while the ability of using gestures (e.g., to point to certain target objects) was part of the interaction with each other. It is not the case that performing gestures benefits the feeling of togetherness. However, it is the case that performing gestures could reinforce the sense of co-presence, because perceived gestures display an interaction with one other. Furthermore, according to Jo, Kim & Kim (2016), behavioural cues as gestures and gaze, play a more significant role in the feeling of togetherness and effective communication than visual details do. In their study, they found that there were no significant differences in perceived co-presence between photo-realistic representations of users and pre-built 3D avatars. Consequently, imitated eyes may not limit perceived co-presence compared to realistic eye representations of users. Interestingly, Jo, Kim and Kim (2016) found that realistic backgrounds in virtual environments benefit the perceived co-presence, as realistic backgrounds contribute to an enhanced information transfer and understandability.

According to Podkosova and Kaufmann (2018), proximity also contributes to the experience of embodied co-presence. As mentioned, co-presence is based on sensory awareness of another individual in VR. In VR, users tend to maintain conventional interpersonal distance, which implies that the users are aware of each other's presence in the same space (Podkosova & Kaufmann, 2018). Moreover, co-placed users perceive more distributed local users as ‘being co-present’ to a lesser extent than more co-located co-users. Moreover, co-placed users are able to perceive a higher amount of distributed local users as being co-present. This also may have contributed to my own feeling of togetherness during the demo, as the others were seated relatively close to me. If we both held out our arms, our arm representations would collide in VR.

5.7. Spatial awareness

VRComm allows users to be typically projected in a virtual immersive environment. It also allows physical head rotations of the user to control their view in VR, which allows users to maintain spatial orientation in the 3D environment. 3D environments that surround the user in which users can orientate themselves can create a sense of direction and space. According to Jones (1993), the sense of space arises from recontextualizing virtual audio, which is also allowed by VRComm. So, in order to reconstruct external sound in the environment to correlate with the known constructs in the environment, users could might ask themselves: What is making this sound and where does this sound come from? So, if someone who is seated on your right in VR, you could perceive the talking sound also coming from the right-side. Thus, VRComm allows users to perceive the direction of the sounds as well as orientate themselves in the environment with head rotations to identify what is making the sound. Consequently, creating a sense of direction and space.

Furthermore, in the current 3D environment of VRComm, there are windows on one side of the room that give a broad nature view, “you can look far over the trees and see mountains on the horizon”. From my own experience, I was standing close to the edge of this window, which gave me feelings of fear of heights when I was looking down. Feelings of anxiety and fear of heights could only be experienced when someone actually perceives themselves confronted with height (Wuehrt et al., 2019). Therefore, VRComm succeeds in making users experience spatial distances between virtual stimuli as if users are actually in the environment. This relates to the concept of presence. Presence in VR refers to the illusion of being there, not to one’s belief that the virtual environment that they are perceiving is real

(Slater, 2018). So even though you know that the virtual environment is not real, it does not change your perception of the environment and how you respond to it.

Proxemics in real environments largely extend to walking in immersive VR (Podkosova & Kaufmann, 2018; Williamson, Vinayagamoorthy, Shamma, & Cesar, 2021). Consequently, users experience feelings of discomfort when their personal space is violated in VR. When a virtual agent enters their personal space, users tend to move away. This was also experienced while exploring VRComm. During the calibration of the user positioning, a user was projected within 1 meter of another user which was experienced as not pleasant for both the users.

6. Discussion

Sharing gaze and gesture cues are currently the most relevant non-verbal behaviours implemented in CSCW tools. Sharing gaze cues enables collaborators to be aware of their partner's attention and to re-direct their attention and it enables the use of deictic references which consequently improves the efficiency of communication. Overall, these capabilities contribute to allowing collaborators to better express themselves and understand each other. Furthermore, gaze cues provide the ability to monitor their partner's attention and provides a sense of togetherness. Based on the evidence of the articles included in the systematic synthesis, the relationships between both gaze and gesture cues and conversational grounding is the most prominent, followed by the relationships between gaze and gesture cues along with co-presence awareness.

Similar to gaze cues, shared pointing cues also enable collaborators to re-direct their partner's attention and use deictic references. Moreover, representational gestures could convey oriented information that is necessary to perform certain actions properly. Overall, collaborators could better express themselves and understand each other better. Therefore, sharing gesture cues also facilitates conversational grounding. Furthermore, shared gesture cues demonstrate an interaction. As a result, there is also a relationship between shared gestures and a sense of togetherness. Lastly, performance could also be perceived by observing the partner's gestures during an object manipulation task. For instance, collaborators could perceive if their partner is correctly rotating specific parts of a target object. When no certain actions in a direction relative to target objects are needed to be performed then representational gestures will not benefit distributed performance monitoring.

Therefore, the perception of non-verbal cues and their influences on the work processes of remote collaboration is dependent on the characteristic of specific tasks. In the following section, the characteristics of tasks will first be made explicit. Afterwards, based on the 'in-principle' evaluation of VRComm, the strengths and areas of improvement of VRComm will be presented and discussed. Subsequently, the recommendations for VRComm are presented.

6.1. How gaze and gesture cues support certain task characteristics

Firstly, sharing non-verbal gaze and gesture cues could benefit tasks characterized by a local worker who is instructed by a remote helper. When the worker's gaze is shared, the remote helper is able to perceive what his/her partner is attending to. Therefore, the helper can re-direct their partner's attention when the partner is not focused on the object of interest. Moreover, the visual cues could provide evidence that previous instructions were understood. When these gaze cues are not in line with the previously given instructions, the helper can intervene and provide additional instructions to establish common ground. Furthermore, when the worker's gestures are visualized, helpers can perceive if they are performing the correct actions needed to be performed with the hands. If not, they can intervene and provide additional instructions and could establish common ground based on the evidence provided by the visual cues. Furthermore, when the worker's gestures are visualized, helpers can perceive if they are performing the correct actions needed to be performed with the hands. If not, they can intervene and provide additional instructions to establish common ground. In addition, when the helper's gestures are visualized into the workspace of the remote worker, he/she can intimate the actions needed to be performed by the local worker. As a result, the actions needed to be performed do not need to be explicitly described. In other words, seeing is learning. Furthermore, when workers need to identify a target object, they can attend to the target object in workspace and ask the helper if they are attending the correct object. For instance, they might ask: Is this the object needed?. The same goes for sharing the worker's gestures, the worker can use pointing gestures to identify the object of interest. Thus, sharing the worker gaze and gestures makes communication more efficient as it is possible to use deictic references. This also applies the other way around in the same manner, thus when the helper's gaze or gestures are visualized into the task space of the local worker.

Secondly, sharing gaze cues could benefit tasks characterized by collaborators who are working in a shared workspace. As mentioned, they have an indication of where their partners are attending to by observing each other's gaze. They can use this information as additional evidence of whether previous instructions were understood correctly. Furthermore, gaze awareness could make communication more efficient because it enables the use of deictic references. Collaborators could also adopt certain strategies when they are aware of each other's gaze. For instance, remotely located partners who are checking a code script for mistakes can simultaneously scan different blocks of coding. If one sees that his/her partner is scanning certain functions that are linked to each other, the other can scan another block of code. While comparing to both programmers checking the whole code simultaneously, such strategies could increase the task efficiency.

6.2. Strengths and areas of improvement of VRComm

We assessed the Social XR demo "VRComm" based on the processes presented in the abstraction hierarchy. We first described the processes and linked our own experiences as a measure for evaluation. Consequently, strengths and areas of improvement of VRComm were identified.

Firstly, the biggest strength of VRComm is that it enables participants to perceive each other's social cues such as posture, gaze directions and gestures. These technological capabilities of VRComm demonstrated to benefit the processes of remote collaboration, where perceiving each other's gestures and gaze directions plays an important role for conversational grounding and co-presence awareness. As such, imitated eyes seem to be useful representations to promote these processes. Another strength of the VRComm is that users are projected in VR which enables collaborators to perceive how others manipulate objects. Considering distributed performance monitoring, this allows collaborators to intervene when they perceive incorrect actions in a direction relative to target objects are being performed. Furthermore, the realistic representations of the users also allow collaborators to perceive each other's posture or user's intensity in performing actions. With this information at hand, users can estimate if someone is experiencing high workload. Consequently, actions can be taken to prevent potential errors in sub-tasks due to a high workload or fatigue. The spatialized audio is also a strength of VRComm because it promotes spatial awareness through spatialized audio that allows users to identify what is making the sound and where it is coming from. Moreover, the virtual environment presented in VRComm enabled the experience of feelings of fear of heights. This implies that the virtual stimuli of the environment succeeded in making users experience the spatial distances as if they were actually in the virtual environment, even though the users' are aware that the virtual environment is not real. Lastly, the realistic background presented also allowed for a better understandability and an enhanced information transfer which further improves co-presence awareness.

As for areas of improvement of VRComm, users are not able to walk in the virtual environment as their current movements are limited by the area of the camera set-up that projects the user in VR. This is considered a drawback as users are not allowed to keep their own interpersonal distances with one other due to lack of space. Consequently, VRComm do not allow richness and flexibility during verbal communication. Furthermore, collaborators are not able to look at each other actions up-close, which limits distributed performance monitoring, because users can only monitor each other in the virtual environment from the user position which they are projected in VR. Moreover, the lack of details of users and real-world items projected in VR currently also limits performance monitoring, as smaller details of target objects may not properly be observed. As a result, wrong actions with target objects may be overlooked which may lead to errors in a sub-task. This also limits conversational grounding, as more dialogue is needed to avoid ambiguity as a consequence of uncertainty. For example, if the others are able to perceive the targets properly. Another consequence of lack of details in projected users is the inability to perceive physical signs that the collaborator is stressed or fatigued.

6.3. VRComm Recommendations

On the basis of my own experiences and the identified areas of improvement, I recommended to develop the VRComm set-up in such a way that users can move naturally in the virtual environment by for example increasing the movement spaces between the cameras to match the virtual environment, updating the user position in VR in such a way that even with a small space between the camera set-ups the users are not limited in the distances they want to travel in VR, or by using a treadmill that can mimic natural movement without letting users move in the in the real world. In turn, this will benefit spatial awareness and co-presence awareness. It may also benefit conversational grounding as users will be able to walk to target objects. In addition, VRComm can benefit the processes of conversational grounding, distributed performance monitoring and workload monitoring if the projections are less noisy in VR. Thus, it would be beneficial to develop the technological aspects of VRComm so that more details can be observed. Furthermore, to avoid feelings of discomfort during communication outside of VR, it is recommended to use a microphone, where the outside input is played back to a sound system that is visualized in VR. This goes well in combination with the spatialized audio, so you no longer get the feeling that someone is talking to you in your head. Lastly, although the imitated eyes benefit the perception of gaze directions and positively influence co-presence awareness, actual user eye representations can provide additional information for collaborators to perceive fatigue.

6.4. Limitations

The current study offers insights into the role of social cues in remote collaboration, although the study presented here has some limitations. First, besides gaze and gesture, there are other social cues like posture and facial expressions that were not included in the initial search string. The reason for this was that including these terms in the search string did not lead to more hits that fit in the qualitative synthesis. A possible cause could be that Scopus may not bring all the articles to the surface that cover these concepts. In the future, it may be better to include more scientific databases that are also sufficient for replicating the current research. Another possible cause is that Scopus only searches for the terms included in the search string in the abstract, keywords and title. Consequently, relevant studies that used other terms in their abstract, keywords and title may not have come under scrutiny.

Secondly, the majority of the papers discussed teamwork in terms of pairs working together. As a result, no conclusion can be drawn on how shared social cues influence the work processes in which groups larger than two collaborators must work together. Furthermore, the relationships established in the abstraction hierarchy are based on the research results reported in the literature. When no effect sizes were reported, the effect sizes for the results to be statistically significant were calculated. In total, 21 required effect sizes have been calculated. Additionally, we classified the effect sizes: small [0.1,0.3], medium [0.3,0.5], large [0.8,1]. It has been found that a 13 results have a small effect size. Therefore, it has to be considered that there is some uncertainty in relationships between the social cues and the work processes. Finally, several questions for rating the articles on quality were highly subjective. The articles have been assessed by one researcher only. Therefore, the outcomes are prone to bias.

6.5. Lessons learned about the abstraction hierarchy

The abstraction hierarchy developed from the literature study demonstrated to be a useful framework for identifying the strengths and points for improvement of the social XR demo. By evaluating the demo based on the process layer of the abstraction hierarchy, we judge VRComm as a baseline of affordance, or what the technology can provide users during remote collaboration. This approach allows us to make statements about the technological capabilities of VRComm. Therefore, recommendations can be made to develop VRComm for broader remote collaboration purposes. Currently, VRComm is developed for remote conferencing. This could explain why some processes were more limited by the technological capabilities of the set-up than others, as some processes are less prominent depending on the task. For example: distributed performance monitoring and workload monitoring are not essential for having meetings with one other. Furthermore, it is also less important that users have the ability to walk during meetings.

There are some limitations by using these methods as a measure for assessment. Firstly, the findings are based on interpretations of personal experiences with the demo rather than data based on user experiences. Therefore, the results could be attributed to our beliefs and expectancies of VRComm. Another limitation is that we did not explore a broad set of tasks. For example, we did have discussions in the virtual environment, and we tried to identify target objects by looking at them or pointing to them, but we did not perform a task where one user has to build something in the virtual environment while the other one gave the instructions. To tackle these limitations, it is therefore recommended to conduct user experiments. For future development, it would be interesting to measure the aforementioned processes for different task scenarios. As a result, statements can be made how the technological capabilities of the set-up affect all the work processes as mentioned in the abstraction hierarchy. Moreover, as mentioned in the literature studies, most CSCW tools were tested with user groups of dyads and triads. Therefore, it is interesting to see how these tools provide an affordance during experiments with larger groups. Consequently, promoting these tools to translate better to remote collaborative work experiences where larger groups of associates have to collaborate. But in the end, it is of great importance that the CSCW tools are studied under everyday work conditions. Only then, true statements can be made whether the tools offer an affordance or not.

6.6. Future research

There still remain several questions on how social cues influence remote collaboration that could not be answered with the current study. To better understand the influences of social cues and how they should be incorporated into new CSCW systems designs, future research with the following considerations is recommended.

As mentioned, the majority of the papers focused on remote collaboration in pairs. Consequently, it is not possible to make a statement how sharing social cues could benefit the work processes of larger groups. Further research is required to design CSCW systems that incorporate social cues intended for groups larger than two people.

Secondly, almost halve of the results from the studies reported small effect sizes (13 out of 27 results). An explanation for this may be that a large part of the included literature consisted of studies with relatively small sample sizes. To tackle this, conducting future studies with a larger sample size or replicating existing studies with larger sample sizes are necessary to clarify the uncertainty in relationships between the social cues and the work processes.

Thirdly, the interplay between different social cues needs to be researched further. For instance, the results of the study of Yang et al. (2020) indicated that implementations of different combinations of visual gaze and gesture cues with spatialized voice and spatialized auditory beacon cues significantly improved the participants' sense of co-presence compared to a no-cue control condition. However, the integration of hand gestures did not significantly improve sense of co-presence when the visual head frustrum cue was already provided. Moreover, results of the study of Piumsomboon, Dey, Ens, Lee and Billinghurst (2019) indicated that when only gestures were shared, a larger number of gestures was shared compared to conditions that also enabled to share different gaze visualisations. More research is required to understand how different combinations of shared social cues influence remote collaboration.

Furthermore, the interplay of workload monitoring and performance monitoring on social cues needs to be studied further. Currently, the research of Ou et al. (2005a) and Ou, Oh, Yang & Fussell (2005b) are the only two studies that provide evidence that gaze cues are used for performance monitoring. No evidence was found that social cues play a role in workload monitoring. Therefore, more research is needed to understand how social cues influence the processes of performance monitoring and workload monitoring.

From a user-centred design perspective, gaze visualizations can be distracting during processes where there is no need to communicate target locations or to monitor each other's gaze behaviour (D'Angelo & Begel, 2017). When gaze is constantly displayed into the workspace, extra effort is needed from others to decide whether, for example, a fixation is meant for indicating a target object or if it reflects an ongoing search. Therefore, for developing future designs, it should be kept in mind to what extent gaze representations benefit certain task characteristics over the traditional implementations. For example, a mouse cursor could also be used for indicating target objects. Ideally, mouse cursors have the only function to highlight target objects. Consequently, no extra effort is needed from others to decide whether the mouse cursor reflects an ongoing search.

Many papers compared different interface designs and their technological capabilities in terms of usability ratings and performance times. Although it may be beneficial to determine which interface is best suited for the task, these designs often differ in many ways from each other, making it difficult to draw conclusions of what specific aspects of the design promoted the collaboration. A deeper understanding of underlying aspects is needed to develop designs that provide an affordance. Therefore, future research should also focus on methods for controlling independent variables and analysing more objective measures.

7. References

Adams, R. B., Albohn, D. N., & Kveraga, K. (2017). Social Vision: Applying a Social-Functional Approach to Face and Expression Perception. Current Directions in *Psychological Science, 26*(3), 243–248. doi:10.1177/0963721417706392

Attfield S, Fields B, Wheat A, Hutton R, Nixon J, Leggatt A, Blackford H. (2015). Distributed sensemaking: A case study of military analysis, In *Proceedings of 12th International Conference on Naturalistic Decision Making* (pp. 9-12).

Bai, H., Sasikumar, P., Yang, J., & Billinghurst, M. (2020). A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-13).

Bohannon, L. S., Herbert, A. M., Pelz, J. B., & Rantanen, E. M. (2013). Eye contact and video-mediated communication: A review. *Displays*, *34*(2), 177-185. doi:10.1016/j.displa.2012.10.009

Boucher, J. D., Pattacini, U., Lelong, A., Bailly, G., Elisei, F., Fagel, S., ... & Ventre-Dominey, J. (2012). I reach faster when I see you look: gaze effects in human–human and human–robot face-to-face cooperation. *Frontiers in neurorobotics*, *6*, 3..

Bradner, E., & Mark, G. (2001, September). Social presence with video and application sharing. In *Proceedings of the 2001 international ACM SIGGROUP conference on Supporting group work* (pp. 154-161).

Bruner, J. (1986). *Actual minds, possible worlds.* Cambridge, MA: Harvard University Press

Bulu, S. T. (2012). Place presence, social presence, co-presence, and satisfaction in virtual worlds. *Computers & Education*, *58*(1), 154-161. doi:10.1016/j.compedu.2011.08.024

Cabibihan, J. J., So, W. C., Saj, S., & Zhang, Z. (2012). Telerobotic pointing gestures shape human spatial cognition. *International Journal of Social Robotics*, *4*(3), 263-272. doi:10.1007/s12369-012-0148-9

Chan, A., MacLean, K., & McGrenere, J. (2008). Designing haptic icons to support collaborative turn-taking. *International Journal of Human-Computer Studies*, *66*(5), 333-355. doi: 10.1016/j.ijhcs.2007.11.002

Costanza, E., Kunz, A., & Fjeld, M. (2009). Mixed reality: A survey. In *Human machine interaction* (pp. 47-68). Springer, Berlin, Heidelberg.

D'Angelo, S., & Begel, A. (2017). Improving communication between pair programmers using shared gaze awareness. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 6245-6290).

D'Angelo, S., & Gergle, D. (2018). An eye for design: gaze visualizations for remote collaborative work. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

D'Angelo, S., & Gergle, D. (2016). Gazed and confused: Understanding and designing shared gaze for remote collaboration. In *34th Annual Conference on Human Factors in Computing Systems, CHI 2016* (pp. 2492-2496). Association for Computing Machinery.

da Silva, F. P. (2014). Mental workload, task demand and driving performance: what relation?. *Procedia-Social and Behavioral Sciences, 162*, 310-319. doi:10.1016/j.sbspro.2014.12.212

Dell'Agnola, F., Momeni, N., Arza, A., & Atienza, D. (2020, July). Cognitive workload monitoring in virtual reality based rescue missions with drones. In *International Conference on Human-Computer Interaction* (pp. 397-409). Springer, Cham.

Earnshaw, R. A. (Ed.). (2014). *Virtual reality systems*. Academic press.

Ellis, A. P., Hollenbeck, J. R., Ilgen, D. R., Porter, C. O., West, B. J., & Moon, H. (2003). Team learning: Collectively connecting the dots. *Journal of applied Psychology, 88*(5), 821. doi:10.1037/0021-9010.88.5.821

Ellis, C. A., Gibbs, S. J., & Rein, G. (1991). Groupware: some issues and experiences. *Communications of the ACM, 34*(1), 39-58. doi:10.1145/99977.99987

Evans, M. A., Feenstra, E., Ryon, E., & McNeill, D. (2011). A multimodal approach to coding discourse: Collaboration, distributed cognition, and geometric reasoning. *International Journal of Computer-Supported Collaborative Learning, 6(2), 253–278.* doi:10.1007/s11412-011-9113-0

Fidel, R., Mark Pejtersen, A., Cleal, B., & Bruce, H. (2004). A multidimensional approach to the study of human-information interaction: A case study of collaborative information retrieval. *Journal of the American Society for information Science and Technology*, *55*(11), 939-953. doi:10.1002/asi.20041

Fussell, S. R., Setlock, L. D., Parker, E. M., & Yang, J. (2003, April). Assessing the value of a cursor pointing device for remote collaboration on physical tasks. In *CHI'03 Extended Abstracts on Human Factors in Computing Systems* (pp. 788-789).

Fussell, S. R., Setlock, L. D., Yang, J., Ou, J., Mauer, E., & Kramer, A. D. I. (2004). Gestures Over Video Streams to Support Remote Collaboration on Physical Tasks. *Human–Computer Interaction*, 19(3), 273–309. doi:10.1207/s15327051hci1903_3

Gao, L., Bai, H., Lee, G., & Billinghurst, M. (2016). An oriented point-cloud view for MR remote collaboration. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications* (pp. 1-4).

Gauglitz, S., Nuernberger, B., Turk, M., & Höllerer, T. (2014, November). In touch with the remote world: Remote collaboration with augmented reality drawings and virtual navigation. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology* (pp.197-205).

Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures?. *Human factors*, *26*(5), 519-532. doi:10.1177/001872088402600504

Gusenbauer, Michael; Haddaway, Neal R. (2019). *Which Academic Search Systems are Suitable for Systematic Reviews or Meta-Analyses? Evaluating Retrieval Qualities of Google Scholar. PubMed and 26 other Resources*. Research Synthesis Methods, (), jrsm.1378–. doi:10.1002/jrsm.1378

Gunkel, S. N., Hindriks, R., Assal, K. M. E., Stokking, H. M., Dijkstra-Soudarissanane, S., Haar, F. T., & Niamut, O. (2021, July). VRComm: an end-to-end web system for real-time photorealistic social VR communication. In *Proceedings of the 12th ACM Multimedia Systems Conference* (pp. 65-79).

Gupta, K., Lee, G. A., & Billinghurst, M. (2016). Do you see what I see? The effect of gaze tracking on task space remote collaboration. *IEEE transactions on visualization and computergraphics*, *22*(11), 2413-2422. doi:10.1109/TVCG.2016.2593778

Gutwin, C., & Greenberg, S. (1999). *A framework of awareness for small groups in shared-workspace groupware* (pp. 99-1). Technical Report 99-1, Department of Computer Science, University of Saskatchewan, Canada.

Hertel, G., Konradt, U., & Orlikowski, B. (2004). Managing distance by interdependence: Goal setting, task interdependence, and team-based rewards in virtual teams. *European Journal of work and organizational psychology*, *13*(1), 1-28. doi:10.1080/13594320344000228

Higuchi, K., Chen, Y., Chou, P. A., Zhang, Z., & Liu, Z. (2015, April). Immerseboard: Immersive telepresence experience using a digital whiteboard. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2383-2392).

Higuch, K., Yonetani, R., & Sato, Y. (2016). Can eye help you? Effects of visualizing eye fixations on remote collaboration scenarios for physical tasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5180-5190).

Hoedemaeker, M. (2002). *Summary Description of Workload Indicators: WP1 Workload Measures. Human Machine Interface and the Safety of Traffic in Europe Growth Project*. GRD1-2000-25361. HASTE. Institute for Transport Studies. Leeds, UK: University of Leeds.

Huang, W., & Alem, L. (2013). Gesturing in the air: supporting full mobility in remote collaboration on physical tasks. *Journal of Universal Computer Science*, *19*(8), 1158-1174.

Huang, W., Alem, L., & Tecchia, F. (2013, September). HandsIn3D: supporting remote guidance with immersive virtual environments. In *IFIP Conference on Human-Computer Interaction* (pp. 70-77). Springer, Berlin, Heidelberg.

Huron, D. (2018). On the functions of sadness and grief. In Heatcher C. Lench (Ed.), *The function of emotions* (pp. 59-91). Springer, Cham.

Irlitti, A., Piumsomboon, T., Jackson, D., & Thomas, B. H. (2019). Conveying spatial awareness cues in xR collaborations. *IEEE transactions on visualization and computer graphics*, *25*(11), 3178-3189. doi:10.1109/TVCG.2019.2932173

Ishii, H., Kobayashi, M., & Grudin, J. (1993). Integration of interpersonal space and shared workspace: ClearBoard design and experiments. *ACM Transactions on Information Systems (TOIS)*, *11*(4), 349-375. doi:10.1145/159764.159762

Jakobsen, M. R., & Hornbæk, K. (2014). Up close and personal: Collaborative work on a high- resolution multitouch wall display. *ACM Transactions on Computer-Human Interaction (TOCHI), 21*(2), 1-34. doi:10.1145/2576099

Jay, C., Glencross, M., & Hubbold, R. (2007). Modelling the effects of delayed haptic and visual feedback in a collaborative virtual environment. ACM Transactions on Computer-Human Interaction, 14(2), 8–es. doi:10.1145/1275511.1275514

Jennings, N. R. (1996). Coordination techniques for distributed artificial intelligence. In, O'Hare, G. M. P. and Jennings, N. R. (Eds.) *Foundations of Distributed Artificial Intelligence. Foundations of Distributed Artificial Intelligence* (pp. 187-210). Wiley.

Jo, D., Kim, K. H., & Kim, G. J. (2016). Effects of avatar and background representation forms to co-presence in mixed reality (MR) tele-conference systems. In *SIGGRAPH ASIA 2016 virtual reality meets physical reality: modelling and simulating virtual humans and environments* (pp.1-4).

Johnson-Glenberg, M. C., Birchfield, D. A., Tolentino, L., & Koziupa, T. (2014). Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *Journal of Educational Psychology, 106*(1), 86–104. doi:10.1037/a0034008

Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C. M., Van Riemsdijk, M. B., & Sierhuis, M. (2014). Coactive design: Designing support for interdependence in joint activity. *Journal of Human-Robot Interaction*, *3*(1), 43-69. doi:10.5898/JHRI.3.1.Johnson

Jones, S. (1993). A sense of space: Virtual reality, authenticity and the aural. *Critical Studies in Media Communication*, *10*(3), 238-252. doi:10.1080/15295039309366866

Kim, S., Lee, G., Huang, W., Kim, H., Woo, W., & Billinghurst, M. (2019, May). Evaluating the combination of visual communication cues for HMD-based mixed reality remote collaboration. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-13).

Klein, C., DiazGranados, D., Salas, E., Le, H., Burke, C. S., Lyons, R., & Goodwin, G. F. (2009). Does team building work?. *Small group research, 40*(2), 181-222. doi:10.1177/1046496408328821

Klein, G., Phillips, J. K., Rall, E. L., & Peluso, D. A. (2007, January). A data-frame theory of sensemaking. In *Expertise out of context: Proceedings of the sixth international conference on naturalistic decision making* (Vol. 113). New York: Lawrence Erlbaum Associates.

Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten challenges for making automation a "team player" In *joint human-agent activity. IEEE Intelligent Systems,* 19(6), 91–95. doi:10.1109/MIS.2004.74

Kmet, L.M., Cook, L.S. and Lee, R.C. (2014) Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields. Edmonton: Alberta Heritage Foundation for Medical Research (AHFMR). https://www.ihe.ca/publications/standard-quality-assessment-criteria-for-evaluating-primary-research-papers-from-a-variety-of-fields

Kunz, A., Nescher, T., & Kuchler, M. (2010, October). Collaboard: a novel interactive electronic whiteboard for remote collaboration with people on content. In *2010 International Conference on Cyberworlds* (pp. 430-437). IEEE.

Kütt, G. H., Tanprasert, T., Rodolitz, J., Moyza, B., So, S., Kenderova, G., & Papoutsaki, A. (2020). Effects of Shared Gaze on Audio-Versus Text-Based Remote Collaborations. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), 1-25. doi:10.1145/3415207

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press.

Le Chénéchal, M., Duval, T., Gouranton, V., Royan, J., & Arnaldi, B. (2016, March). Vishnu: virtual immersive support for helping users an interaction paradigm for collaborative remote guiding in mixed reality. In *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE)* (pp. 9-12). IEEE.

Lee, G. A., Kim, S., Lee, Y., Dey, A., Piumsomboon, T., Norman, M., & Billinghurst, M. (2017a, November). Improving Collaboration in Augmented Video Conference using Mutually Shared Gaze. In *ICAT-EGVE* (pp. 197-204).

Lee, G. A., Teo, T., Kim, S., & Billinghurst, M. (2017b). Mixed reality collaboration through sharing a live panorama. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications* (pp. 1-4).

Lee, G. A., Teo, T., Kim, S., & Billinghurst, M. (2018). A user study on mr remote collaboration using live 360 video. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 153-164). IEEE.

Lesser, V. R. (1998). Reflections on the nature of multi-agent coordination and its implications for an agent architecture. *Autonomous agents and multi-agent systems*, *1*(1), 89-111. doi:10.1023/A:1010046623013

Lind, M. (2003). Making sense of the abstraction hierarchy in the power plant domain. *Cognition, Technology & Work*, *5*(2), 67-81. doi:10.1007/s10111-002-01094

Lissermann, R., Huber, J., Schmitz, M., Steimle, J., & Mühlhäuser, M. (2014, April). Permulin: mixed-focus collaboration on multi-view tabletops. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3191-3200).

Llobera, J., Spanlang, B., Ruffini, G., & Slater, M. (2010). Proxemics with multiple dynamic characters in an immersive virtual environment. *ACM Transactions on Applied Perception (TAP)*, *8*(1), 1-12. doi:10.1145/1857893.1857896

Miller, S. (2001). Workload measures. *National Advanced Driving Simulator. Iowa City, United States*.

Müller, R., Helmert, J. R., Pannasch, S., & Velichkovsky, B. M. (2013). Gaze transfer in remote cooperation: Is it always helpful to see what your partner is attending to?. *Quarterly journal of experimental psychology*, *66*(7), 1302-1316. doi:10.1080/17470218.2012.737813

Neale, D. C., Carroll, J. M., & Rosson, M. B. (2004, November). Evaluating computer-supported cooperative work: models and frameworks. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work* (pp. 112-121).

Nuernberger, B., Lien, K. C., Höllerer, T., & Turk, M. (2016, March). Interpreting 2d gesture annotations in 3d augmented reality. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)* (pp. 149-158). IEEE.

Olwal, A., Frykholm, O., Groth, K., & Moll, J. (2011, September). Design and evaluation of interaction technology for medical team meetings. In *IFIP Conference on Human-Computer Interaction* (pp. 505-522). Springer, Berlin, Heidelberg.

Ou, J., Oh, L. M., Fussell, S. R., Blum, T., & Yang, J. (2005a). Analyzing and predicting focus of attention in remote collaborative tasks. In *Proceedings of the 7th international conference on Multimodal interfaces* (pp. 116-123).

Ou, J., Oh, L. M., Fussell, S. R., Blum, T., & Yang, J. (2008). Predicting visual focus of attention from intention in remote collaborative tasks. *IEEE Transactions on Multimedia*, *10*(6), 1034-1045. doi:10.1109/TMM.2008.2001363

Ou, J., Oh, L. M., Yang, J., & Fussell, S. R. (2005b). Effects of task properties, partner actions, and message content on eye gaze patterns in a collaborative task. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 231-240).

Oyekoya, O., Steptoe, W., & Steed, A. (2012, May). Sphereavatar: A situated display to represent a remote collaborator. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2551-2560).

Pauchet, A., Coldefy, F., Lefebvre, L., Picard, S. L. D., Bouguet, A., Perron, L., ... & Collobert, M. (2007a, September). Mutual awareness in collocated and distant collaborative tasks using shared interfaces. In *IFIP Conference on Human-Computer Interaction* (pp. 59-73). Springer, Berlin, Heidelberg.

Pauchet, A., Coldefy, F., Lefebvre, L., Picard, S. L. D., Perron, L., Bouguet, A., ... & Corvaisier, D. (2007b, October). TableTops: Worthwhile experiences of collocated and remote collaboration. In *Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'07)* (pp. 27-34). IEEE.

Piumsomboon, T., Dey, A., Ens, B., Lee, G., & Billinghurst, M. (2019). The effects of sharing awareness cues in collaborative mixed reality. *Frontiers in Robotics and AI, 6*, 1-47. doi:10.3389/frobt.2019.00005

Piumsomboon, T., Lee, G. A., Hart, J. D., Ens, B., Lindeman, R. W., Thomas, B. H., & Billinghurst, M. (2018). 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).

Podkosova, I., & Kaufmann, H. (2018, November). Co-presence and proxemics in shared walkable virtual environments with mixed colocation. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (pp. 1-11).

Prouzeau, A., Bezerianos, A., & Chapuis, O. (2016). Evaluating multi-user selection for exploring graph topology on wall-displays. *IEEE Transactions on Visualization and Computer Graphics, 23*(8), 1936-1951. doi:10.1109/TVCG.2016.2592906

Rasmussen. I. (1986*). information processing and human-machine interactions An approach to cognitive engineering*. New York: North-Holland.

Reising, D. V. C. (2000, July). The abstraction hierarchy and its extension beyond process control. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 44, No. 1, pp. 194-197). Sage CA: Los Angeles, CA: SAGE Publications

Sallnäs, E. L., Rassmus-Gröhn, K., & Sjöström, C. (2000). Supporting presence in collaborative environments by haptic force feedback. *ACM Transactions on Computer-Human Interaction (TOCHI), 7*(4), 461-476. doi:10.1145/365058.365086

Sauppé, A., & Mutlu, B. (2014, February). How social cues shape task coordination and communication. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 97-108). Baltimore, Maryland, USA

Sherman, W. R., & Craig, A. B. (Eds.) (2003). Understanding virtual reality. *San Francisco, CA: Morgan Kauffman*.

Shin, S. J., Kim, T. Y., Lee, J. Y., & Bian, L. (2012). Cognitive team diversity and individual team member creativity: A cross-level interaction. *Academy of management journal, 55*(1), 197-212. doi:10.5465/amj.2010.0270

Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British Journal of Psychology, 109*(3), 431-433. doi:10.1111/bjop.12305

Talja, S., & Hansen, P. (2006). Information sharing. In *New directions in human information behavior* (pp. 113-134). Springer, Dordrecht.

Terken, J., & Sturm, J. (2010). Multimodal support for social dynamics in co-located meetings. *Personal and Ubiquitous Computing, 14*(8), 703-714. doi:10.1007/s00779-010-0284-x

Verwey, W. B. (2000). On-line driver workload estimation. Effects of road situation and age on secondary task measures. *Ergonomics, 43*(2), 187-209. doi:10.1080/001401300184558

Wang, P., Bai, X., Billinghurst, M., Zhang, S., Han, D., Sun, M., ... & Han, S. (2020a). Haptic feedback helps me? A VR-SAR remote collaborative system with tangible interaction. *International Journal of Human–Computer Interaction, 36*(13), 1242-1257. doi:10.1080/10447318.2020.1732140

Wang, P., Bai, X., Billinghurst, M., Zhang, S., He, W., Han, D., Wang, Y., Min, H., Lan, W., & Han, S. (2020b). Using a Head Pointer or Eye Gaze: The Effect of Gaze on Spatial AR Remote Collaboration for Physical Tasks. *Interacting with Computers, 32*(2), 153-169. doi: 10.1093/iwcomp/iwaa012

Wang, P., Bai, X., Billinghurst, M., Zhang, S., Wei, S., Xu, G., ... & Zhang, J. (2020c). 3DGAM: using 3D gesture and CAD models for training on mixed reality remote collaboration. *Multimedia Tools and Applications,* 1-26. doi:10.1007/s11042-020-09731-7

Wang, P., Zhang, S., Bai, X., Billinghurst, M., He, W., Sun, M., Lv, H. & Ji, H. (2019). 2.5 DHANDS: a gesture-based MR remote collaborative platform. *The International Journal of Advanced Manufacturing Technology, 102*(5), 1339-1353. doi:10.1007/s00170-018-03237-1

Williamson, J., Li, J., Vinayagamoorthy, V., Shamma, D. A., & Cesar, P. (2021, May). Proxemics and social interactions in an instrumented virtual reality workshop. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Wolff, R., Roberts, D. J., Steed, A., & Otto, O. (2007). A review of telecollaboration technologies with respect to closely coupled collaboration. *International Journal of Computer Applications in Technology, 29*(1), 11. doi:10.1504/ijcat.2007.014056

Woods, D. D., & Hollnagel, E. (2006). *Joint cognitive systems: Patterns in cognitive systems engineering*. CRC Press. doi:10.1201/9781420005684

Wuehr, M., Breitkopf, K., Decker, J., Ibarra, G., Huppert, D., & Brandt, T. (2019). Fear of heights in virtual reality saturates 20 to 40 m above ground. *Journal of neurology, 266*(1), 80-87. doi:10.1007/s00415-019-09370-5

Xiao, X., & Ishii, H. (2010, January). MirrorFugue: communicating hand gesture in remote piano collaboration. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction* (pp. 13-20).

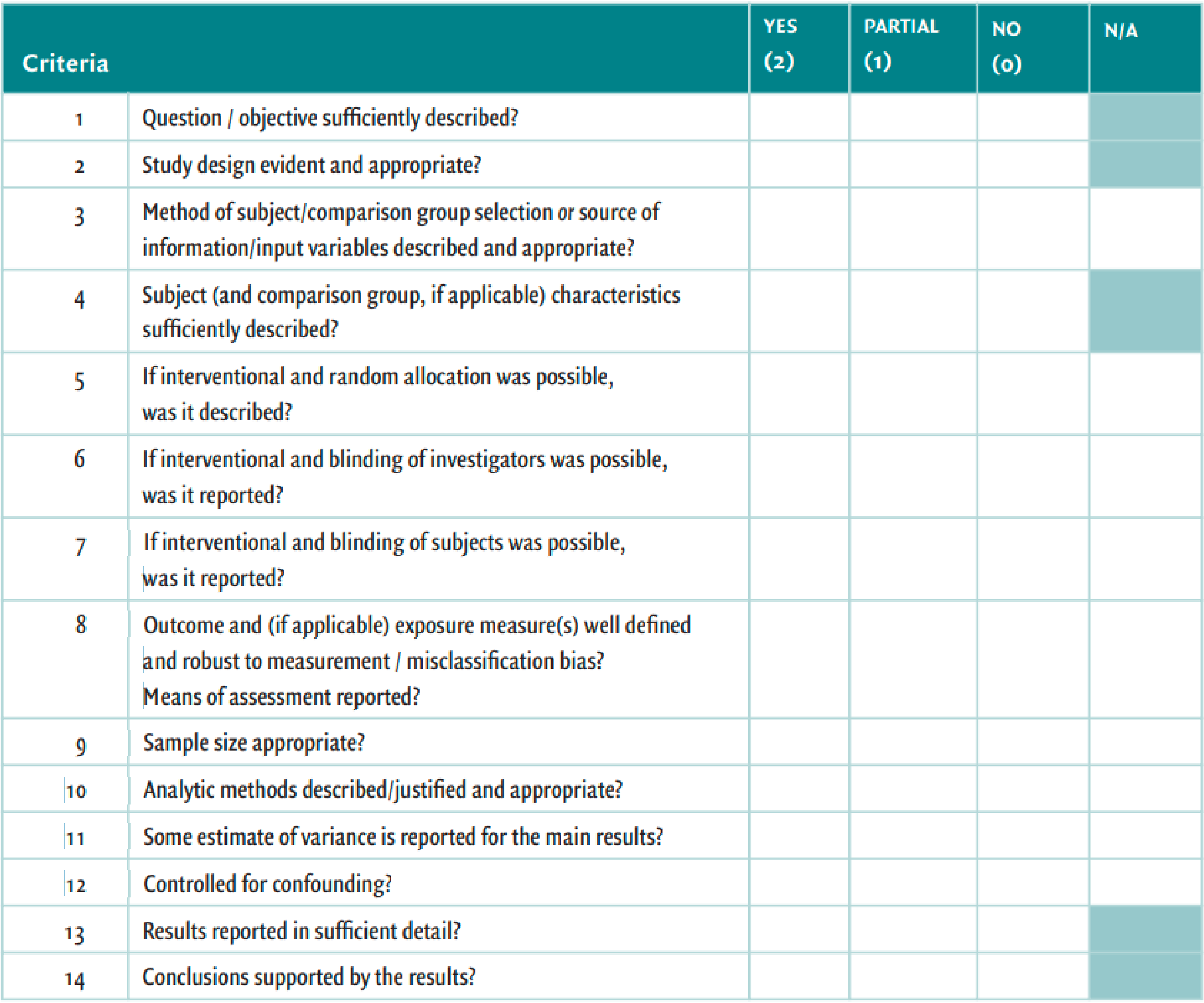
Yamashita, N., Kaji, K., Kuzuoka, H., & Hirata, K. (2011, March). Improving visibility of remote gestures in distributed tabletop collaboration. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work* (pp. 95-104).

Yang, J., Sasikumar, P., Bai, H., Barde, A., Sörös, G., & Billinghurst, M. (2020). The effects of spatial auditory and visual cues on mixed reality remote collaboration. *Journal on Multimodal User Interfaces, 14*(4), 337-352. doi:10.1007/s12193-020-00331-1

Zhang, Y., Pfeuffer, K., Chong, M. K., Alexander, J., Bulling, A., & Gellersen, H. (2017). Look together: using gaze for assisting co-located collaborative search. *Personal and Ubiquitous Computing*, *21*(1), 173-186. doi:10.1007/s00779-016-0969-x

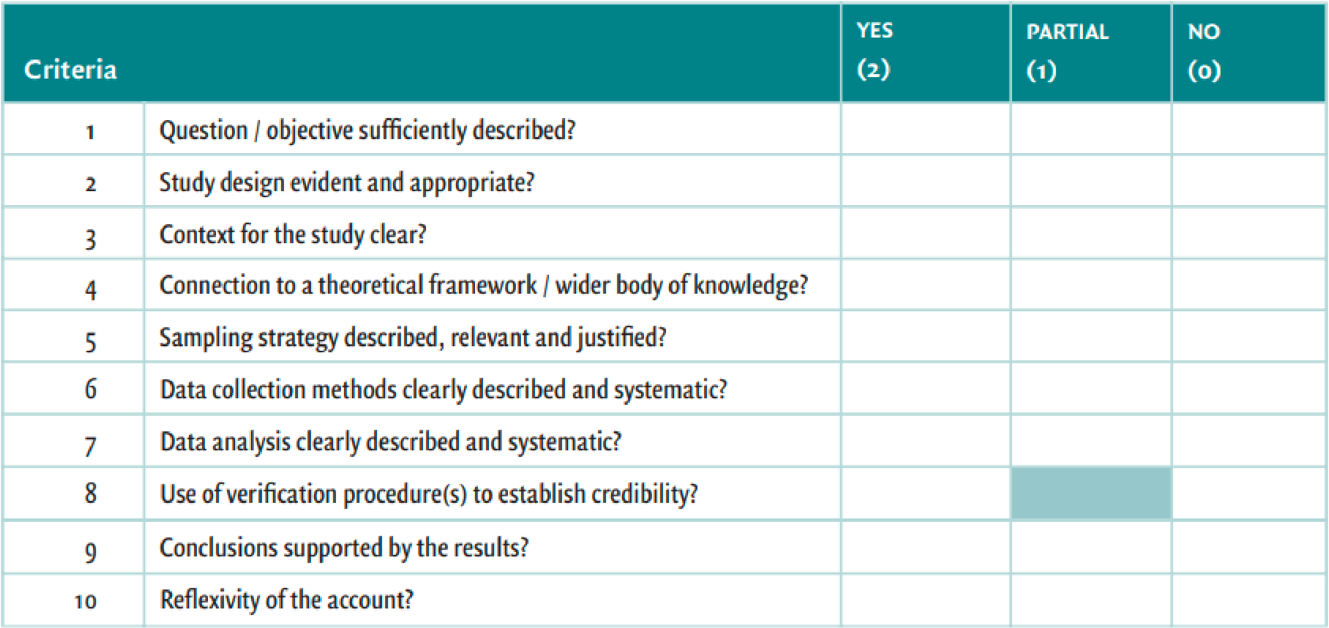
8. Appendices

Appendix A. Checklist for assessing the quality of quantitative studies



Criteria		YES (2)	PARTIAL (1)	NO (0)	N/A
1	Question / objective sufficiently described?				▨
2	Study design evident and appropriate?				▨
3	Method of subject/comparison group selection or source of information/input variables described and appropriate?				
4	Subject (and comparison group, if applicable) characteristics sufficiently described?				▨
5	If interventional and random allocation was possible, was it described?				
6	If interventional and blinding of investigators was possible, was it reported?				
7	If interventional and blinding of subjects was possible, was it reported?				
8	Outcome and (if applicable) exposure measure(s) well defined and robust to measurement / misclassification bias? Means of assessment reported?				
9	Sample size appropriate?				
10	Analytic methods described/justified and appropriate?				
11	Some estimate of variance is reported for the main results?				
12	Controlled for confounding?				
13	Results reported in sufficient detail?				▨
14	Conclusions supported by the results?				▨

Appendix B. Checklist for assessing the quality of qualitative studies



Criteria		YES (2)	PARTIAL (1)	NO (0)
1	Question / objective sufficiently described?			
2	Study design evident and appropriate?			
3	Context for the study clear?			
4	Connection to a theoretical framework / wider body of knowledge?			
5	Sampling strategy described, relevant and justified?			
6	Data collection methods clearly described and systematic?			
7	Data analysis clearly described and systematic?			
8	Use of verification procedure(s) to establish credibility?			
9	Conclusions supported by the results?			
10	Reflexivity of the account?			

Appendix C. Mean scores of the included qualitative and quantitative studies

Title Article	Type of Study	Mean Scores
Collaborative embodied learning in mixed reality motion-capture environments: Two science studies	Quantitative	2.00
Multimodal support for social dynamics in co-located meetings	Qualitative	2.00
Gestures over video streams to support remote collaboration on physical tasks	Quantitative	1.92
Modeling the effects of delayed haptic and visual feedback in a collaborative virtual environment	Quantitative	1.92
Gazed and confused: Understanding and designing shared gaze for remote collaboration	Quantitative	1.92
The effects of sharing awareness cues in collaborative mixed reality	Quantitative	1.92
A multimodal approach to coding discourse: Collaboration, distributed cognition, and geometric reasoning	Qualitative	1.90
A User Study on MR Remote Collaboration Using Live 360 Video	Quantitative	1.83
A User Study on Mixed Reality Remote Collaboration with Eye Gaze and Hand Gesture Sharing	Quantitative	1.83
Mini-me: An adaptive avatar for Mixed Reality remote collaboration	Qualitative	1.82
Supporting Presence in Collaborative Environments by Haptic Force Feedback	Quantitative	1.75
Do you see what i see? the effect of gaze tracking on task space remote collaboration	Quantitative	1.75
Improving communication between pair programmers using shared gaze awareness	Quantitative	1.75
Gaze transfer in remote cooperation: Is it always helpful to see what your partner is attending to?	Quantitative	1.75
An eye for design: Gaze visualizations for remote collaborative work	Quantitative	1.75
Haptic Feedback Helps Me? A VR-SAR Remote Collaborative System with Tangible Interaction	Quantitative	1.75
Effects of Shared Gaze on Audio-Versus Text-Based Remote Collaborations	Quantitative	1.75
Telerobotic Pointing Gestures Shape Human Spatial Cognition	Quantitative	1.73

Appendix C. (Continued).

Title Article	Type of Study	Mean Scores
Analyzing and predicting focus of attention in remote collaborative tasks	Qualitative	1.70
Designing haptic icons to support collaborative turn-taking	Quantitative	1.67
Immerse board: Immersive tele presence experience using a digital whiteboard	Quantitative	1.67
Evaluating the Combination of Visual Communication Cues for HMD-based Mixed Reality Remote Collaboration	Quantitative	1.67
Improving collaboration in augmented video conference using mutually shared gaze	Quantitative	1.67
2.5DHANDS: a gesture-based MR remote collaborative platform	Quantitative	1.67
Using a head pointer or eye gaze: The effect of gaze on spatial AR remote collaboration for physical tasks	Quantitative	1.67
Can eye help you?: Effects of visualizing eye fixations on remote collaboration scenarios for physical tasks	Quantitative	1.58
Look together: using gaze for assisting co-located collaborative search	Quantitative	1.58
Improving visibility of remote gestures in distributed tabletop collaboration	Quantitative	1.58
Assessing the value of a cursor pointing device for remote collaboration on physical tasks	Quantitative	1.58
MirrorFugue: Communicating hand gesture in remote piano collaboration	Quantitative	1.58
Gesturing in the air: Supporting full mobility in remote collaboration on physical tasks	Quantitative	1.58
The effects of spatial auditory and visual cues on mixed reality remote collaboration	Quantitative	1.58
Interpreting 2D gesture annotations in 3D augmented reality	Quantitative	1.50
Effects of task properties, partner actions, and message content on eye gaze patterns in a collaborative task	Quantitative	1.50
How social cues shape task coordination and communication	Quantitative	1.50
Mutual awareness in collocated and distant collaborative tasks using shared interfaces	Quantitative	1.50

Appendix C. (Continued).

Title Article	Type of Study	Mean Scores
Improving co-located collaboration with show-through techniques	Quantitative	1.50
Mixed reality collaboration through sharing a live panorama	Quantitative	1.45
Permulin: Mixed-focus collaboration on multi-view tabletops	Quantitative	1.42
TableTops: Worthwhile experiences of collocated and remote collaboration	Quantitative	1.42
3DGAM: using 3D gesture and CAD models for training on mixed reality remote collaboration	Quantitative	1.42
In touch with the remote world: Remote collaboration with augmented reality drawings and virtual navigation	Qualitative	1.40
Design and evaluation of interaction technology for medical team meetings	Qualitative	1.40
Predicting visual focus of attention from intention in remote collaborative tasks	Quantitative	1.40
Vishnu: Virtual immersive support for HelpiNg users an interaction paradigm for collaborative remote guiding in mixed reality	Quantitative	1.36
An oriented point-cloud view for MR remote collaboration	Quantitative	1.36
I reach faster when i see you look: Gaze effects in human-human and human-robot face-to-face cooperation	Quantitative	1.33
HandsIn3D: Supporting remote guidance with immersive virtual environments	Quantitative	1.33
CollaBoard: A novel interactive electronic whiteboard for remote collaboration with people on content	Quantitative	1.33
Coordinating turn-taking with gaze	Qualitative	1.30
Up close and personal: Collaborative work on a high-resolution multitouch wall display	Quantitative	1.27
Exploring enhancements for remote mixed reality collaboration	Quantitative	1.25
Effective cooperative haptic interaction over the Internet	Quantitative	1.25
Stitching: Pen gestures that span multiple displays	Qualitative	1.20

Appendix C. (Continued).

Title Article	Type of Study	Mean Scores
ShadowPuppets: Supporting collocated interaction with mobile projector phones using hand shadows	Qualitative	1.20
Where are you pointing at?' A study of remote collaboration in a wearable videoconference system	Quantitative	1.18
CoReach: Cooperative gestures for data manipulation on wall-sized displays	Quantitative	1.18
3D helping hands: A gesture based MR system for remote collaboration	Quantitative	1.10
Simultaneous remote haptic collaboration for assembling tasks	Qualitative	1.10
MultiView: Spatially faithful group video conferencing	Quantitative	1.08
Augmented 3D hands: a gesture-based mixed reality system for distributed collaboration	Quantitative	1.08
Supporting hand gestures in mobile remote collaboration: A usability evaluation	Quantitative	1.08
WeSpace: The design, development, and deployment of a walk-up and share multi-surface collaboration system	Quantitative	1.00
Study of augmented gesture communication cues and view sharing in remote collaboration	Quantitative	0.92
HandsInAir: A wearable system for remote collaboration on physical tasks	Qualitative	0.90
GazeTorch: Enabling gaze awareness in collaborative physical tasks	Qualitative	0.90
Distributed pointing for multimodal collaboration over sketched diagrams	Quantitative	0.90
PhyShare: Sharing physical interaction in virtual reality	Quantitative	0.83
Design and Implementation of TeleAdvisor: a Projection-Based Augmented Reality System for Remote Collaboration	Quantitative	0.75
Clearboard: a seamless medium for shared drawing and conversation with eye contact	Qualitative	0.70
Designing shared gaze awareness for remote collaboration	Qualitative	0.60
A proposal of body movement-based interaction towards remote collaboration for concurrent engineering	Quantitative	0.58

Appendix C. (Continued).

Title Article	Type of Study	Mean Scores
Lark: Coordinating co-located collaboration with information visualization	Qualitative	0.50
GestureCam: A video communication system for sympathetic remote collaboration	Qualitative	0.40
GroupSketch. A multi-user sketchpad for geographically-distributed small groups	Qualitative	0.30