

Dimitra Anastasiou*, Adrien Coppens and Valérie Maquil

Gesture combinations during collaborative decision-making at wall displays

<https://doi.org/10.1515/icom-2023-0037>

Received December 15, 2023; accepted March 3, 2024;

published online March 25, 2024

Abstract: This paper describes an empirical user study with 24 participants during collaborative decision-making at large wall displays. The main objective of the user study is to analyze combinations of mid-air pointing gestures with other gestures or gaze. Particularly, we investigate gesture sequences (having pointing gestures as an initiator gesture) and gaze-pointing gesture misalignments. Our results show that most pointing gestures are part of gesture sequences and more precise gestures lead to touch gestures on the wall display, likely because they are associated with precise concepts. Regarding combinations of pointing gestures and gaze, misalignments often happen when users touch the display to make a change and want to observe the effect of that change on another display. The analyses conducted as part of this study clarify which natural awareness cues are more frequent in face-to-face collaboration, so that appropriate choices can be made regarding the transmission of equivalent cues to a remote location.

Keywords: gaze-gesture alignment; gesture sequences; wall displays; collaborative decision-making

1 Introduction

Large interactive wall displays offer unique advantages for collaborative data analysis and decision-making. Due to their considerable size and resolution, they can present large amounts of data in different scales and views, and side-by-side, supporting users to better identify details and gain more insights about the data [1, 2]. They have been shown

to be useful, for instance, in road traffic management [3], automotive design [4], medical coordination [5], and architectural design [6]. Furthermore, large wall displays support collaboration as they can easily accommodate multiple users that are able to access and view content at the same time, and follow the respective actions of each other [7]. In these situations of collaborative decision-making, users are naturally making use of a large number of hand gestures which provide them with “workspace awareness”, i.e., an “up-to-the-moment understanding of another person’s interaction with a shared space” [8]. In this paper, we consider the concept of **natural awareness cues**, which provide such awareness information to collocated collaborators. These include non-verbal communication indicators, such as body position and movement, hand gestures, and gaze as natural awareness cues.

Gesture-based human machine interaction has been an application field of work both from researchers and designers and it has evolved with regards to sensing and processing the data. An analysis of recent taxonomies and a literature classification of gesture-based interfaces can be found in Carfi and Mastrogiovanni [9]. While gestures have been deeply investigated in human computer interaction mainly on multi-touch or tangible user interfaces, research on gestures on large wall displays still needs more empirical studies, particularly of gestures produced under realistic conditions. Such empirical studies including multiple participants simultaneously exploring data and discussing a concept at wall displays can provide insights not only on collaboration patterns between users, but also on data visualization methods and techniques that can be used on large displays.

In the ReSurf¹ project, we seek to support mixed-presence decision-making between two wall displays. To remedy the lack of awareness information that can be transmitted by conventional audio-video links, we seek to design **synthetic awareness cues** that track awareness information and make them accessible over distance. For instance, in Figure 1, users interact with the data (i.e. filter, query, zoom) and explain their thoughts to both collocated and remote group members. During their interactions, gaze, posture and hand gestures are tracked and visualized using

*Corresponding author: Dimitra Anastasiou, Luxembourg Institute of Science and Technology, Human Modelling Group, Esch-sur-Alzette, Luxembourg, E-mail: dimitra.anastasiou@list.lu. <https://orcid.org/0000-0002-9037-0317>

Adrien Coppens and Valérie Maquil, Luxembourg Institute of Science and Technology, Visualisation and Interaction Group, Esch-sur-Alzette, Luxembourg, E-mail: adrien.coppens@list.lu (A. Coppens), valerie.maquil@list.lu (V. Maquil). <https://orcid.org/0000-0002-2841-6708> (A. Coppens). <https://orcid.org/0000-0002-0198-3729> (V. Maquil)

¹ <https://www.list.lu/en/informatics/project/resurf/>.

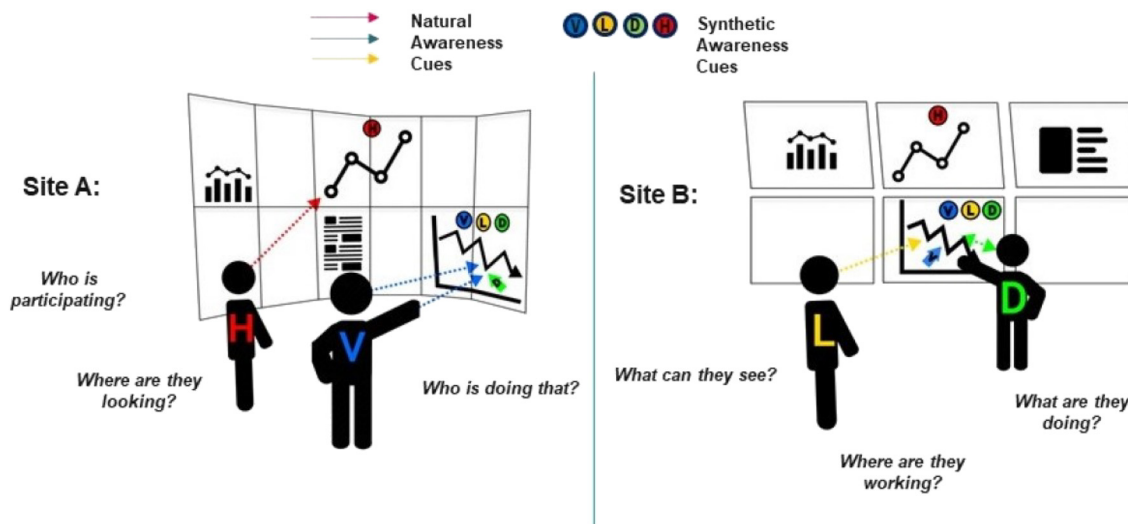


Figure 1: Tracking and visualizing awareness cues based on gestures and gaze.

pointers. These synthetic awareness cues allow the group members to more easily follow the actions of the speaker by being informed of where s/he is standing, and where s/he is pointing and looking at. The speaker is, in turn, notified of subtle reactions by other group members, such as attention shifts. Overall, with such a system, the group can collaborate in a fluid and natural way, without time-consuming and fatigue-related distractions, such as the need to interrupt each other to indicate where to look or to check whether others (and who) are following. In this context, we define synthetic awareness cues as *digital indicators conveying awareness information*; they are the digital equivalent to natural awareness cues.

We planned various user studies during the project, the initial ones taking place at one site with the aim to observe behavioural patterns of acquiring and providing natural awareness cues during collaboration at wall displays. Indeed, we cannot convey all non-verbal communication through synthetic awareness cues, as this would be overwhelming and distracting for users interacting with the display. Therefore, we need to evaluate which behavioural patterns are most efficient for collaborative work and should be prioritized.

The motivation for this study particularly, is to observe gesture sequences and gesture-gaze misalignments on wall displays. We define a **gesture sequence** as *a combination of an initial pointing gesture (the initiator) followed by one or more other gestures that are part of the same discussion point*.

While the sequence must be initiated by a pointing gesture, subsequent (or follow-up) gestures may be of any

type, i.e. same or other type of pointing, touch, emblem or adaptor gestures (see Section 2.1).

Our gestures sequences relate to Morris et al. [10], who defined the relative timing of each contributor's actions as "parallelism". In particular, they defined a sequence as "parallel", if all users perform their gesture simultaneously and "serial", if a user's gesture immediately follows another's gesture. However, we differentiate our concept of gesture sequence from Morris et al. [10], since in their definition of *serial*, the entire sequence accomplishes nothing unless everyone finishes their action, whereas in our case, each gesture could also be stand-alone.

The present study took place in Luxembourg in October 2022 with a circular wall-sized display. Our research method includes the analysis of pointing gestures not in isolation, but as part of gesture sequences. Moreover, in this paper we explore the cases of gaze-pointing gesture misalignments, i.e. when a user performs a pointing gesture not toward where (s)he is frontally viewing.

Our results allow researchers to better understand which gesture sequences lead to touch actions and are therefore most important for collaborative work. In addition, gesture-gaze misalignments indicate to what extent gestural and gaze related information can be considered as complementary. The descriptive statistical results will facilitate investigating the design of awareness cues that visualise specific gestural and gaze information on a distributed wall display at a remote location.

The paper is laid out as follows: in the literature review we briefly report on existing works on gesture (Section 2.1) as well as gaze and visual attention (Section 2.2) as two distinct natural awareness cues that can be

mediated as synthetic ones on wall displays. In Section 3, we describe our study, including our research questions and research method (participants, scenario, apparatus, analysis). Section 4 focuses on the results of the frequency and types of the gesture sequences and the gaze-pointing misalignments. A discussion about the implications and limitations of this work and our future projects are found in Section 5.

2 Literature review

Gestures (see Section 2.1) and gaze (Section 2.2) are two distinct natural awareness cues of non-verbal behavior that multiple users exchange while they view content at the same time on wall displays. The awareness information makes collaborators' actions and intentions clear and allows them to seamlessly align and integrate their activities with other group members.

2.1 Gestures on wall displays

According to Gutwin and Greenberg [8], the main sources of awareness information are (i) people's bodies, (ii) workspace artifacts, and (iii) conversation and gestures. As for the third point, there is a distinction between intentional explicit gestures and consequential communication. While the former are the stereotypical gestures with clear intention of pointing somewhere or something, the latter is information transfer that emerges as a consequence of a person's activity within an environment [8]. Pointing gestures have been examined by many scholars, such as linguists, semioticians, psychologists, anthropologists, and primatologists. In psycholinguistics, the most prominent gesture taxonomy is that of McNeill [11], who categorized the gestures into gesticulation, emblems, pantomimes, and sign language. Gesticulation is further classified into iconic, metaphoric, rhythmic, cohesive, and deictic or pointing gestures. The prototypical pointing gesture (analogous to the "stereotypical" mentioned before) is a communicative body movement that projects a vector from a body part, with this vector indicating a certain direction, location, or object [12].

Gestures are an indispensable part of embodied cognition, which is cognitive science that implies that thinking and perception are shaped by interactions with the physical environment [13]. Regarding the relation between gestures and embodied cognition, Soni et al. recently identified four types of gesture interactions that promote scientific discussion and collaborative meaning-making through

embodied cognition [14]: (T1) gestures for orienting the group; (T2) cooperative gestures for facilitating group meaning-making; (T3) individual intentional gestures for facilitating group meaning-making; and (T4) gestures for articulating conceptual understanding to the group. In our opinion, T3 is analogous to the intentional gestures and T4 to the consequential communication based on the framework of Gutwin and Greenberg [8].

The relationship between gesture and thought is described in the literature review paper of Goldin-Meadow and Beilock [15]. They state that gesture actively brings action into a speaker's mental representations, and those mental representations then affect behavior – at times more powerfully than the actions on which the gestures are based. Gestures have been examined specifically in relation to problem-solving and decision-making [16–19]. In the study of Alibali et al. [16], they examined whether gestures play a functional role in problem-solving. In their study, participants in two experiments solved problems requiring the prediction of gear movement, either with gesture allowed or with gesture prohibited. They found that participants in the gesture-allowed condition were more likely to use perceptual-motor strategies in the gesture-prohibited condition. Both rotation and ticking gestures tended to accompany perceptual-motor strategies.

As far as gestures specifically in relation to wall-displays are concerned, Liu et al. introduced CoReach [20], a set of collaborative gestures that combine input from multiple users to manipulate content, facilitate data exchange and support communication. In their experiment on a wall display, they asked participants to find similarities and connections between pictures and to arrange them in a meaningful way that they could agree on. In the experiment of Liu et al. [20], it is mainly touch gestures being analyzed compared to our study, which is about mid-air pointing gestures as initiators for gesture sequences. Gesture elicitation studies are also a common and efficient way to create gesture sets [21] and as for eliciting mid-air gestures for wall displays, it has been observed by Wittorf and Jakobsen [22] that the size or extent of gestures is related to the size of the display. In other words, users make larger and more physically-based gestures in wall displays than in smaller displays [22].

According to Hinrichs and Carpendale [23], gestures should not be considered in isolation from previous and subsequent gestures. They have explored gesture sequences in multi-touch interfaces in-the-wild, but to our knowledge, an empirical study with gesture sequences under realistic collaborative conditions at large wall displays, as described in this paper, has not been explored until now.

Noteworthy is the distinction of our definition of gesture sequences from the so-called cooperative gestures. According to Morris et al. [10], cooperative gestures can be used to enhance users' sense of teamwork, increase awareness of important system events, facilitate reachability and access control on large, shared displays. However in Morris et al.'s approach [10], the gestures of the users were considered as a single, combined command. The case study, which is presented in this paper, builds upon the work of Maquil et al. [24]. In particular, we look at a pointing gesture produced by a single user, consider this as a single command, and then explore the reaction(s) of the other users towards this initial one-user command. Therefore, we differentiate our work from cooperative gestures as defined in Morris et al. [10]. We follow the categorization of Maquil et al. [24] about pointing gestures (see Table 1) and analyze accordingly the gestures produced by the participants in our user study. All three categories in Table 1 are mid-air gestures and not touch gestures.

The first type is narrative pointing (NP), where according to Maquil et al. [24], a user points sharply with the index finger, but also moves the finger towards a larger area of the display (up/down or left/right). The second type of pointing gesture is loose pointing (LP). Here, the user is not looking at the screen and holds the hand usually open or the palm up. LP gestures often happen when a user describes a concept as a whole. A third type is that of sharp pointing (SP). This is the "stereotypical" pointing gesture with an index finger where a user points to a very specific area of the display (e.g., a specific word or number). Its duration is usually much shorter than NP.

In this paper we extend the taxonomy of pointing gestures to explore which gestures come as subsequent gestures to NP, LP, and SP. The research question that we aim to answer through our studies is "*Which are the most frequent subsequent gestures after narrative, loose, and sharp pointing respectively?*"

The potential gestural reactions following NP, LP, and SP are:

- Lack of action;
- Other pointing gestures of any type: NP, LP, SP;
- Emblems or adaptors;
- Touch gestures.

These reactions can be produced by the same user or anybody else and we considered this aspect as well in our gesture analysis (see Section 4.1, Figure 6b). In this paper, we first focus on deictic gestures and actions on the system. Hereafter, we also briefly describe emblems and adaptors as they may also come up as reactions and should be considered at a later stage.

Emblems are those nonverbal acts (a) which have a direct verbal translation usually consisting of a word or two, or a phrase, (b) for which this precise meaning is known by most or all members of a group, class, subculture, or culture, (c) which are most often deliberately used with the conscious intent to send a particular message to the other person(s), (d) for which the person(s) who sees the emblem usually not only knows the emblem's message but also knows that it was deliberately sent to him, and (e) for which the sender usually takes responsibility for having made that communication [25].

Adaptors are movements first learned as part of an effort to satisfy self needs or body needs, or to perform certain bodily actions, or to manage and cope with emotions, or to develop or maintain prototypic interpersonal contacts, or to learn instrumental activities [25].

As touch gestures, we regard all gestures that include touching a display independently of where or what this touch was targeted at. Our hypothesis regarding our aforementioned research question is that there are more touch gestures initiated by sharp pointing (Hypothesis 1), since this is the more stereotypical intentional type of gesture ([12]) compared to narrative and loose pointing.

2.2 Gaze and visual attention

In the next paragraphs, we review the literature on gaze at tangible and digital artifacts, since it is another important natural awareness cue about visual attention that can provide important insights about collaborative data visualization on wall displays. Gaze can be aligned or misaligned with hand gestures. Our focus is on pointing gestures; most often, gaze and pointing gestures are indeed aligned, since humans usually look at objects in one local area at a time. However, humans shift their gaze to scan the surrounding visual environment, because they cannot process all the

Table 1: Pointing categories on wall displays (excerpt from [24]).

Type	Hand usage	Referent	Duration
Narrative pointing (NP)	Pointing and moving index finger up/down or left/right	Full sentence/description of a value increase/decrease	Long
Loose pointing (LP)	Open palm, two fingers	Concept in total	Mid
Sharp pointing (SP)	Index finger	Specific value/text	Short

information simultaneously. Recently Lystbæk et al. [26] examined what happens when a target is presented in the peripheral visual field and made various experiments by manipulating the participants' head direction and fixation position: the head was directed to the fixation location, the target position, or the opposite side of the fixation. The performance was highest when the head was directed to the target position even when there was misalignment of the head and eye, suggesting that visual perception can be influenced by both head direction and fixation position [26].

Moreover, some studies proved that more fixations on a particular area indicate that it is more noticeable, or more important to the viewer than other areas. This is in line with "The More You Look The More You Get" paradigm [27] where users focusing their gaze on a specific work of art or part of it can be interested to receive some additional content about that specific item. Others [28] have highlighted wall-sized displays as a viable solution to present artworks that are difficult or impossible to move, and presented a Natural User Interface to explore 360° digital artworks shown on wall-sized displays. That solution allowed visitors to look around and explore virtual worlds using only their gaze, stepping away from the boundaries and limitations of keyboard and mouse.

Recently Cheng et al. [29] summarized empirical patterns of interdisciplinary work in organizational behavior, primatology, and social, developmental, and cognitive psychology to analyze visual attention as a window to leadership. Among others, they highlighted that shared gaze may facilitate teammate coordination and performance and eye gaze provides a reliable, behavioral, and under-utilized source of information about the hierarchical structure and functioning of a team [29]. The connection between eye gaze data and cognitive styles has been examined by Raptis et al. [30]. They revealed that individual differences in cognitive styles are quantitatively reflected on eye gaze data (gaze entropies, fixation duration and count) while users perform visual activities of varying type (e.g., visual search, visual decision-making) and varying characteristics. The authors suggested as the next step of this research to conduct more feasibility studies, considering other cognitive styles, activity characteristics, and application domains.

Moreover, Sharma et al. [31] investigated the causal relationship between individual and collaborative cognitive processes with gaze measures as a proxy to provide more insight into the collaborative learning process. They found that collaborative gaze patterns drive the individual focus when participants are engaged in problem-solving dialogues using an intelligent tutoring system and that the

nature of the causal relationship changes depending upon the context of the learning.

Furthermore, as far as gaze on displays is concerned, GazeProjector is a system that combines (1) natural feature tracking on displays to determine the mobile eye tracker's position relative to a display with (2) accurate point-of-gaze estimation [32]. A related work to our research is that of Lystbæk et al. [26], who suggested gaze-hand alignment as a principle for both modalities for pointing in Augmented Reality, and alignment of their input as selection trigger. Their idea is based on Zhai et al. [33] whose key idea is to leverage that the eyes naturally look ahead to a pointing target, followed by the hands. The results of our study, though, show many misalignments which come to contradict the statement of Zhai et al. [33], when it comes to large wall displays (see Section 4.2).

To sum up, gaze is often researched in relation both to the sensing (i.e. how it is measured) and the cue design aspects, but the literature gap is that gaze research often occurs in isolation and not with regards to its alignment with gestures. The research question that arises from the literature review and we seek to answer through our user studies is the following: "*When are pointing gestures misaligned with gaze?*" It is noteworthy that in the setting of large wall displays and collaborative decision-making between multiple users, there are certain challenges, such as the large size wall displays, the high amount of data being visualized, and crossing users in visual and peripheral field. Our hypothesis is that the gaze-pointing gestures misalignments happen when a group is split into subgroups (Hypothesis 2) and users have to share their attention between frontal viewing on displays and pointing gestures, because they cannot process all the information simultaneously ([26]).

3 Research method

In the following subsections we present our research method providing information on participants, apparatus, decision-making scenario, and analysis.

3.1 Participants

24 users, divided into six groups of four, participated in our study. Twenty of them were male, three female, and one participant preferred not to answer the question. Participants' ages ranged from 20 to 38 ($M = 24.2$, $SD = 4.5$, $Mdn = 22$). All participants were students in either computer science or geography. They were recruited through the University of Luxembourg and a Technical University in Luxembourg.

The study participants acted as a team of decision-makers from a hospital that needs to ensure that the stock of protective equipment meets the hospital's needs for the next three months. The participants needed to interpret and analyse the data, and identify the best solution given the existing constraints.

The study took place in consultation with our ethics committee. We explained to the participants what type of data we would record and how we would process it. To respect privacy, participants would immediately be assigned pseudonyms. We also informed participants about their rights to withdraw their consent and ask for the deletion of the data at any time and without giving reasons. We provided them with an information sheet and a consent form.

3.2 Apparatus

The experiment took place in our 360° Immersive Arena, 2 m high, composed of 12 screens (4 K resolution each) that are spatially positioned in a circle of 3.64 m diameter (see Figure 2a). Eight screens were used (therefore covering 240°) to display data visualizations. Three fixed cameras (top, front, and back cameras) were used to record the user study and the participants were using clip-on microphones for audio recording. Participants did not wear any motion tracking system or eye gaze trackers.

3.3 Decision-making scenario

The decision-making scenario had four distinct tasks: (i): estimate future COVID numbers, (ii) select the protective equipment to restock, (iii) select one offer in the overview,

and (iv) select delivery option. Each task included at least one type of (interactable) data visualisation, some of which are shown in Figure 2b. The current task was displayed on the first (leftmost) screen (see Figure 2a). When a group considered that a task was complete, they clicked on a confirmation button to continue to the next task.

3.4 Analysis

We performed our analyses on video recordings from the experiments, adding up to a total of 2 h 23 m 52 s. We analyzed and manually annotated a total of 578 pointing gestures, 201 touch gestures and 86 gaze misalignments made by participants using the ELAN software [34], as shown in Figure 3. Such gestures were then organized in sequences of annotations using the same software. We relied on Python scripts to process these annotations, compute the descriptive and general statistics (sums, percentages, averages) mentioned in this document, and used the Matplotlib [35] library to generate the corresponding plots. Both the annotations² and the notebook that contains the code used for computing statistics and generating visualisations³ are available online.

4 Results

In the following subsections we present our results based on descriptive statistics with regards to gesture sequences as well as gaze-pointing misalignments.

² <https://zenodo.org/records/10686346>.

³ <https://zenodo.org/records/10694228>.

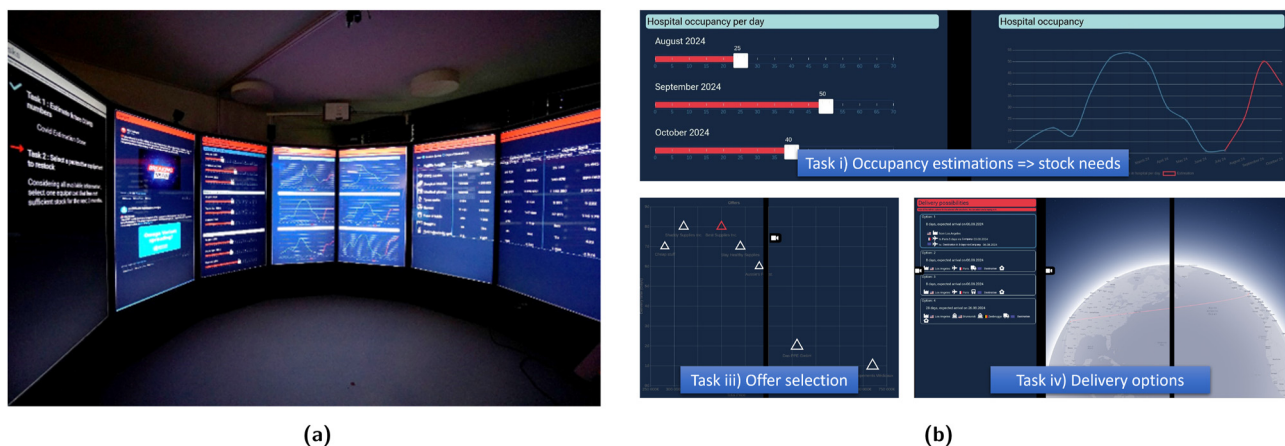


Figure 2: The mixed-presence collaboration setup and scenario used for the present study. (a) The 360° Immersive Arena, the multi-display setup we used for the study, consisting of 12 screens of 4 K resolution each, displaying the scenario with data visualizations. (b) A closer look at some of our data visualisations. The corresponding tasks are indicated on the Figure. Note that the black bars represent “bezels” i.e. gaps between adjacent screens in the display setup.

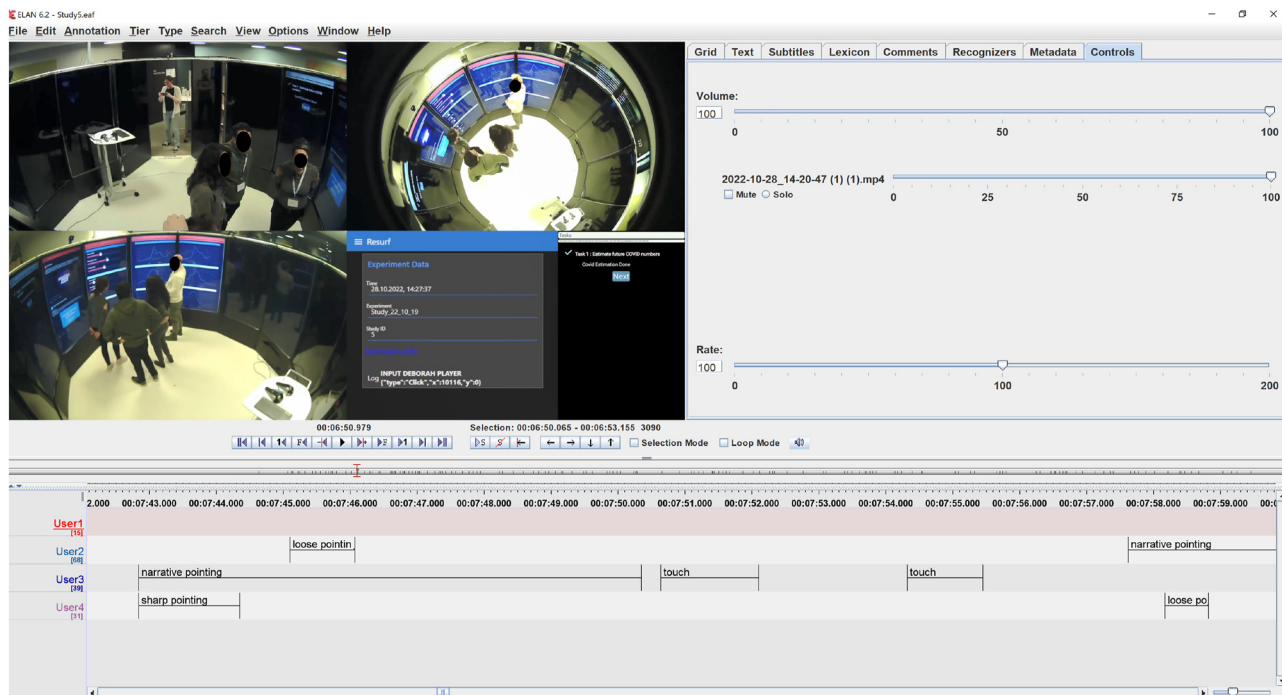


Figure 3: Our study analysis and annotation setup using the ELAN software.

4.1 Gesture sequences

Generally speaking, most of our annotated gestures are part of gesture sequences. If we include touch gestures, we had a total of 779 gestures, 64 % were part of gesture sequences and 36 % isolated gestures. The gesture sequences are generally short with no more than 9 gestures (see Figure 4).

In Figure 5a we see the distribution of all annotated gesture types. LP ranked first with 29 % followed by SP (27 %). Touch gestures come at the third place (26 %) and NP is the least common type of gesture at 18 %.

Figure 5b shows the share of the gesture types (NP, LP, SP) that served as initiators for the annotated gesture

sequences. The results show that SP is the most frequent initiator gesture type (45 % of all initiators were SP gestures, and 72/209 i.e. 34 % of all SP gestures were initiators of actual sequences, containing at least one follow-up gesture). LP and NP were indeed less frequent initiators, respectively accounting for 31 % and 24 % of the initiators, with 49/227 (22 %) of all LP gestures being initiators, and 39/142 (27 %) of NP gestures. This seems to indicate that SP gestures tend to produce more gestural reactions than other types of gestures, which may be explained by the fact that they typically concern a precise choice to be made, which may itself lead to discussions and counter-proposals by other team members.

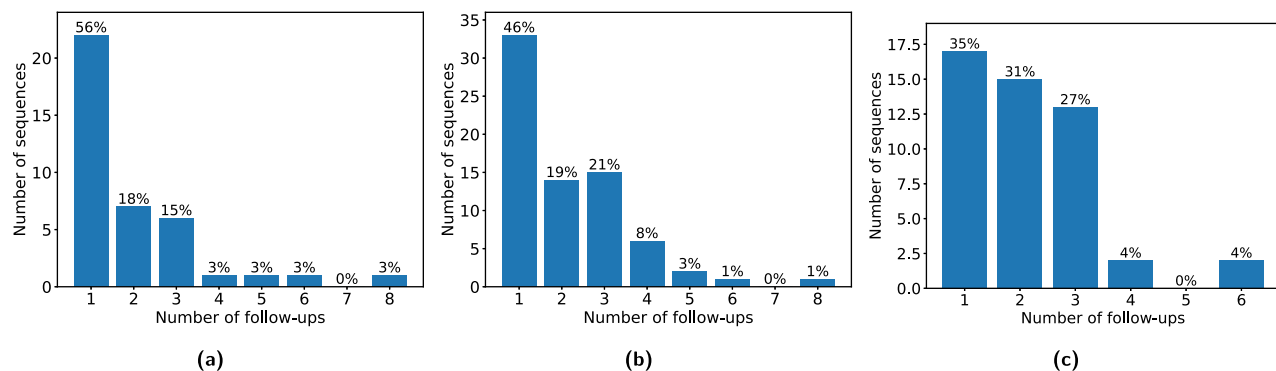


Figure 4: Length distributions of follow-up sequences of tracked gestures per initiator type. (a) Starting with a NP initiator gesture. (b) Starting with a SP initiator gesture. (c) Starting with a LP initiator gesture.

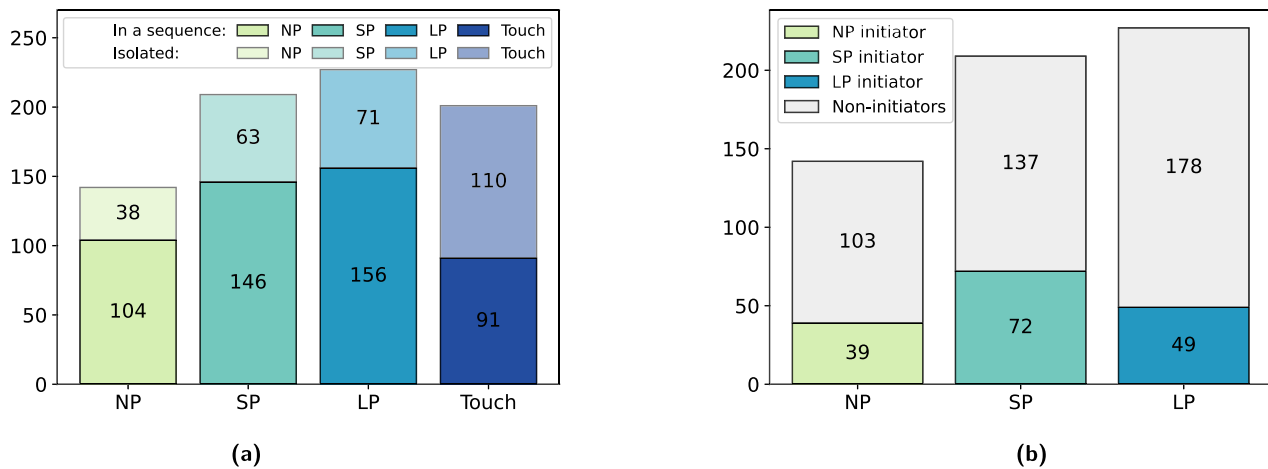


Figure 5: Distributions of annotated gesture types overall and amongst initiators. (a) Distribution of annotated gesture types. (b) Share of initiator gesture types for sequences of annotations.

Figure 4 depicts the length of sequences, which shows that sequences typically do not consist of many subsequent gestures. On average, NP-initiated sequences have 1.97 follow-up gestures, SP-initiated sequences 2.14, and LP-initiated sequences 2.16. Overall, the longest sequences are composed of 9 gestures, and 88 % (NP), 87 % (SP), and 92 % (LP) of the sequences are maximum 4 gestures long. As seen in Figure 5a, while the majority of gestures were part of a sequence, some were isolated. Isolated SP, LP, and NP gestures add up to 172 instances out of the total 578 pointing gestures (i.e. 30 %).

In Figure 6a we see the share of the follow-up gestures for sequences respectively having NP, LP, or SP as initiator. We can observe that there is no particular tendency

to keep the type of subsequent gestures the same as the initiator. However, it seems like more precise initiators (SP and NP) tend to produce more LP (that are less precise) follow-ups. This is especially the case for sequences starting with a SP initiator, the most precise type of pointing gesture, that produces 50 % LP follow-ups. Conversely, LP-initiated sequences are more often followed by the more precise gesture types (compared to SP- and NP-initiated sequences).

In terms of users producing subsequent gestures, we notice in Figure 6b that 24 % of the follow-ups after a NP initiator are made by the same user, which roughly corresponds to the “expected” 25 % (that would correspond to a random distribution of these follow-ups among the four participants). LP- and SP-initiated sequences however

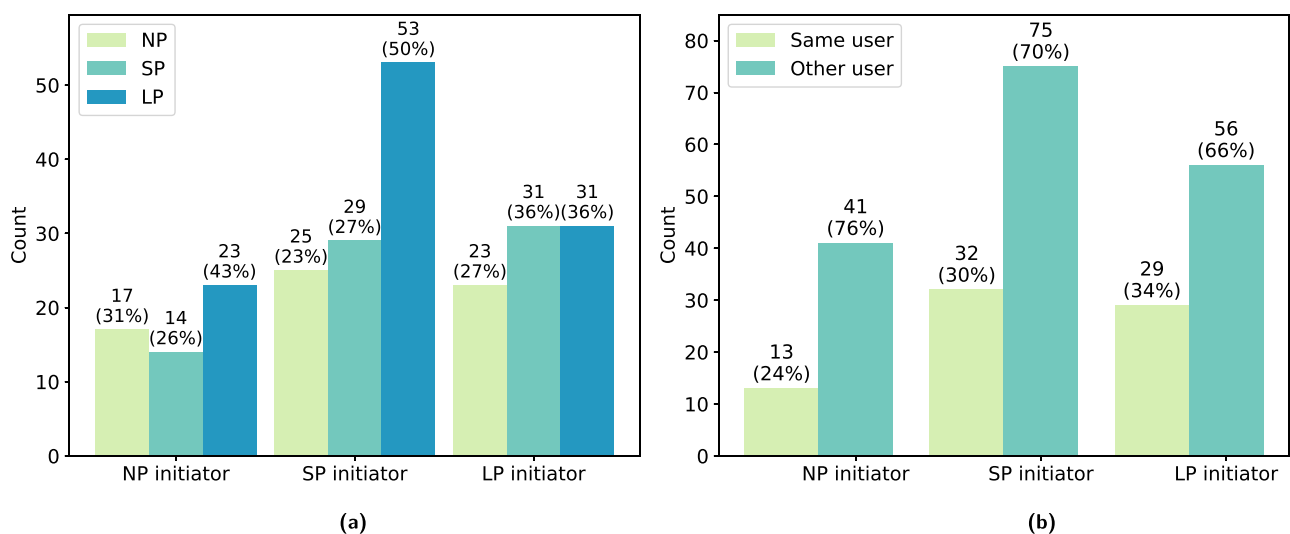


Figure 6: Follow-up gesture types and authors, depending on sequence initiator type. (a) Share of follow-up gesture types per initiator type. (b) Share of follow-up gesture authors per initiator type.

reach higher values (respectively 34 %, and 30 %) and therefore show a slight tendency to have more of the subsequent follow-ups made by the same author as the initiator, although a large portion of the follow-ups are still made by other users.

Another element we wanted to quantify relates to touch events. In fact, Figure 7 shows how many of the sequences led to a touch event (the touch does not need to be the last element of the sequence, but at least one touch event must be included for that sequence to be counted as leading to a touch). While sequences starting with a NP or SP gesture respectively led to a touch 46 % and 47 % of the time, LP-initiated sequences only led to a touch event 31 % of the time. We can deduce that more-precise gestures (SP and NP), which are likely associated with concepts being described in more details, tend to lead to subsequent actions on the system more often. LP regularly occurs as an unintentional communication artifact whereas NP and especially SP are typically intentional pointing gestures. This result is in line with our first hypothesis (see Hypothesis 1 in Section 2.1).

4.2 Gaze-pointing misalignment

In our study the participants were not wearing an eye-tracking device, which means that the precise Point of Gaze could not be obtained. However, for the purpose of our research, we can deduce certain results about gaze and pointing gestures by analyzing and manually coding the raw video data. As aforementioned, we manually annotated the head-pointing gestures misalignments. The results showed that out of the total of 578 pointing gestures (including SP, LP,

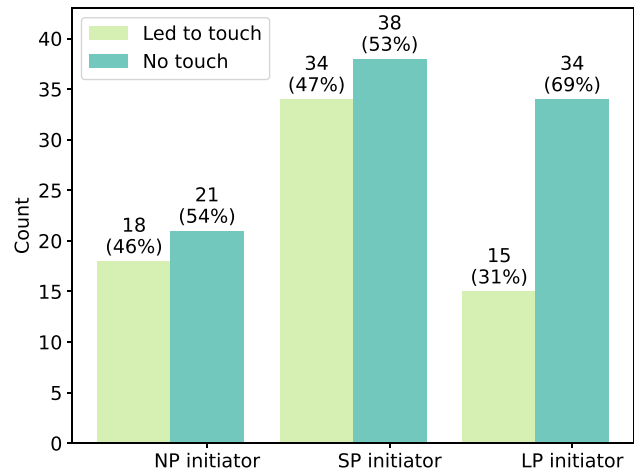


Figure 7: Share of sequences that led to at least one touch event.

and NP) that were performed during the study, in 86 cases (15 %) was the gaze misaligned with the pointing gesture.

We also observed that most misalignments were produced in relation to the screens with large amounts of data visualizations, such as options to touch/select from, e.g. kinds of protective equipment or delivery options. In these cases, the users sharply pointed (SP) and then touched the display while simultaneously looking at another display (the one closest on the right) to see the impact of their action on data visualizations.

In other fewer cases, users misaligned their gaze with their pointing gestures when they referred orally to the tasks related to the collaborative decision-making scenario (as in Figure 8a). In these cases, the user pointing to the



(a)



(b)

Figure 8: Examples of gaze-pointing misalignments. (a) Misalignment – far from screen. (b) Misalignment – close to screen.

display with the screens reminded the other participants of the current task. The most frequent pointing type when this happened was loose pointing, because the user did not refer to a specific element, but to the whole task-reminding screen in order to support their verbal comment.

In the rest of cases, the user's hand was held stretched towards the screen and while the user described a concept, they shifted their gaze towards a participant (example shown on Figure 8b). In these cases, the pointing was narrative. It should be noted that the misalignment here is not what usually happens in the retraction phase of a gesture, where a user shortly keeps the arm stretched, but we talk about a longer in duration, intentional gesture-gaze misalignment.

As a last point, the misalignments happened both close and far from the screen, so the position of the user therefore does not seem to be of much impact to misalignments.

5 Discussion and future work

The present research is part of the ReSurf project which seeks to enhance mixed-presence collaboration on distributed wall displays and investigates the use of collaborative awareness cues in this context. Through the exploratory case study described in this paper, we aimed to first observe what kind of awareness cues happen naturally onsite with collocated users. Based on the gained understanding, we aim to create in the future a technical system that automatically detects the most important awareness information and transmits it through pointers, icons, or annotations onto another wall display at a remote location (see Figure 1). Such cues have been explored and proposed for smaller workspaces in the past (desktops, tabletops, or physical tasks) but have only seldom been investigated in the context of remote collaboration across two or more wall displays.

There is a literature gap when it comes to empirical human computer interaction studies on gestures, and particularly mid-air gestures, at interactive wall displays. Therefore, in our case study we analyzed gesture sequences as well as gaze-pointing gestures misalignments under realistic collaborative conditions. The take-away messages based on the results of our user study are:

- (i) The majority of pointing gestures (64 %) are part of gesture sequences. These are generally short and no longer than 9 gestures.
- (ii) SP, as a more precise gesture, leads to touch gestures/actions on the wall display, possibly because they are associated with a more precise concept.
- (iii) More precise initiators (NP and especially SP) tend to produce less precise (LP) follow-ups.

- (iv) Gaze-gestures misalignments often happen when users touch the display to make a change and want to observe the effect of that change on another display.

Comparing our observed results to our hypotheses, Hypothesis 1 (see Section 2.1) is supported, because out of the 67 sequences that led to touch, 34 were initiated with an SP gesture, 18 with an NP gesture and 15 with an LP gesture. However, when comparing the distribution within sequence types, sequences initiated with a SP gesture led to a touch 47 % of the time, compared to NP-initiated sequences (46 %) and LP-initiated ones (31 %). Hypothesis 2 is not supported (see Section 4.2), since based our observations, groups did not split into subgroups.

The results described in this paper provided insights which help us prioritize and differentiate the synthetic awareness cues we will design to transmit the identified gesture types to other collaborators in a mixed-presence collaboration context. More specifically, our study has shown that most gestures are part of gesture sequences, which is in line with the statement of Hinrichs and Carpendale [23]. We distinguish our gesture sequences, because they can be standalone, from the “serial” ones of Morris et al. [10], which are regarded as a single, combined command. This suggests that, instead of simply transmitting synthetic cues on an individual basis, it might be interesting to look at transmitting a sequence of gestures as a single grouped synthetic cue. For instance, instead of showing a pointer for each individual pointing gesture, the system would, in case of a gesture sequence, show a line or path including all the follow-up gestures. This would help in limiting the clutter that would result from a high number of synthetic cues being displayed. Furthermore, since SP most frequently leads to touch actions on the wall, these seem to be most crucial as awareness information. As far as gaze is concerned, we revealed that narrative pointing is usually combined with intentional gaze misalignment and that the distance of the user from the screen does not impact the misalignments. In our study, there were indeed individual differences in cognitive styles with regards to visual decision-making as stated also by [30]. To sum up, the results have shown that pointing is not always aligned with gaze, and that therefore, both channels provide distinct information that needs to be considered as complementary.

Regarding the limitations of this work, the main one is the lack of precise estimation both of the gesture points and the Point of Gaze (PoG); it would be possible to extract both of these through specific hardware, such as motion capture body suits and eye trackers. However, in order to avoid the obtrusiveness of such tracking devices and to create a

more realistic setting of collaborative decision-making, we decided not to use such trackers that would impact the type of gestures being made. We are currently experimenting with the Azure Kinect depth camera that allows to track multiple body skeletons in action and can be used to recognize specific hand gestures as well as general head gaze. Another limitation was that we only relied on one annotator. Although the manual annotation was made by an experienced annotator, all videos should be annotated by two or more annotators to increase the reliability of the results through a high inter-annotator agreement ([36]).

This work is a first step into understanding gesture sequences at wall-sized interactive displays. Future work can build upon our results to measure the impact of gaze-gesture misalignments in relation to the response time and to identify what types of tasks and which kinds of display setups lead to an increase in the number of misalignments.

Future work should also evaluate whether the type of task, prior familiarity of users with each other, and task completion time had an impact on the gesture sequences that were produced. Furthermore, it should consider gesture sequences having touch as initiator gesture or also other mid-air gestures, such as emblems or adaptors. Our work showed that many touch gestures were isolated (110/201 i.e. 55 % were made without prior pointing gestures) and this needs further analysis when exactly this happened, by considering other natural awareness cues e.g. joint attention or agreement gestures. Looking at the order of follow-up gestures would allow identifying whether the probability of performing a certain type of gesture varies depending on all preceding gestures of the sequence, and not only on the initiator gesture.

On a more general note, the emergence of Immersive Analytics (IA), i.e. within augmented reality (AR) and virtual reality (VR) environments, provides new opportunities for gesture data analysis. Li et al. [37] designed Gesture Explorer, an immersive visualisation tool that uses 3D spatial arrangements to support gesture analysis and grouping in gesture elicitation studies. In this concept, we will also explore gaze in VR/AR following the principles that bi-directional gaze visualization influences symmetric collaboration ([38]).

In this paper, we showed how mutual awareness is shared in a collocated setting and the results help us create synthetic awareness cues that can enhance collaborative decision-making in a distributed wall-display setup. A comparative user study about remote collaboration across two physically distributed wall displays is already planned. In our study, we used a scenario of juxtapositioning data from various sources, as it is a common scenario that can be

found in many different domains and involves both low and high complexity. Through the research presented in this paper, we contributed to a next generation of mixed-presence decision-making tools, where people can collaborate smoothly in the context of data intensive decision-making, and enjoy an experience that is as close as possible to collocated collaboration.

Acknowledgments: We would like to thank Hoorieh Afkari, Johannes Hermen, Christian Moll, and Lou Schwartz for their contributions in developing the scenario and conducting the study. Furthermore, we thank all participants to our user study.

Research ethics: All ethical procedures have been strictly followed during the user study. The project ReSurf followed the fundamental ethical principles of the European Code of Conduct for Research Integrity and the LIST Code of Ethics throughout all stages of our research process, including planning, undertaking research and publishing results. Prior to the main user studies, the planned procedures were reviewed by LIST's Ethical Committee and by the Data Protection Officer to ensure GDPR compliance. The primary consideration were the rights, dignity, safety, health, and welfare of all the participants. Before proceeding to the data collection, consent in written form was asked from all participants regarding storage, processing, and use of the audio-visual recordings. There was no payment, inducement, or other financial-like benefits given to participants and the confidentiality of personal data of participants was ensured through specific technical means.

Author contributions: The author(s) have (has) accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: The author(s) state(s) no conflict of interest.

Research funding: This research is funded by the Luxembourg National Research Fund (FNR) under the FNR CORE ReSurf project (Grant nr C21/IS/15883550).

Data availability: The raw data (annotations) can be obtained on request from the corresponding author.

References

1. Jakobsen M. R., Hornbæk K. Up close and personal: collaborative work on a high-resolution multitouch wall display. *ACM Trans. Comput.-Hum. Interact.* 2014, 21, 1–34.
2. Langner R., Kister U., Dachsel R. Multiple coordinated views at large displays for multiple users: empirical findings on user behavior, movements, and distances. *IEEE Trans. Visualization Comput. Graphics* 2019, 25, 608–618.
3. Prouzeau A., Bezerianos A., Chapuis O. Towards road traffic management with forecasting on wall displays. In *Proceedings of*

- the 2016 ACM International Conference on Interactive Surfaces and Spaces: Nature Meets Interactive Surfaces, ISS 2016: New York, NY, USA, 2016; pp. 119–128.
4. Buxton W., Fitzmaurice G., Balakrishnan R., Kurtenbach G. Large displays in automotive design. *IEEE Comput. Graph. Appl.* 2000, 20, 68–75.
 5. Simonsen J., Karasti H., Hertzum M. Infrastructuring and participatory design: exploring infrastructural inversion as analytic, empirical and generative. *CSCW* 2020, 29, 115–151.
 6. Kubicki S., Guerriero A., Schwartz L., Daher E., Idris B. Assessment of synchronous interactive devices for BIM project coordination: prospective ergonomics approach. *Autom. Constr.* 2019, 101, 160–178.
 7. Yuill N., Rogers Y. Mechanisms for collaboration. *ACM Trans. Comput.-Hum. Interact.* 2012, 19, 1–25.
 8. Gutwin C., Greenberg S. A descriptive framework of workspace awareness for real-time groupware. *Comput. Support. Coop. Work* 2002, 11, 411–446.
 9. Carfi A., Mastrogiovanni F. Gesture-based human-machine interaction: taxonomy, problem definition, and analysis. *IEEE Trans. Cybern.* 2021, 53, 497–513.
 10. Morris M. R., Huang A., Paepcke A., Winograd T. Cooperative gestures: multi-user gestural interactions for co-located groupware. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 2006; pp. 1201–1210.
 11. McNeill D. *Hand and Mind: What Gestures Reveal about Thought*; University of Chicago Press: Chicago, 1992.
 12. Kita S. *Pointing: Where Language, Culture, and Cognition Meet*; Lawrence Erlbaum Associates: Mahwah, 2003.
 13. Shapiro L. *Embodied Cognition*; Routledge: London, 2019.
 14. Soni N., Darrow A., Luc A., Gleaves S., Schuman C., Neff H., Chang P., Kirkland B., Alexandre J., Morales A., Stofer K. A., Anthony L. Affording embodied cognition through touchscreen and above-the-surface gestures during collaborative tabletop science learning. *Int. J. Comput.-Support. Collab. Learn.* 2021, 16, 105–144.
 15. Goldin-Meadow S., Beilock S. L. Action's influence on thought: the case of gesture. *Perspect. Psychol. Sci.* 2010, 5, 664–674.
 16. Alibali M. W., Spencer R. C., Knox L., Kita S. Spontaneous gestures influence strategy choices in problem solving. *Psychol. Sci.* 2011, 22, 1138–1144.
 17. Chu M., Kita S. The nature of gestures' beneficial role in spatial problem solving. *J. Exp. Psychol. Gen.* 2011, 140, 102–116.
 18. Lozano S. C., Tversky B. Communicative gestures facilitate problem solving for both communicators and recipients. *J. Mem. Lang.* 2006, 55, 47–63.
 19. Çapan D., Furman R., Göksun T., Eskenazi T. Hands of confidence: when gestures increase confidence in spatial problem-solving. *Q. J. Exp. Psychol.* 2024, 77, 257–277.
 20. Liu C., Chapuis O., Beaudouin-Lafon M., Lecolinet E. CoReach: Cooperative gestures for data manipulation on wall-sized displays. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017; pp. 6730–6741.
 21. Villarreal-Narvaez S., Vanderdonckt J., Vatavu R.-D., Wobbrock J. O. A systematic review of gesture elicitation studies: what can we learn from 216 studies? In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, 2020; pp. 855–872.
 22. Wittorf M. L., Jakobsen M. R. Eliciting mid-air gestures for wall-display interaction. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, 2016; pp. 1–4.
 23. Hinrichs U., Carpendale S. Gestures in the wild: studying multi-touch gesture sequences on interactive tabletop exhibits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011; pp. 3023–3032.
 24. Maquil V., Anastasiou D., Afkari H., Coppens A., Hermen J., Schwartz L. Establishing awareness through pointing gestures during collaborative decision-making in a wall-display environment. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023; pp. 1–7.
 25. Ekman P., Friesen W. V. The repertoire of nonverbal behavior: categories, origins, usage, and coding. *Semiotica* 1969, 1, 49–98.
 26. Lystbæk M. N., Rosenberg P., Pfeuffer K., Grønbæk J. E., Gellersen H. Gaze-hand alignment: combining eye gaze and mid-air pointing for interacting with menus in augmented reality. *Proc. ACM Hum.-Comput. Interact.* 2022, 6, 1–18.
 27. Milekic S. The more you look the more you get: intention-based interface using gaze-tracking. In *Museums and the Web 2003*; Archives & Museum Informatics: Toronto, 2003.
 28. Calandra D. M., Di Mauro D., Cutugno F., Di Martino S. Navigating wall-sized displays with the gaze: a proposal for cultural heritage. In *Proceedings of the 1st Workshop on Advanced Visual Interfaces for Cultural Heritage*; CEUR-WS, 2016; pp. 36–43.
 29. Cheng J. T., Gerpott F. H., Benson A. J., Bucker B., Foulsham T., Lansu T. A., Schülke O., Tsuchiya K. Eye gaze and visual attention as a window into leadership and followership: a review of empirical insights and future directions. *Leadersh. Q.* 2022, 34, 101654.
 30. Raptis G. E., Katsini C., Belk M., Fidas C., Samaras G., Avouris N. Using eye gaze data and visual activities to infer human cognitive styles: method and feasibility studies. In *proceedings of the 25th conference on user modeling, Adaptation and Personalization*, 2017; pp. 164–173.
 31. Sharma K., Olsen J. K., Alevén V., Rummel N. Measuring causality between collaborative and individual gaze metrics for collaborative problem-solving with intelligent tutoring systems. *J. Comput. Assist. Learn.* 2021, 37, 51–68.
 32. Lander C., Gehring S., Krüger A., Boring S., Bulling A. Gaze projector: accurate gaze estimation and seamless gaze interaction across multiple displays. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, 2015; pp. 395–404.
 33. Zhai S., Morimoto C., Ihde S. Manual and gaze input cascaded (MAGIC) pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1999; pp. 246–253.
 34. Wittenburg P., Brugman H., Russel A., Klassmann A., Sloetjes H., ELAN: A professional framework for multimodality. In *5th International Conference on Language Resources and Evaluation (LREC 2006)*, 2006; pp. 1556–1559.
 35. Hunter J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* 2007, 9, 90–95.
 36. Artstein R., Poesio M. Inter-coder agreement for computational linguistics. *Comput. Linguist.* 2008, 34, 555–596.

37. Li A., Liu J., Cordeil M., Topliss J., Piumsomboon T., Ens B. GestureExplorer: immersive visualisation and exploration of gesture data. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023; pp. 1–16.
38. Jing A., May K., Lee G., Billinghamurst M. Eye see what you see: exploring how bi-directional augmented reality gaze visualisation influences co-located symmetric collaboration. *Front. Virtual Real.* 2021, 2, 697367.

Supplementary Material: This article contains supplementary material (<https://doi.org/10.1515/icom-2023-0037>)

Bionotes



Dimitra Anastasiou
Luxembourg Institute of Science and Technology, Human Modelling Group, Esch-sur-Alzette, Luxembourg
dimitra.anastasiou@list.lu
<https://orcid.org/0000-0002-0198-3729>

Dr. Dimitra Anastasiou (female) is a Senior R&T Associate. She holds her doctoral degree since 2010 and has worked in 10 national and European research projects since then. Overall, she has expertise in Computational Linguistics and Language Technologies as well as in Human-Computer Interaction, Tangible Interfaces, and User-Centred Design. She has also been editor of 8 workshop proceedings. She is member of several programme committees of renowned journals and conferences and is often invited to review project proposals.



Adrien Coppens
Luxembourg Institute of Science and Technology, Visualisation and Interaction Group, Esch-sur-Alzette, Luxembourg
adrien.coppens@list.lu
<https://orcid.org/0000-0002-2841-6708>

Dr. Adrien Coppens (male) is a Junior R&T Associate. He holds a PhD in Computer Science from the University of Mons, where he worked on the integration of immersive technologies for architectural design. His current research is focused on interconnecting large displays. More specifically and as part of the ReSurf project, he looks at ways to preserve the awareness information that collocated collaborators can naturally gather in a remote setting.



Valérie Maquil
Luxembourg Institute of Science and Technology, Visualisation and Interaction Group, Esch-sur-Alzette, Luxembourg
valerie.maquil@list.lu
<https://orcid.org/0000-0002-0198-3729>

Dr. Valérie Maquil (female) is Senior R&T Associate at the Visualisation and Interaction research group at LIST. Her research focus is on tangible, tabletop and large surface interaction in the context of collaborative problem-solving, decision-making and learning. She has more than 50 publications, many in the top conferences and journals of these fields and continuously serves in program committees. Currently, she is leading the FNR CORE project ReSurf which designs awareness cues in interactive wall displays for mixed-presence collaboration.