



**HAL**  
open science

# Eliciting Multimodal and Collaborative Interactions for Data Exploration on Large Vertical Displays

Gabriela Molina León, Petra Isenberg, Andreas Breiter

## ► To cite this version:

Gabriela Molina León, Petra Isenberg, Andreas Breiter. Eliciting Multimodal and Collaborative Interactions for Data Exploration on Large Vertical Displays. *IEEE Transactions on Visualization and Computer Graphics*, 2024, 30 (2), pp.1624-1637. 10.1109/TVCG.2023.3323150 . hal-04365019

**HAL Id: hal-04365019**

**<https://inria.hal.science/hal-04365019v1>**

Submitted on 11 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Eliciting Multimodal and Collaborative Interactions for Data Exploration on Large Vertical Displays

Gabriela Molina León, Petra Isenberg, *Member, IEEE*, and Andreas Breiter

**Abstract**—We examined user preferences to combine multiple interaction modalities for collaborative interaction with data shown on large vertical displays. Large vertical displays facilitate visual data exploration and allow the use of diverse interaction modalities by multiple users at different distances from the screen. Yet, how to offer multiple interaction modalities is a non-trivial problem. We conducted an elicitation study with 20 participants that generated 1015 interaction proposals combining touch, speech, pen, and mid-air gestures. Given the opportunity to interact using these four modalities, participants preferred speech interaction in 10 of 15 low-level tasks and direct manipulation for straightforward tasks such as showing a tooltip or selecting. In contrast to previous work, participants most favored unimodal and personal interactions. We identified what we call *collaborative synonyms* among their interaction proposals and found that pairs of users collaborated either unimodally and simultaneously or multimodally and sequentially. We provide insights into how end-users associate visual exploration tasks with certain modalities and how they collaborate at different interaction distances using specific interaction modalities. The supplemental material is available at [https://osf.io/m8zuh/?view\\_only=34bfd907d2ed43bbe37027fdf46a3fa](https://osf.io/m8zuh/?view_only=34bfd907d2ed43bbe37027fdf46a3fa).

**Index Terms**—Multimodal interaction, collaborative work, large vertical displays, elicitation study, spatio-temporal data.

## 1 INTRODUCTION

THE standard mouse and keyboard devices used to interact with desktop computers are not as well suited to interact with large vertical displays due to the larger screen size, potentially changing distances between the users and the screen [1], and collaborative work scenarios that require awareness of each others' actions [2]. By *large vertical displays*, we refer to displays fixed in their vertical position and significantly larger than desktop displays. In our work, we set out to explore alternative interaction modalities: touch, pen, speech, and mid-air gestures. We were specifically interested in multimodal interaction: interaction where these four types of input can be used in combination to perform certain actions. Combining interaction modalities can have various benefits, such as allowing to support user input from different distances, offering multiple degrees of freedom, and providing better support for particular tasks [3], [4], [5]. The best possible combinations of these interaction modalities for large vertical displays, however, are not immediately obvious: touch and pen interaction require standing close to the screen, and speech and mid-air gestural interaction are often not easily discoverable [6]. Also, we do not yet know how specific interaction modalities should be combined for different tasks and to support collaborative work. The preference for specific modality combinations and the order in which they are used may change depending on the modalities and tasks [7]. As large vertical displays provide more interaction space, group work is an important scenario to consider. Combining

multiple modalities and users leads to a more complex scenario. As such, many questions are still open in the space of multimodal interaction for large vertical displays. Here, we focus on the following two research questions:

- RQ1 What interaction modalities are preferred for exploring data visually on a large vertical display?
- RQ2 How can groups benefit from using multimodal interaction for collaborative data exploration?

In order to explore these questions, we conducted and analyzed an interaction elicitation study with 20 participants — in groups of two — in which we asked them to come up with interaction proposals for 15 tasks, giving the option of using touch, pen, speech, and mid-air gestures. With this methodology, we examine what end-users propose intuitively and assess the elicited interactions.

We found that people preferred unimodal interactions with either speech, touch, or pen to perform the exploration-focused tasks we gave them, which differs from previous findings about multimodality being preferable [7], [8]. When acting with more than one modality, participants opted for using touch first and speech later. When collaborating, participants worked closely together and acted either simultaneously using the same modality or in a sequence with two different modalities. With our work, we provide design knowledge on user preferences for multimodal and collaborative interactions with large vertical displays. We contribute the elicited gesture set, the top proposals, the interaction patterns, and our analysis on what interaction modalities were chosen in specific scenarios.

- G. Molina León and A. Breiter are with the University of Bremen, 28359 Bremen, Germany. E-mail: {molina, abreiter}@uni-bremen.de.
- P. Isenberg is with the CNRS, Inria, LISN, Université Paris-Saclay, 91405 Orsay, France. E-mail: [petra.isenberg@inria.fr](mailto:petra.isenberg@inria.fr).

Manuscript received February 21, 2023; revised July 12, 2023.



Fig. 1. The two most common collaborative interactions. (a) One participant selects a view element via touch first. Then, the other participant indicates the annotation text via speech. (b) Two participants use the pens simultaneously to annotate.

## 2 BACKGROUND: ELICITATION STUDIES

Past work has proposed multimodal interactions based on intuition and related work (e.g. [3], [9]). Here, we tackle the subject with a different methodology called the *elicitation study*. The elicitation study methodology was first proposed by Wobbrock et al. [10]. It is an interaction design methodology in which end-users are presented with the effect of an action on a computing system and are asked to propose the action to trigger the effect. The effect of the interaction is known as the *referent* and the proposed command or gesture is known as a *symbol* [10]. After the elicitation, symbols are classified into clusters of *signs* based on their similarity [11]. Elicitation studies are mainly used to inform the design of interactions for a system [12]. The main outcome of elicitation studies is the *consensus set* which is the set of interaction proposals that reached the highest agreement per referent [13]. Usually, it is called *consensus gesture set* because standard elicitation studies tend to be about mid-air gestures. In this article, we refer to it as *consensus set* because our study involves multiple modalities. Recently, Villarreal-Narvaez et al. [13] conducted a literature review on elicitation studies. Based on their findings, they suggest future studies explore other modalities besides mid-air gestures and elicit more than one symbol per referent, to investigate further the design space for interacting with smart environments.

The first elicitation study on multimodal interaction was conducted by Morris [14], without data visualizations. She elicited voice and mid-air gesture commands for interacting with a web browser on a living room TV and found that gestures had more commonalities among participants than speech. The results suggest that specific modalities fit better for certain referents. Willett et al. [15] conducted the first elicitation study for post-WIMP interaction with data visualizations. The researchers elicited multi-touch gestures for selection in four types of data charts. They found that participants strongly preferred simple, one-handed selection gestures, mainly using only one finger. According to Lee et al. [16], an open research direction for post-WIMP interaction with data visualizations is exploring creative adaptations from broader human-computer interaction (HCI) research. As

the authors suggested, we take this successful HCI method — the elicitation study — to investigate multimodal and collaborative interactions for data visualizations.

### 2.1 Benefits of Elicitation Studies

Researchers have cited multiple benefits of elicitation studies. Elicitation studies allow us to understand user proclivities and preferences for interactive technologies [11]. They serve not only to define a set of preferred interactions but also to characterize the diversity of the proposed interactions, aiming to understand better how people associate (or not) some types of interactions with specific tasks. Elicitation studies are considered a type of participatory design [14], as they allow end-users to get closely involved in the design process of interactive systems. Although designing with end-users may be more complex and time-consuming than the alternative, it leads to developing more usable and satisfying designs [17]. User-defined gestures tend to be preferred and more memorable than gestures predefined by a professional designer [18]. In the experiment of Nacenta et al. [19], participants considered the user-defined gestures less effortful and less time-consuming. In the field of HCI, more generally, the design of novel interactive systems is often based on elicitation studies [20], as eliciting interactions without the technical limitations of a gesture recognizer facilitates the exploration of the design space.

### 2.2 Challenges of Elicitation Studies

One of the main challenges of elicitation studies is legacy bias. This type of bias describes the tendency for users to propose commands they know from previous interaction experiences. Morris et al. [21] recommend three techniques to reduce legacy bias: *production*, *priming*, and *partners*. We follow their recommendation by applying these three techniques in our study, as explained in Section 4.6. However, this bias is not always seen as a disadvantage [22]. Legacy interactions can be more discoverable and therefore lead to a consensus set that feels intuitive to the users.

Tsandilas and Dragicevic critiqued that the standard formulas used for agreement calculation in elicitation studies (e.g. [18], [23]) do not consider *chance agreement* [12]. Chance agreement is the likelihood that two or more participants propose the same type of interaction by chance. While Tsandilas proposed agreement indices that take chance agreement into account [24], these indices do not consider our scenario where participants make more than one proposal per referent. More recently, Vatavu and Wobbrock argued that chance agreement should not affect agreement but also focuses only on studies with single proposals [11].

Other challenges may arise when applying the findings of an elicitation study to a real-world system. For example, technical limitations may prevent the detection of the elicited interactions, or these interactions may conflict with other existing interactions in the system. Those are issues that the researchers do not necessarily encounter in the study, as the methodology does not consider implementation details. Instead, the results of the elicitation study are meant to serve as a basis for navigating the design space of interaction techniques for new systems. Moreover, having no technical limitations allows end-users to be more creative and, accordingly, to propose innovative ways of interaction.

### 3 RELATED WORK

In this section, we present related work on interacting with large vertical displays, and collaborative data exploration.

#### 3.1 Interaction design for Large Vertical Displays

Working on large vertical displays has various benefits and challenges. While the display size and resolution facilitate sensemaking [25] and collaborative work [26], the extreme viewing angles up close can impact perception accuracy for certain data encodings [27], and users may have difficulty reaching some display areas [28]. Consequently, researchers have investigated multiple ways of interacting with these displays: direct manipulation through touch [28] or pen [29], gaze [30], using mobile devices, such as smartwatches [31], tablets [32], and augmented reality displays [33], through mid-air gestures [34] and body movements [3]. Other researchers, like Baudisch et al. [35], have proposed to apply focus and context techniques to visualize information at different resolution levels without the need for additional actions. This diversity is reasonable as users often physically move in front of the screen [36]: they tend to stand far from it to get an overview and move closer to access the details [31]. As such, supporting interaction modalities that allow both close-up and distant interaction is crucial. We explore four possible modalities: the use of speech and mid-air gestures from afar as they have proved helpful in other contexts (e.g., [3], [37]) and do not require additional screens—and pen and touch for close-up interaction.

We are not the first to propose using touch and pen interaction for visualization. Lee et al. [38] proposed leveraging touch and pen interaction for authoring and annotating visualizations. While touch and pen are often used interchangeably, touch is more pervasive thanks to the popularity of smartphones, but the pen is more precise, as it does not have the fat-finger problem [39]. Walny et al. [40] found

that although touch is preferred to move objects, both pen and touch were used for selecting menu items. Badam et al. [3] proposed using mid-air gestures and proxemics for visual exploration with interactive lenses. They found that people preferred using proxemics for navigation and mid-air gestures for “direct” actions, such as terminating a lens composition. Pointing from a distance, resembling a laser pointer, was considered a mid-air gesture that involved extending the hand using a special glove. However, these findings come from comparing pairs of interaction modalities (e.g., pen with touch, mid-air gestures with proxemics). We extend their work by considering four modalities and their combinations.

Previous work has already proposed ways of interacting with the visualization techniques we included in the study. For example, Drucker et al. [39] suggested sorting data items in a bar chart by dragging the finger along the corresponding axis, while Srinivasan et al. [37] recommended using speech commands for filtering. Nevertheless, we wanted to discover whether the study participants would propose similar actions or go in a different direction, given that they were free to choose among and combine multiple modalities. We also included the symbol map for which there are no multimodal interaction proposals yet.

Badam et al. [41] suggested mapping modalities to specific interaction techniques based on their affordances in immersive environments. Inspired by their work, we seek to identify the preferred modality combinations for visual exploration tasks. Srinivasan et al. [9] proposed to interact with unit visualizations on a vertical display, mixing touch, pen, and speech interaction. The authors recommend using direct manipulation to interact with single items and natural language to interact with item groups. However, more recently, experts preferred the pen over touch and speech for exploring data on tablets [42]. On a large display, we investigate user preferences about these three modalities combined with mid-air gestures, including more visualization techniques. While other modalities like gaze and proxemics are also worth investigating, we limit the scope of our research to four modalities, as it is already complex to consider them in combination with collaborative work.

#### 3.2 Collaborative data exploration

Isenberg et al. [43] define *collaborative visualization* as the “shared use of computer-supported, interactive visualizations by more than one person to perform joint information processing activities.” Collaboration can be co-located or distributed and synchronous or asynchronous. It can go from loosely coupled to closely coupled depending on how much information participants share and how much they interact with each other [44]. Collaborative systems should support not only *taskwork* (actions to complete the task) but also *teamwork* (actions to complete the task as a group) [45]. In this paper, we focus on the co-located scenario and study participant choices regarding timing, collaboration style, and the use of different interaction modalities for achieving taskwork and teamwork.

Interaction challenges on large displays during collaboration have been subject of research. While comparing

horizontal and vertical displays, Rogers and Lindley [46] found that vertical ones make it easier to show content to an audience. However, sharing devices is harder than on a horizontal display because it requires moving closer to the screen or a table to put the device down and give the opportunity to someone else to pick it up afterward. That may represent an added challenge for interactions that require an additional device, such as pen input. Based on their findings, Rogers and Lindley suggest providing the option of adding annotations and performing calculations directly on the vertical display to facilitate collaboration. The need for coordination [47] and privacy [48], for example, can impact interaction while sharing the screen space. In the experiment of Prouzeau et al. [47], pairs consistently divided space while working on a wall-sized display, even if the task was not spatially divisible. In the study of Isenberg et al. [2], participants solved interaction conflicts (e.g., two users trying to drag the same element) by talking or establishing rules. Moreover, participants asked for dedicated features to help group members be aware of what the others were doing. Adding annotations is a helpful awareness feature that we included in our study. Dostal et al. [49] proposed measuring user attention via gaze tracking to adapt the visualizations according to the status of each collaborator.

When working next to each other, participants can closely collaborate through cooperative gestures, i.e., gestures by multiple users that contribute to a single joint command [50]. Liu et al. [51] found that these gestures can reduce the physical effort required to manipulate data items on a large display. In the tabletop system Cambiera [52], Isenberg and Fisher proposed an interaction technique called *collaborative brushing and linking* that helped each user to be aware of the interactions of others in their personal views of the data. We elicit collaborative interactions to learn when end-users favor multi-user interactions and how they coordinate their work when interacting multimodally.

## 4 STUDY DESIGN

To conduct our elicitation study, we recruited researchers who explore data in their everyday work life. We asked them to participate in pairs and to brainstorm together about diverse ways to interact with data visualizations through touch, speech, pen, and/or mid-air gestures. While they could discuss all proposals together, the final proposals were individual and did not need to overlap. Working together and making multiple proposals per referent were two strategies to combat legacy bias [21] (see details in Sect. 4.6). We recruited participants who already worked together to add ecological validity to our study [14].

We conducted a pilot study with an additional pair of experts (P1 and P2) that helped to adjust the prompts and the minimum number of proposals required. Then, we proceeded to conduct the main study with 10 pairs of participants. We followed common practice in elicitation studies and recruited 20 participants [13]. The study took around 90 minutes for each pair.

### 4.1 Apparatus

We conducted the study with an 86-inch Promethean ActivePanel display of 4K resolution in a meeting room of 28

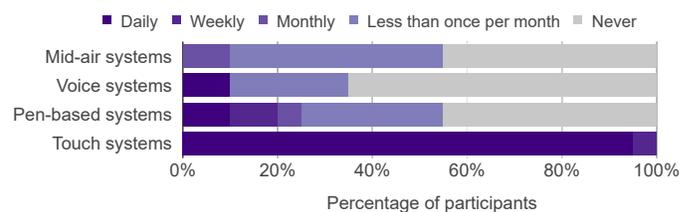


Fig. 2. Participants' reported prior experience with each of the tested interaction modalities, expressed by frequency of use.

$m^2$ . We created the visualizations in Python with the Plotly Express library [53]. The experiment was video recorded with the informed consent of the participants.

### 4.2 Participants

In total, we recruited 20 researchers and research assistants (eight female, aged 20–52) through university mailing lists. They had already worked with spatio-temporal data in diverse scientific domains, mainly in the social sciences. The main domains of expertise were political science and geography. There were 11 doctoral students, six postdoctoral researchers, two bachelor students, and one professor.

All participants reported that they interacted with data visualizations and explored spatio-temporal data as part of their job. Twelve of 20 participants worked with visualizations at least once per week. All pairs of experts were either working together or had collaborated in the past.

When asked about how frequently they had interacted with the proposed modalities, participants had most experience with touch and least experience with speech, as shown in Fig. 2. While everyone had experience with touch interaction, 45% had never interacted with mid-air gestures, another 45% had never used a pen as an input device, and 65% had never interacted with speech commands. When asked about their experience interacting with large vertical displays, 75% of the participants reported to have worked with them before. Everyone was right-handed except for two ambidextrous people.

### 4.3 Dataset

The data we used to create the visualizations was a set of development indicators published by the Gapminder foundation [54]. They included the life expectancy, GDP per capita, and population of 142 countries from 1952 to 2007. Given that the experts worked in different fields within scientific research, we decided to use a real-world dataset that everyone would understand to ensure ecological validity.

### 4.4 Referents

The referents in our study were low-level data interaction tasks. We chose low-level tasks because every more complex exploration task is composed of these low-level tasks and requires combinations of interactions to be completed. Specifically, we examine 15 low-level tasks relevant to the exploration of spatio-temporal data inspired by the typology of Andrienko and Andrienko [55]. We focused on tasks relevant for working collaboratively with a large vertical display based on the interaction taxonomy of Yi et al. [56], the task

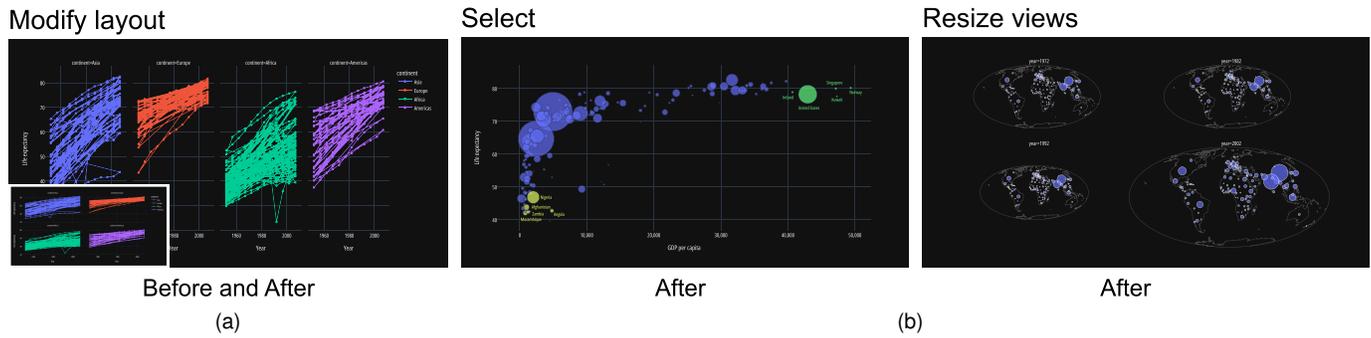


Fig. 3. (a) Prompts showing the data visualizations before and after the interaction for the *Modify layout* referent. (b) After prompts of the *Select* and *Resize views* referents.

TABLE 1  
Our 15 Referents and Their Classification. Citations Refer to the Corresponding Task Typologies and Taxonomies.

Referent	Task type
Show item details	Direct lookup [55], Abstract/Elaborate [56]
Select	Inverse lookup [55], Select [56], [57]
Deselect	Select [56], [57]
Activate B & L	Select [57], Connect [56]
Deactivate B & L	Select [57], Connect [56]
Data-centric filter	Filter [56], Behavior characterization [55]
View-driven filter	Filter [56], Pattern search [55]
Sort	Reconfigure [56], Arrange [57]
Change encoding	Encode [56], [57]
Merge views	Aggregate [57], Direct comparison [55]
Split view	Aggregate [57], Direct comparison [55]
Resize views	Abstract/Elaborate [56]
Modify layout	Reconfigure [56], Arrange [57]
Show regression line	Connection discovery [55]
Adding annotation	Annotate [57], Reconfigure graphics [41]

typology of Brehmer and Munzner [57], and the interactions with multiple coordinated views investigated by Langner et al. [58]. Six of our 15 referents were associated with managing multiple coordinated views (MCV): activating and deactivating brushing & linking (B & L), merging views, splitting a view, resizing views, and rearranging views. We added multi-selection and annotation authoring as referents to support visual awareness across users [52]. Moreover, we differentiate between data-centric and view-driven filtering based on the different interactions proposed by Sadana and Stasko [59].

We present the final list of referents and their classification according to related work in Table 1. We excluded zooming and panning from the list because previous work has consistently found successful interactions to perform them using touch [60], mid-air gestures [3], and proxemics [1]. The referent images we used in the experiment can be found in the supplemental material. The supplemental material is publicly available on OSF at [https://osf.io/m8zuh/?view\\_only=34bfd907d2ed43bbbe37027fdf46a3fa](https://osf.io/m8zuh/?view_only=34bfd907d2ed43bbbe37027fdf46a3fa).

#### 4.5 Visualization techniques

We chose the visualization techniques based on the recommendations of Andrienko et al. [61] for exploring spatio-temporal data. Accordingly, we included the following techniques: line charts, bar charts, scatterplots, bubble charts, and symbol maps. We selected these charts for the variety of

visual channels they use to encode data. Additionally, eleven referents involved multiple views to display spatial entities or temporal steps via a *small multiples* technique.

#### 4.6 Study procedure

The elicitation study was composed of four parts. First, we explained its structure to the participants and asked for their informed consent to record their interaction proposals through video and audio. Then, they filled out a demographics questionnaire that included questions about their previous experience with each of the interaction modalities.

To start the elicitation, we asked participants to picture themselves in a scenario where they wished to explore a new dataset together, and they needed to perform a series of actions as part of the exploration process. As our goal was to investigate how multimodal interaction can benefit group work (RQ2), it was important that the participants would see themselves as a team. We asked the experts to propose individually at least three interactions for each of the 15 referents, including at least one collaborative proposal and at least one that was multimodal. One of the three proposals being both collaborative and multimodal was also sufficient. We presented each referent graphically through a pair of images showing the visualization before and after the interaction (see examples in Fig. 3). We considered using animated prompts that would show transitions but decided against them to avoid biasing the participants by implicitly suggesting specific ways of interaction (e.g., resizing a view could start by dragging one corner if the animation showed the view getting enlarged in a specific direction first). For each referent, the experimenter read a question out loud presented above the pair of images of the form *How would you...?*, such as “How would you merge two views into one?”, before switching to the first image on full screen to start eliciting. We show how the elicitation took place in the supplemental video.

Participants were free to come up with any proposal that they felt was best suited without restricting themselves to any set of “allowed” interactions. There was no time limit. We set the order of the tasks according to how they complemented each other, e.g., deselect after select. We allowed participants to use any of the four interaction modalities: touch, pen, speech, and mid-air gestures. As we aimed to investigate what modalities were preferred (RQ1), participants could propose using any modality alone or combined with others.

Participants were also free to add interface elements to the screen if they wished to have them, so we could observe and analyze what they preferred. For each referent, we encouraged participants to consult with each other and to show their ideas by performing the corresponding actions. Each participant had to make their own proposals that could be similar to or different than those of their partner. For each referent, each person described their final proposals on paper and picked a favorite among them. Most participants started each task from a table around three meters away from the display where they had written down their proposals for the previous task. The experimenter then asked the participants to move back towards the display for each task but did not prescribe a specific starting distance to take on. During the study, participants could stand where they wanted and relocate freely.

We applied the Wizard of Oz technique for changing between the referent images (the *before* and *after* images) when participants made an interaction proposal to demonstrate the effect of the interaction [62]. We made clear that the technical interaction recognition of the system would hypothetically work perfectly. This was done to avoid that participants would not propose interaction techniques out of fear that they might not be technically realizable. The study concluded with a short questionnaire asking participants to rate the perceived effectiveness of each modality with a five-point Likert scale, as in the study of Morris [14].

To reduce legacy bias, we applied the *priming*, *production*, and *partners* techniques, as recommended by Morris et al. [21]. After answering the demographics questions, we primed participants by asking them to report three life situations where they had behaved creatively in the past, as suggested by Sassenberg and Moskowitz [63] and successfully tested in previous elicitation studies [64]. During the elicitation, we applied *production* by asking participants to produce at least three proposals for each referent [65]. Moreover, we asked them to make at least one multimodal proposal and at least one collaborative proposal, given the small number of multimodal interactions elicited in Morris' study [14]. We applied the *partners* technique by inviting the experts to brainstorm and interact in pairs. We asked them to come with someone they already knew or worked with.

#### 4.7 Data analysis

We first extracted the interaction proposals from the list that each participant wrote down during the experiment. Then, we completed or corrected the details of each proposal based on the video recordings that were analyzed by two researchers separately. For each proposal, we documented the referent, the participant, the sequence of steps and their modality, whether it was performed by one person or two (if two, whether the steps happened in parallel), and whether it was a favorite. Afterward, we grouped the interaction proposals into *signs* based on their similarity according to the modality, the data attributes, and the target involved.

For analyzing the signs, we calculated the metrics *max-consensus* and *consensus-distinct ratio* proposed by Morris [14] for each referent, and we report the consensus set based on the most popular proposal per referent, according to frequency. The max-consensus indicates the percentage of

TABLE 2  
Descriptive Statistics of the Proposals per Referent and Participant.

Metric	All	Multimodal (M)	Collaborative (C)	M & C
Min	3.00	1.00	1.00	0.00
Median	3.00	1.00	1.00	1.00
Mean	3.38	1.41	1.08	0.85
Std	0.61	0.60	0.28	0.43
Max	6.00	4.00	3.00	2.00

participants that proposed the most common interaction for a given referent. The max-consensus is 100% if all participants recommended the most common proposal. The consensus-distinct ratio indicates the proportion of distinct interactions proposed by a minimum number of participants (the *consensus threshold*). The consensus-distinct ratio is 1.0 when every interaction proposed for the referent is over the threshold. Although most researchers calculate an agreement score or rate for elicited interactions [13], our study included multiple proposals per participant and was conducted in pairs. Thus, it required different measures [11]. Vatavu [66] proposed other metrics for this type of studies, such as consensus and growth rate, but said calculations are based on spatio-temporal coordinates of body gestures and do not consider multiple interaction modalities.

## 5 RESULTS

We elicited a total of 1015 interaction proposals for the 15 referents. Each proposal included at least one step (action) completed with one modality. The multimodal and collaborative proposals included at least two steps taking place either sequentially or simultaneously. We present a summary of the proposals per referent and participant in Table 2. In elicitation studies involving speech input, the speech proposals whose text overlaps with the referent name are sometimes excluded because the participants tend to use those words first. We did not remove those commands, given that, in referents like *sort*, ignoring commands using that verb would radically limit the possibilities of appropriate terms [5]. Each participant successfully produced at least three interaction proposals per referent, with one person (P9) even proposing six ways for selecting. Participants made more multimodal proposals than required (424 instead of 300), leading to 42% of the elicited interactions being multimodal. They also proposed slightly more collaborative interactions than required (324 instead of 300), resulting in 32% of the interactions being collaborative.

After grouping the proposals into clusters based on their similarity (e.g., grouping speech commands such as “Deselect group 1” and “Deselect yellow group”), we identified 360 *distinct* interactions or *signs* among the 1015 proposals. Of the 360 signs, 215 were multimodal, and 161 were collaborative. More specifically, 127 proposals were both multimodal and collaborative (35.28%), 111 were unimodal performed by a single person (30.83%), 88 were multimodal and performed by a single person (24.44%), and 34 were collaborative and unimodal (9.44%).

TABLE 3

Consensus Set and Metrics per Referent for the Four Most Common Modality Combinations. MC Stands for Max-Consensus and CDR Stands for Consensus-Distinct Ratio. The Highest Value for Each Metric per Modality Combination Is in a Blue Cell.

Referent	Most Common Interaction	All		Speech only		Touch only		Touch-Speech		Pen only	
		MC	CDR	MC	CDR	MC	CDR	MC	CDR	MC	CDR
Show details	 Tap on mark	80.0%	0.45	55.6%	0.67	94.1%	0.50	44.4%	0.75	75.0%	0.50
Select	 Lasso around marks	40.0%	0.29	35.7%	0.29	50.0%	0.60	100.0%	0.50	66.7%	0.67
Deselect	 "Deselect group 1"	70.0%	0.11	77.8%	0.25	15.4%	0.00	28.6%	0.00	42.9%	0.33
Activate B & L	 "Extend yellow to all years"	60.0%	0.11	100.0%	1.00	30.0%	0.17	37.5%	0.17	50.0%	0.00
Deactivate B & L	 "Deselect group 1 except on 1992"	55.0%	0.09	68.8%	0.67	22.2%	0.00	25.0%	0.14	100.0%	0.00
Data-centric filter	 "Show only Asia"	75.0%	0.20	78.9%	0.67	87.5%	0.33	33.3%	0.29	-	-
View-driven filter	 "Show me the outliers"	50.0%	0.17	62.5%	0.25	57.1%	1.00	22.2%	0.00	66.7%	0.25
Sort	 "Sort by population"	60.0%	0.38	63.2%	1.00	45.5%	0.50	71.4%	0.33	100.0%	0.00
Change encoding	 "Set population size as point size"	65.0%	0.13	81.2%	0.25	15.4%	0.00	36.4%	0.29	50.0%	0.00
Merge views	 "Merge graphs"	75.0%	0.27	78.9%	0.67	57.1%	1.00	50.0%	0.00	100.0%	0.00
Split view	 "Split by country"	55.0%	0.18	68.8%	1.00	53.8%	0.25	100.0%	0.00	-	-
Resize views	 Pinch on top of the view	50.0%	0.19	69.2%	0.20	66.7%	0.50	40.0%	0.00	100.0%	0.00
Modify layout	 Drag a view	60.0%	0.18	68.8%	0.20	80.0%	0.67	-	-	-	-
Show regression line	 "Add regression line"	95.0%	0.21	95.0%	0.33	55.6%	0.20	38.5%	0.25	100.0%	1.00
Add annotation	 Write text with the pen	65.0%	0.35	50.0%	0.25	100.0%	0.00	60.0%	0.67	81.2%	0.50

## 5.1 The consensus set is unimodal and personal

On average, participants proposed 27 distinct interactions per referent. We present the *consensus set* of our study in Table 3, together with the metrics for the four most common modality combinations among the top proposals. As the standard agreement scores for elicitation studies do not consider the case of multiple proposals per referent [11], we calculated the top proposals based on their frequency and present the agreement metrics *max-consensus* and *consensus-distinct ratio* proposed by Morris for this case [14]. As mentioned in Sect. 4.7, the max-consensus indicates the percentage of participants that proposed the most common interaction for a given referent. A high consensus suggests that the interaction was considered the most intuitive for the task. Consistently, participants agreed most on interactions involving only one modality (speech, touch, and pen) and performed by a single person. Despite previous evidence suggesting that multimodal interaction may lead to a more fluid experience [37], participants preferred to explore the data with simple unimodal interactions. Yet, a system would require to support speech, touch, and pen input to include the most commonly proposed interactions. Accordingly, it has to enable both distant interaction and direct manipulation. Only speech was common for interacting from a distance and should likely be supported at a minimum, together with touch or pen for close interaction. Participants proposed more mid-air gestures than pen interactions, but only pen interactions made it to the consensus set.

For 10 of the 15 referents, participants preferred speech interaction. Overall, we noticed that when given the freedom to propose any interaction with any of the four modalities, participants often came up with a speech command as the first proposal. Accordingly, 591 of the 1015 interactions proposed (58.23%) included speech interaction. Of them,

243 (23.94% of all) were speech commands only. During the study, several participants commented that using speech felt like the easiest option. In contrast, the most common proposals for the referents *show details*, *resize views*, and *modify layout* were standard touch gestures, while the most common proposals for the referents *select* and *add annotation* were with pen interaction. For these five referents, depending on the task, about 20-60% (39% on average) of the first proposals made by the participants became part of the consensus set. Participants preferred direct manipulation for lookup tasks [55] and for modifying the position and size of the views. Speech interaction was instead favored for more abstract tasks aimed to find patterns across sets of data items, such as regression, and for synoptic tasks [55] related to comparing the sets across views. We did not find any evidence of a relationship between the visualization techniques we used and the interaction modalities of the consensus set.

On average, participants tended to agree most on speech interactions. Speech interactions had a mean max-consensus of 70%, mid-air gestures of 63%, pen interactions of 62%, and touch interactions had a mean max-consensus of 55%. While pen and mid-air interactions had a max consensus higher than touch, participants made no proposals at all using those modalities for three (data-centric filter, split view, and modify layout) and two referents (view-driven filter and add annotation), respectively. Multimodal proposals starting with a touch gesture, followed by a speech command, had a mean max-consensus of 46%. Touch-speech interactions were preferred over interactions using pen-only and mid-air gestures among the top proposals. We present more details about the multimodal proposals in Sect. 5.2. Overall, the proposal with the highest max-consensus (95%) was a speech command to apply a regression model.

Now we look at the consensus-distinct ratio, which gives a sense of the spread of the agreement. Unlike Morris [14], we used a *consensus threshold* of three instead of two because we elicited many more interactions due to applying the *production* principle [21], and pairs often agreed on their proposals after brainstorming, so reaching a consensus between two people was common. So the consensus-distinct ratio is 1.0 when every interaction proposed for that referent was proposed by at least three participants. On average, the interactions that made it to the consensus set were proposed by at least half the pairs in agreement (i.e. per referent, 5.6 groups proposed the same interaction twice), in contrast to each person proposing something different than their partner. The referent with the highest ratio was *show details* suggesting that most interaction proposals for invoking a tooltip reached a high agreement among the participants. The lowest ratio was 0.09 for *deactivate brushing & linking (B & L)* which suggests a higher diversity of proposals overall, with less agreement.

The difference between the metrics across the top four modality combinations suggests that participants mapped some referents to specific modalities. For example, the referent *activate B & L* reached the highest consensus among speech proposals, in contrast to a low one with other modalities. Some referents like *data-centric filter* and *show regression line* reached a high consensus with touch and pen, respectively. Still, the popularity of speech commands overall determined the top proposal of those referents. Accordingly, the second top proposals for *data-centric filter* and *show regression line* were touch-only and pen-only, respectively (see the list of top three proposals in the supplemental material). Moreover, although touch and pen both serve for direct manipulation, participants favored touch interaction for *resize views* and *modify view layout*.

### 5.1.1 Multimodal synonyms

When looking beyond the consensus set, we find more diversity regarding modalities and collaboration among the top proposals per referent (see the list of top three proposals). We detected what Morris [14] calls *multimodal synonyms* among the top proposals of the referents *show details*, *select*, and *activate B & L*. Multimodal synonyms are equivalent interactions with different modalities that participants propose as alternatives for the same command. For example, participants proposed to perform lasso selection with either the pen or touch. For brushing & linking, they wished to drag and drop a selected group of items via touch or a mid-air gesture. During the experiment, several participants commented that having modalities to choose from made the system more accessible and allowed them to select a modality depending on the situation.

### 5.1.2 Favorites matched consensus

We asked participants to propose at least three interactions for each referent and to select a favorite among them. We were interested in finding out what interactions participants considered best, given that the most common proposal in an elicitation study may not necessarily be considered the most appropriate in practice. However, when comparing the favorites with the consensus set, the most common interaction was also the most commonly named favorite

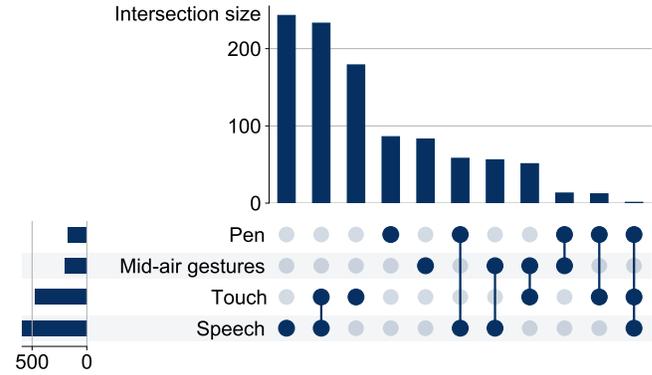


Fig. 4. UpSet plot showing the frequency of the different combinations of the four interaction modalities across the 1015 proposals.

interaction for all referents. The only exceptions were the multiple top favorite interactions for the referents *select* and *merge views*. For *select*, participants had four favorite proposals besides the most common one. The first was the collaborative version of the lasso selection with pen, the second was a query via speech command, the third was the collaborative version of the second, and the fourth was a multimodal and collaborative interaction with speech and touch. To *merge views*, participants favored dragging one view towards the other besides using a speech command. Although multimodal and collaborative interactions did not make it to the consensus set, we examine them in the following sections to investigate in which situations and how these interactions can support data exploration on a large vertical display.

## 5.2 Multimodal interaction: Mainly touch and speech

We show the distribution of the modality combinations among the 1015 proposals in Fig. 4. Speech and touch interaction were the most used interaction modalities, with the total of touch-only proposals even surpassing all the pen interactions and mid-air gestures combined. However, the second largest group of interactions was multimodal, combining touch and speech (23% of all proposals). Such multimodal commands also appeared in nine top proposals.

Participants tended to divide tasks into multiple steps and associate each step with a different modality. 61% of the proposals consisted of a touch gesture followed by a speech command. Using touch followed by speech was most suggested for *add annotation*, *show regression line*, and *change encoding*. Participants used touch first for choosing a view or data items of interest. Then, they specified an action to apply to them orally. For example, someone first tapped on a view element to select it and then added an annotation via voice, or they selected a view with touch first and then asked for the calculation of a regression model via speech. In the inverse order (39%), participants first used a speech command to select data items, activate a mode, or invoke a menu. Then, they performed the main action with touch. For example, using speech to activate a rearranging mode and then approaching the screen to drag and drop multiple views until reaching the desired view layout. Using speech

followed by touch was most common among the proposals for *resize views*, *modify layout*, and *select*.

Nineteen of 20 participants chose at least one multimodal interaction as their favorite. The single participant (P21) who did not choose any multimodal interaction as his favorite said, during the experiment, that he preferred the simplest interactions. In contrast, his partner (P22), always chose a multimodal interaction as his favorite. Among the favorite multimodal interactions, the most common combination was a touch gesture followed by a speech command (26%) and a speech command followed by a touch gesture (21%) in line with their frequent occurrence among the multimodal proposals.

### 5.3 Collaborative interactions

During data analysis, collaboration styles can range from being closely coupled to loosely coupled [44]. Most of the collaborative proposals we elicited belong to the closely coupled case as they involve two persons working together to perform a low-level task in a co-located scenario. There were a few exceptions showing loosely coupled collaboration, where the persons stood next to each other but worked in parallel interacting with different data items.

The most common collaborative proposals were two distinct interactions for the *add annotation* referent, shown in Fig. 1. These interactions demonstrate the patterns we found: the first consisted of a sequence where one person started by tapping on a bar inside a bar chart, and then, their partner used a speech command to attach an annotation to the bar. The second interaction involved two users writing different annotations simultaneously with the pen.

#### 5.3.1 Collaborative work was either unimodal and simultaneous or multimodal in sequence

We distinguish between two interaction types in collaborative proposals: *sequential* and *simultaneous* interactions. In the sequential case, one person performed the first step of the interaction, and their partner waited for that step to be over before proceeding to execute the next one. Although each step may have targeted different objects on screen, their actions were part of a single joint command. In the *simultaneous* case, both persons interacted simultaneously to perform two steps in parallel without conflict. Of the 161 distinct collaborative interactions, 90% were sequential, and 10% were simultaneous.

The sequential and simultaneous types of collaboration were often paired with specific modality combinations. When two participants interacted simultaneously, they mostly interacted using the same modality (63%). For example, to *resize views*, the third most common proposal was that both users drag a view border to adjust the size. Others performed mid-air gestures in synchrony to merge two views into one, as shown in the supplemental video.

In the sequential case, partners mostly interacted multimodally (83%). Each person became responsible for one modality. Overall, we identified three types of multimodal sequences in collaborative interactions. They demonstrated that groups often performed their actions at different distances: one person stayed close to the screen for direct manipulation, and the partner stood farther away and used speech or

mid-air gestures. The most preferred form of collaboration consisted of a two-step sequence where one person performed a touch gesture, and their partner interacted via voice afterward. Such interactions were proposed for all referents but happened most often for compound tasks [67]. Filtering is such a task that may seem to be a single entity, but in reality, it can be divided into two sub-tasks: choosing (or selecting) a set of items and then subtracting items based on the selection (as it works on Tableau [68]). Using touch followed by speech was most proposed for calculating a regression model, adding annotations, and data-centric filtering. The most common proposal of this kind was the tap-and-speech sequence to add an annotation, shown in Fig. 1a.

The second most common combination of multimodal sequences used speech, followed by touch. It was proposed most for tasks associated with managing multiple views. A person looking at the screen from afar spoke to select an element or activate a mode (e.g., modification mode). Then, the partner completed the joint action by dragging view elements with touch. The third combination involved touch and mid-air gestures. A person started the interaction by tapping or tap-and-holding to select a view element. Then, the partner, who stood further away from the screen, air-dragged other elements or performed a dedicated gesture, e.g., extending arms to split a view into two.

#### 5.3.2 Collaborative synonyms

When comparing interactions of single users and pairs, we identified what we call *collaborative synonyms*. Among the distinct interactions, we often found proposals that were identical except for being done by a single participant or by two people. Sixty-six interactions followed this pattern: 33 described a sequence of steps performed by a single person, and for each one, an equivalent existed, performed by two people. We show two collaborative synonyms illustrating that in Fig. 5. We found at least one example of this case for every referent, mainly for tasks associated with selection and using the B & L technique. There are two pairs of collaborative synonyms among the top proposals: one to *change encoding* and one to *add annotation*. Most synonyms were multimodal, and the collaborative version meant that the second person would introduce the second modality.

### 5.4 Perceived effectiveness

After the elicitation, we asked participants to rate each interaction modality according to how they perceived its effectiveness for the given scenario, as Morris [14] did. Fig. 6 shows these ratings that reflect the overall assessment of the participants about each of the four modalities for exploring data visually on a large vertical display. Participants considered touch the most effective way to interact with the data visualizations, followed closely by speech. The pen and mid-air gestures were rated neutral by 45% and 40% of the participants respectively. The fact that speech was rated second most effective is surprising, given that 65% of the participants initially reported that they had never used speech interaction before. Although people had more experience with the pen and mid-air gestures, they still perceived speech as more effective. Nevertheless, these ratings of perceived effectiveness roughly fit the appearances



Fig. 5. Collaborative synonyms proposed for merging and splitting views. (a) Two participants perform two gestures to merge the views together. (b) One participant performs the gesture alone to split the view back into two.

of each modality in the top proposals per referent. However, when interpreting these results, it is necessary to bear in mind that the assessment of the interaction modalities comes from the experience of the participants with the 15 referents we chose according to the scenario (see Sect. 4.4). Therefore, these findings might be expanded with the study of other referents and scenarios.

## 6 DISCUSSION

In this section, we discuss and interpret our findings.

### 6.1 Is touch and speech all we need?

Based on previous work, we expected participants to associate the exploration tasks with specific interaction modalities. In our study, they chose to focus on speech and touch interactions. Similar to the findings of Mignot et al. [69], our results suggest that people prefer using speech commands for tasks that have no or only a loose connection to specific screen coordinates (i.e. a location on the screen). The preference of speech to filter also matches the recommendations of Badam et al. [41]. However, they propose using touch interaction to select data items, and in contrast, our participants deemed the pen most suitable for selection and speech for deselection. Thus, we hypothesize that participants considered deselecting different than selecting because it did not require looking for the marks on the display.

Touch interaction was one of the two interaction modalities we offered for close interaction. Although large vertical displays provide a wide surface to interact with, participants sometimes wished to interact with the pen instead of touch. For example, they proposed to select with the pen instead of the finger due to the higher precision. The pen was also a clear favorite for adding annotations. Therefore, both touch and pen might have their place for large display interaction, especially when precision becomes critical. We also find the dominance of speech commands for 10 referents intriguing. In the demographics questionnaire, 65% of the participants indicated that they had never used speech interaction, although most web browsers and smartphones recognize speech commands nowadays. So, how come the scientists preferred speech in the study but had rarely used it before? A potential explanation is that the lack of technical limitations during the elicitation gave participants confidence to brainstorm speech commands. Another reason might be that the prevalence of physical navigation in front of large vertical displays gives priority to speech as a natural way to interact from a distance. Thus, speech interaction becomes more relevant as the display size increases. Compared to mid-air gestures, expressing more complex commands was easier — more so when they did not involve the definition of a spatial component in the data (e.g., adding a regression line). Also, some mid-air gesture proposals required standing closer to the screen (e.g., air-dragging a view). Among the top interactions, the most popular mid-air gestures allowed the user to stand further away. Therefore, mid-air gestures were actually proposed both for interacting from afar and at a close range, depending on the characteristics of the specific gesture, but they were still the least proposed among the four modalities. At the end of the study, one participant indicated that mid-air gestures were best combined with another modality to add precision. For example, one proposal involved pointing to a mark from a distance, such as with a laser pointer, and specifying the data attributes to show via voice. The use of voice commands to provide more

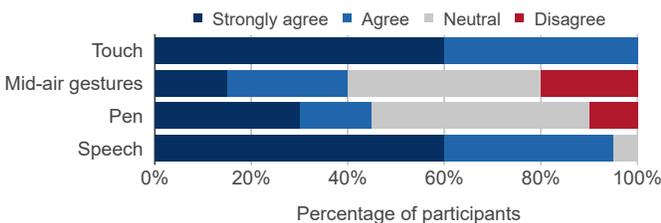


Fig. 6. Ratings of how participants perceived each modality as an effective way to interact with the visualizations on a large vertical display.

details suggests that participants felt they could be more specific with speech. Future work should look into more referents to better understand user preferences when having the possibility to interact through multiple modalities with data visualizations. Investigating diverse tasks is necessary as some may lead to clear tendencies (e.g., participants will most likely prefer using the pen for high-precision tasks).

## 6.2 Should we offer multimodal interaction?

The consensus set of our study is unimodal despite previous evidence of user preferences for multimodal interaction [8]. However, as Oviatt [70] points out, having the possibility to interact multimodally does not mean that users will take it. Although 42% of the elicited interactions were multimodal, the unimodal alternatives had the highest frequency. Participants appreciated the expressiveness of multimodal interaction but favored the simplicity of single modalities for low-level tasks. Moreover, multimodal interaction is still rare in industry products, and legacy bias may have influenced the preference for single modalities.

Nevertheless, multimodal interactions were among the top proposals for nine referents. Those multimodal interactions were mainly sequences of direct manipulation (touch or pen) followed by an action at a distance (speech or mid-air gesture). That suggests that given the freedom to stand at any distance from the screen, participants preferred to combine modalities that would work at different distances, especially when collaborating. For eight of the nine referents with top multimodal proposals, the preferred proposal that surpassed the multimodal proposal in popularity was a speech command. That suggests that participants opted to express the whole task through speech instead of dividing it into two steps. In a real-world scenario where speech recognition errors are common [8], multimodal interaction may be more reliable and precise than a speech command. For example, for view-driven filtering, defining a query orally to define a group of data items that the user noticed visually may be more challenging than tapping on the data items and then applying the filter with speech. Thus, supporting multimodal interactions would make the visualization system more robust. Future work should study what factors may influence the choice of the participants to combine specific modalities, such as physical movement and interaction costs.

## 6.3 Should cooperative input be supported?

Collaborative interactions were the smallest group among the top proposals. For referents like *sort*, two participants expressed that the task was too simple to interact collaboratively. For referents like selecting and splitting views, participants appreciated working collaboratively. That fits the finding of Morris et al. [50] about cooperative gestures not being performed too often to avoid interrupting their partner. Participants favored collaboration when there were two item groups or two views to interact with. When we examine the collaborative interactions that made it to the top proposals, we mostly find combinations of a modality suitable for direct manipulation and a modality for distant interaction. Those combinations correspond to the findings of Hinrichs and Carpendale [71] on interaction with tabletops. They found that the actions performed by multiple users

were strongly influenced by the social context. When our participants proposed collaborative interactions combining close and distant interaction, they were often already in position: one person was standing close to the display and the other further away. Proposing such multimodal sequences may therefore be a direct consequence of their placement. That suggests that participants may divide not only the screen space between them [47] but also the larger area in front of the display. As we did not ask participants to start the task at a specific distance to the screen, future work should analyze how users position themselves in the 3D space in front of the screen, with the help of a motion tracking system (e.g., [49]), to investigate how the initial position and movement may influence their choices. The visualization design choices should also be considered to investigate whether they influence the interaction distance. Identifying the reasons why people choose to interact collaboratively, taking interaction cost and engagement into account, is also an interesting research question for future work.

## 6.4 Are elicitation studies helpful for designing interactive data visualizations?

One of the main goals of an elicitation study is to define a consensus set. In ours, there was no conflict among interactions, i.e., participants did not map the same interaction to two different referents. Thus, we could implement a system that would have no problem distinguishing between tasks. That lack of conflict was potentially due to the dominance of speech and natural language being more expressive than other modalities. Implementing the consensus set would require speech recognition combined with touch and pen input. Given that the touch and pen interactions were standard actions (e.g., tap, drag and drop, draw a line), the main technical challenge would be having reliable speech recognition. If the implemented version struggled with speech recognition errors as in previous work [8], the extended list of top proposals provides alternatives to speech commands. In this respect, the elicitation was a successful methodology for us. In future work, conducting a study with a system enabling the consensus set is necessary to assess how effectively the elicited interactions can support data exploration on a vertical display when put together. The system should offer multiple interaction options per referent (including *multimodal synonyms*) so that participants can choose among different modalities or modality combinations to perform a task. Including *collaborative synonyms* would also give participants the opportunity to choose between personal and collaborative work. Testing the system in different contexts (e.g., meeting room, public space) would help to assess how the circumstances may influence the participant choices.

Participants consistently associated complementary referents with the same interaction modality. For example, activating and deactivating the *brushing & linking* technique were both preferred via speech. *Select* and *deselect* were the only exception. While the pen prevailed for selection, speech interaction was preferred to *deselect*, with a high max-consensus. The study results gave us insights into how users perceived the tasks and how the interaction techniques could be designed based on the groups by modality and directness we found.

In the study design, we applied all recommended techniques to avoid legacy bias. The dominance of speech commands despite the lack of experience of the participants with it suggests that we mitigated the bias successfully. However, three standard touch gestures made it to the consensus set which might not be problematic as the support of standard operations will be expected by future users of an interactive large display system. Participants often started speech commands with phrases like “Hey Siri”, suggesting an influence of their knowledge about voice assistants but also pointing out that they wished for speech input to be given explicitly rather than having their speech analyzed throughout their work in front of the display. Although Morris et al. [21] introduced the three principles for reducing legacy bias almost a decade ago, the standard analysis and agreement calculation recommendations still focus on single elicitation [11], i.e., when one person participates alone and makes only one proposal per referent. The study of Morris [14] is the only known example of group elicitation with multiple modalities. Elicitation research has not yet considered how to incorporate the three principles in the data analysis, and therefore, the options for quantitatively analyzing agreement are still limited.

## 7 CONCLUSION

In this work, we explored how different interaction modalities can be used to explore data visually on large vertical displays. Our results suggest that unimodal and personal interactions are preferred, but a system should enable touch, pen, and speech interaction to support data exploration with direct manipulation and natural language according to user preferences. Participants favored touch and speech, alone or in combination, to perform low-level exploration tasks with diverse visualizations of spatio-temporal data. However, when taking the top proposals into account, the choices and combinations of modalities are diverse. An evaluation with a real-world system would help assess whether and how the interaction choices of users may match the results of the elicitation study. The interface design also needs to be considered, as it will influence the interaction cost and the positioning of the users. In our study, we used an interface as simple as possible and encouraged participants to suggest interface elements to add if they wished. When working collaboratively, participants either used a single modality in parallel or used two modalities in a sequence, one for direct manipulation and another for distant interaction. We provide the consensus set and our analysis of the interaction proposals elicited in the study, to inform the interaction design of visual systems for collaborative data exploration. Future work should consider other interaction modalities relevant to large vertical displays, such as proxemics, and participant groups from other application scenarios with different data and task types.

## ACKNOWLEDGMENTS

The authors would like to thank the study participants, Helen Seitzer, and Keonhi Son for their support. This work was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project number 374666841—SFB 1342.

## REFERENCES

- [1] M. R. Jakobsen, Y. Sahlemariam Haile, S. Knudsen, and K. Hornbæk, “Information visualization and proxemics: Design opportunities and empirical findings,” *IEEE Trans. Visual Comput. Graphics*, vol. 19, no. 12, pp. 2386–2395, 2013. [Online]. Available: <https://doi.org/10.1109/TVCG.2013.166>
- [2] P. Isenberg, A. Bezerianos, N. Henry, S. Carpendale, and J.-D. Fekete, “CoCoNutTrix: Collaborative Retrofitting for Information Visualization,” *IEEE Comput. Graphics Appl.: Special Issue on Collaborative Visualization*, vol. 29, no. 5, pp. 44–57, Sep./Oct. 2009. [Online]. Available: <https://doi.org/10.1109/MCG.2009.78>
- [3] S. K. Badam, F. Amini, N. Elmqvist, and P. Irani, “Supporting visual exploration for multiple users in large display environments,” in *IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2016, pp. 1–10. [Online]. Available: <https://doi.org/10.1109/VAST.2016.7883506>
- [4] K. Hinckley, K. Yatani, M. Pahud, N. Coddington, J. Rodenhouse, A. Wilson, H. Benko, and B. Buxton, “Pen + touch = new tools,” in *Proc. Annu. ACM Symp. on User Interface Software and Technol.* New York: ACM, 2010, pp. 27–36. [Online]. Available: <https://doi.org/10.1145/1866029.1866036>
- [5] A. S. Williams, J. Garcia, and F. Ortega, “Understanding multimodal user gesture and speech behavior for object manipulation in augmented reality using elicitation,” *IEEE Trans. Visual Comput. Graphics*, vol. 26, no. 12, pp. 3479–3489, 2020. [Online]. Available: <https://doi.org/10.1109/TVCG.2020.3023566>
- [6] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani, “Consumed endurance: A metric to quantify arm fatigue of mid-air interactions,” in *Proc. CHI*. New York: ACM, 2014, pp. 1063–1072. [Online]. Available: <https://doi.org/10.1145/2556288.2557130>
- [7] S. Oviatt, A. DeAngeli, and K. Kuhn, “Integration and synchronization of input modes during multimodal human-computer interaction,” in *Referring Phenomena in a Multimedia Context and Their Computational Treatment*. USA: Association for Computational Linguistics, 1997, pp. 1–13.
- [8] A. Saktheeswaran, A. Srinivasan, and J. Stasko, “Touch? speech? or touch and speech? Investigating multimodal interaction for visual network exploration and analysis,” *IEEE Trans. Visual Comput. Graphics*, vol. 26, no. 6, pp. 2168–2179, 2020. [Online]. Available: <https://doi.org/10.1109/TVCG.2020.2970512>
- [9] A. Srinivasan, B. Lee, and J. Stasko, “Interweaving multimodal interaction with flexible unit visualizations for data exploration,” *IEEE Trans. Visual Comput. Graphics*, vol. 27, no. 8, pp. 3519–3533, 2021. [Online]. Available: <https://doi.org/10.1109/TVCG.2020.2978050>
- [10] J. O. Wobbrock, M. R. Morris, and A. D. Wilson, “User-defined gestures for surface computing,” in *Proc. CHI*. New York: ACM, 2009, pp. 1083–1092. [Online]. Available: <https://doi.org/10.1145/1518701.1518866>
- [11] R.-D. Vatavu and J. O. Wobbrock, “Clarifying agreement calculations and analysis for end-user elicitation studies,” *ACM Trans. Comput.-Hum. Interact.*, vol. 29, no. 1, Jan. 2022. [Online]. Available: <https://doi.org/10.1145/3476101>
- [12] T. Tsandilas and P. Dragicevic, “Accounting for Chance Agreement in Gesture Elicitation Studies,” LRI - CNRS, University Paris-Sud, Research Report 1584, Feb. 2016. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-01267288>
- [13] S. Villarreal-Narvaez, J. Vanderdonckt, R.-D. Vatavu, and J. O. Wobbrock, “A systematic review of gesture elicitation studies: What can we learn from 216 studies?” in *Proc. DIS*. New York: ACM, 2020, pp. 855–872. [Online]. Available: <https://doi.org/10.1145/3357236.3395511>
- [14] M. R. Morris, “Web on the wall: Insights from a multimodal interaction elicitation study,” in *Proc. ITS*. New York: ACM, 2012, pp. 95–104. [Online]. Available: <https://doi.org/10.1145/2396636.2396651>
- [15] W. Willett, Q. Lan, and P. Isenberg, “Eliciting Multi-touch Selection Gestures for Interactive Data Graphics,” in *EuroVis - Short Papers*, N. Elmqvist, M. Hlawitschka, and J. Kennedy, Eds. The Eurographics Association, 2014. [Online]. Available: <https://doi.org/10.2312/eurovisshort.20141161>
- [16] B. Lee, A. Srinivasan, P. Isenberg, and J. Stasko, “Post-wimp interaction for information visualization,” *Foundations and Trends® in Human-Computer Interaction*, vol. 14, no. 1, pp. 1–95, 2021. [Online]. Available: <http://dx.doi.org/10.1561/11000000081>

- [17] C. Abras, D. Maloney-Krichmar, J. Preece, and W. Bainbridge, "Encyclopedia of human-computer interaction," *Thousand Oaks: Sage Publications*, vol. 37, pp. 445–456, 2004.
- [18] J. O. Wobbrock, H. H. Aung, B. Rothrock, and B. A. Myers, "Maximizing the guessability of symbolic input," in *CHI Extended Abstracts*. New York: ACM, 2005, pp. 1869–1872. [Online]. Available: <https://doi.org/10.1145/1056808.1057043>
- [19] M. A. Nacenta, Y. Kamber, Y. Qiang, and P. O. Kristensson, "Memorability of pre-designed and user-defined gesture sets," in *Proc. CHI*. New York: ACM, 2013, pp. 1099–1108. [Online]. Available: <https://doi.org/10.1145/2470654.2466142>
- [20] M. Nebeling, A. Huber, D. Ott, and M. C. Norrie, "Web on the wall reloaded: Implementation, replication and refinement of user-defined interaction sets," in *Proc. ITS*. New York: ACM, 2014, pp. 15–24. [Online]. Available: <https://doi.org/10.1145/2669485.2669497>
- [21] M. R. Morris, A. Danielescu, S. Drucker, D. Fisher, B. Lee, m. c. schraefel, and J. O. Wobbrock, "Reducing legacy bias in gesture elicitation studies," *Interactions*, vol. 21, no. 3, pp. 40–45, May 2014. [Online]. Available: <https://doi.org/10.1145/2591689>
- [22] A. S. Williams and F. R. Ortega, "A concise guide to elicitation methodology," 2021. [Online]. Available: <https://arxiv.org/abs/2105.12865>
- [23] R.-D. Vatavu and J. O. Wobbrock, "Formalizing agreement analysis for elicitation studies: New measures, significance test, and toolkit," in *Proc. CHI*. New York: ACM, 2015, pp. 1325–1334. [Online]. Available: <https://doi.org/10.1145/2702123.2702223>
- [24] T. Tsandilas, "Fallacies of agreement: A critical review of consensus assessment methods for gesture elicitation," *ACM Trans. Comput.-Hum. Interact.*, vol. 25, no. 3, Jun. 2018. [Online]. Available: <https://doi.org/10.1145/3182168>
- [25] C. Andrews, A. Endert, and C. North, "Space to think: Large high-resolution displays for sensemaking," in *Proc. CHI*. New York: ACM, 2010, pp. 55–64. [Online]. Available: <https://doi.org/10.1145/1753326.1753336>
- [26] M. R. Jakobsen and K. Hornbæk, "Up close and personal: Collaborative work on a high-resolution multitouch wall display," *ACM Trans. Comput.-Hum. Interact.*, vol. 21, no. 2, Feb. 2014. [Online]. Available: <https://doi.org/10.1145/2576099>
- [27] A. Bezerianos and P. Isenberg, "Perception of Visual Variables on Tiled Wall-Sized Displays for Information Visualization Applications," *IEEE Trans. Visual Comput. Graphics*, vol. 18, no. 12, pp. 2516–2525, Dec. 2012. [Online]. Available: <https://doi.org/10.1109/TVCG.2012.251>
- [28] P. Riehm, G. Molina León, J. Reibert, F. Ehtler, and B. Froehlich, "Short-contact touch-manipulation of scatterplot matrices on wall displays," *Comput. Graphics Forum*, vol. 39, no. 3, pp. 265–276, 2020. [Online]. Available: <https://doi.org/10.1111/cgf.13979>
- [29] F. Guimbretière, M. Stone, and T. Winograd, "Fluid interaction with high-resolution wall-size displays," in *Proceedings of the 14th Annu. ACM Symp. on User Interface Software and Technol.* New York: ACM, 2001, pp. 21–30. [Online]. Available: <https://doi.org/10.1145/502348.502353>
- [30] S. Herholz, L. L. Chuang, T. G. Tanner, H. H. Bühlhoff, and R. W. Fleming, "Libgaze: Real-time gaze-tracking of freely moving observers for wall-sized displays," in *13th International Fall Workshop on Vision, Modeling, and Visualization*. Akademische Verlagsgesellschaft AKA, 2008, pp. 101–110.
- [31] T. Horak, S. K. Badam, N. Elmquist, and R. Dachselt, "When david meets goliath: Combining smartwatches with a large vertical display for visual data exploration," in *Proc. CHI*. New York: ACM, 2018, pp. 1–13. [Online]. Available: <https://doi.org/10.1145/3173574.3173593>
- [32] U. Kister, K. Klamka, C. Tominski, and R. Dachselt, "Grasp: Combining spatially-aware mobile devices and a display wall for graph visualization and interaction," *Comput. Graphics Forum*, vol. 36, no. 3, pp. 503–514, 2017. [Online]. Available: <https://doi.org/10.1111/cgf.13206>
- [33] P. Reipschlager, T. Flemisch, and R. Dachselt, "Personal augmented reality for information visualization on large interactive displays," *IEEE Trans. Visual Comput. Graphics*, vol. 27, no. 2, pp. 1182–1192, 2021. [Online]. Available: <https://doi.org/10.1109/TVCG.2020.3030460>
- [34] M. Nancel, J. Wagner, E. Pietriga, O. Chapuis, and W. Mackay, "Mid-air pan-and-zoom on wall-sized displays," in *Proc. CHI*. New York: ACM, 2011, pp. 177–186. [Online]. Available: <https://doi.org/10.1145/1978942.1978969>
- [35] P. Baudisch, N. Good, and P. Stewart, "Focus plus context screens: Combining display technology with visualization techniques," in *Proc. Annu. ACM Symp. on User Interface Software and Technol.* New York: ACM, 2001, pp. 31–40. [Online]. Available: <https://doi.org/10.1145/502348.502354>
- [36] R. Ball, C. North, and D. A. Bowman, "Move to improve: Promoting physical navigation to increase user performance with large displays," in *Proc. CHI*. New York: ACM, 2007, pp. 191–200. [Online]. Available: <https://doi.org/10.1145/1240624.1240656>
- [37] A. Srinivasan, B. Lee, N. Henry Riche, S. M. Drucker, and K. Hinckley, *InChorus: Designing Consistent Multimodal Interactions for Data Visualization on Tablet Devices*. New York: ACM, 2020, pp. 1–13. [Online]. Available: <https://doi.org/10.1145/3313831.3376782>
- [38] B. Lee, G. Smith, N. H. Riche, A. Karlson, and S. Carpendale, "Sketchinsight: Natural data exploration on interactive whiteboards leveraging pen and touch interaction," in *Proc. PacificVis*, 2015, pp. 199–206. [Online]. Available: <https://doi.org/10.1109/PACIFICVIS.2015.7156378>
- [39] S. M. Drucker, D. Fisher, R. Sadana, J. Herron, and m. schraefel, "Touchviz: A case study comparing two interfaces for data analytics on tablets," in *Proc. CHI*. New York: ACM, 2013, pp. 2301–2310. [Online]. Available: <https://doi.org/10.1145/2470654.2481318>
- [40] J. Walny, B. Lee, P. Johns, N. Henry Riche, and S. Carpendale, "Understanding pen and touch interaction for data exploration on interactive whiteboards," *IEEE Trans. Visual Comput. Graphics*, vol. 18, no. 12, pp. 2779–2788, 2012. [Online]. Available: <https://doi.org/10.1109/TVCG.2012.275>
- [41] S. K. Badam, A. Srinivasan, N. Elmquist, and J. Stasko, "Affordances of input modalities for visual data exploration in immersive environments," in *2nd Workshop on Immersive Analytics*, 2017.
- [42] G. Molina León, M. Lischka, W. Luo, and A. Breiter, "Mobile and multimodal? A comparative evaluation of interactive workplaces for visual data exploration," *Comput. Graphics Forum*, vol. 41, no. 3, pp. 417–428, 2022. [Online]. Available: <https://doi.org/10.1111/cgf.14551>
- [43] P. Isenberg, N. Elmquist, J. Scholtz, D. Cernea, K.-L. Ma, and H. Hagen, "Collaborative Visualization: Definition, Challenges, and Research Agenda," *Inf. Visualization Journal (IVS), Special Issue on Information Visualization: State of the Field and New Research Directions*, vol. 10, no. 4, pp. 310–326, Oct. 2011. [Online]. Available: <https://doi.org/10.1177/1473871611412817>
- [44] P. Isenberg, D. Fisher, M. Ringel Morris, K. Inkpen, and M. Czerwinski, "An Exploratory Study of Co-located Collaborative Visual Analytics around a Tabletop Display," in *Proceedings of Visual Analytics Science and Technology (VAST)*. Los Alamitos: IEEE, 2010, pp. 179–186. [Online]. Available: <https://doi.org/10.1109/VAST.2010.5652880>
- [45] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale, "Empirical studies in information visualization: Seven scenarios," *IEEE Trans. Visual Comput. Graphics*, vol. 18, no. 9, pp. 1520–1536, 2012. [Online]. Available: <https://doi.org/10.1109/TVCG.2011.279>
- [46] Y. Rogers and S. Lindley, "Collaborating around vertical and horizontal large interactive displays: which way is best?" *Interact. Comput.*, vol. 16, no. 6, pp. 1133–1152, Sep. 2004. [Online]. Available: <https://doi.org/10.1016/j.intcom.2004.07.008>
- [47] A. Prouzeau, A. Bezerianos, and O. Chapuis, "Evaluating multi-user selection for exploring graph topology on wall-displays," *IEEE Trans. Visual Comput. Graphics*, vol. 23, no. 8, pp. 1936–1951, 2017. [Online]. Available: <https://doi.org/10.1109/TVCG.2016.2592906>
- [48] F. Brudy, E. Ledo, S. Greenberg, and A. Butz, "Is anyone looking? Mitigating shoulder surfing on public displays through awareness and protection," in *Proceedings of The International Symposium on Pervasive Displays*. New York: ACM, 2014, pp. 1–6. [Online]. Available: <https://doi.org/10.1145/2611009.2611028>
- [49] J. Dostal, U. Hinrichs, P. O. Kristensson, and A. Quigley, "Spidereyes: Designing attention- and proximity-aware collaborative interfaces for wall-sized displays," in *Proc. IUI*. New York: ACM, 2014, pp. 143–152. [Online]. Available: <https://doi.org/10.1145/2557500.2557541>
- [50] M. R. Morris, A. Huang, A. Paepcke, and T. Winograd, "Cooperative gestures: Multi-user gestural interactions for co-located groupware," in *Proc. CHI*. New York: ACM, 2006, pp. 1201–1210. [Online]. Available: <https://doi.org/10.1145/1124772.1124952>
- [51] C. Liu, O. Chapuis, M. Beaudouin-Lafon, and E. Lecolinet, "Coreach: Cooperative gestures for data manipulation on wall-sized displays," in *Proc. CHI*. New York: ACM, 2017, pp. 6730–6741. [Online]. Available: <https://doi.org/10.1145/3025453.3025594>

- [52] P. Isenberg and D. Fisher, "Collaborative Brushing and Linking for Co-located Visual Analytics of Document Collections," *Comput. Graphics Forum*, vol. 28, no. 3, pp. 1031–1038, Jun. 2009. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8659.2009.01444.x>
- [53] Plotly, "Plotly express in python," 2022. [Online]. Available: <https://plotly.com/python/plotly-express/>
- [54] G. Foundation. (2022, Jul.) Free data via gapminder.org, cc-by license. [Online]. Available: <https://www.gapminder.org/data/>
- [55] N. Andrienko and G. Andrienko, *Exploratory analysis of spatial and temporal data: a systematic approach*. Springer Science & Business Media, 2006.
- [56] J. S. Yi, Y. a. Kang, J. Stasko, and J. Jacko, "Toward a deeper understanding of the role of interaction in information visualization," *IEEE Trans. Visual Comput. Graphics*, vol. 13, no. 6, pp. 1224–1231, 2007. [Online]. Available: <https://doi.org/10.1109/TVCG.2007.70515>
- [57] M. Brehmer and T. Munzner, "A multi-level typology of abstract visualization tasks," *IEEE Trans. Visual Comput. Graphics*, vol. 19, no. 12, pp. 2376–2385, 2013. [Online]. Available: <https://doi.org/10.1109/TVCG.2013.124>
- [58] R. Langner, U. Kister, and R. Dachselt, "Multiple coordinated views at large displays for multiple users: Empirical findings on user behavior, movements, and distances," *IEEE Trans. Visual Comput. Graphics*, vol. 25, no. 1, pp. 608–618, 2019. [Online]. Available: <https://doi.org/10.1109/TVCG.2018.2865235>
- [59] R. Sadana and J. Stasko, "Designing and implementing an interactive scatterplot visualization for a tablet computer," in *Proc. AVI*. New York: ACM, 2014, pp. 265–272. [Online]. Available: <https://doi.org/10.1145/2598153.2598163>
- [60] —, "Designing multiple coordinated visualizations for tablets," *Comput. Graphics Forum*, vol. 35, no. 3, pp. 261–270, 2016. [Online]. Available: <https://doi.org/10.1111/cgf.12902>
- [61] N. Andrienko, G. Andrienko, and P. Gatalsky, "Exploratory spatio-temporal visualization: an analytical review," *Journal of Visual Languages & Computing*, vol. 14, no. 6, pp. 503–541, 2003, visual Data Mining. [Online]. Available: [https://doi.org/10.1016/S1045-926X\(03\)00046-6](https://doi.org/10.1016/S1045-926X(03)00046-6)
- [62] M. Perera, T. Gedeon, M. Adcock, and A. Haller, "Towards self-guided remote user studies - feasibility of gesture elicitation using immersive virtual reality," in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2021, pp. 2576–2583. [Online]. Available: <https://doi.org/10.1109/SMC52423.2021.9658673>
- [63] K. Sassenberg and G. B. Moskowitz, "Don't stereotype, think different! overcoming automatic stereotype activation by mindset priming," *J. Exp. Social Psychol.*, vol. 41, no. 5, pp. 506–514, 2005. [Online]. Available: <https://doi.org/10.1016/j.jesp.2004.10.002>
- [64] A. Ali, M. Ringel Morris, and J. O. Wobbrock, "I am iron man': Priming improves the learnability and memorability of user-elicited gestures," in *Proc. CHI*. New York: ACM, 2021. [Online]. Available: <https://doi.org/10.1145/3411764.3445758>
- [65] A. S. Williams, J. Garcia, F. De Zayas, F. Hernandez, J. Sharp, and F. R. Ortega, "The cost of production in elicitation studies and the legacy bias-consensus trade off," *Multimodal Technologies and Interaction*, vol. 4, no. 4, 2020. [Online]. Available: <https://www.mdpi.com/2414-4088/4/4/88>
- [66] R.-D. Vatavu, *The Dissimilarity-Consensus Approach to Agreement Analysis in Gesture Elicitation Studies*. New York: ACM, 2019, pp. 1–13. [Online]. Available: <https://doi.org/10.1145/3290605.3300454>
- [67] W. Buxton, "Chunking and phrasing and the design of human-computer dialogues," in *Readings in Human-Computer Interaction*. Elsevier, 1995, pp. 494–499.
- [68] T. Y. Channel. How to revert keep only or exclude on tableau desktop. Tableau. [Online]. Available: [https://www.youtube.com/watch?v=bK3MGEV0OIU&ab\\_channel=Tableau](https://www.youtube.com/watch?v=bK3MGEV0OIU&ab_channel=Tableau)
- [69] C. Mignot, C. Valot, and N. Carbonell, "An experimental study of future "natural" multimodal human-computer interaction," in *INTERACT and CHI Conference Companion*, 1993, pp. 67–68.
- [70] S. Oviatt, "Ten myths of multimodal interaction," *Commun. ACM*, vol. 42, no. 11, pp. 74–81, Nov. 1999. [Online]. Available: <https://doi.org/10.1145/319382.319398>
- [71] U. Hinrichs and S. Carpendale, "Gestures in the wild: Studying multi-touch gesture sequences on interactive tabletop exhibits," in *Proc. CHI*. New York: ACM, 2011, pp. 3023–3032. [Online]. Available: <https://doi.org/10.1145/1978942.1979391>



**Gabriela Molina León** is a doctoral researcher at the University of Bremen. Her main research interests include information visualization and human-computer interaction, with a focus on interaction design for data visualizations.



**Petra Isenberg** is a research director with Inria, France in the Aviz team. Her main research interests include visualization and visual analytics. She is interested in exploring how people can most effectively work when analyzing large and complex data sets—often on novel display technology.



**Andreas Breiter** is Full Professor for Informatics, Department for Mathematics and Computer Science, at the University of Bremen and Scientific Director of the Institute for Information Management Bremen (ifib). His research focuses on research data management, computational social science, and the digital transformation of education.