IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 April 2024, accepted 10 May 2024, date of publication 23 May 2024, date of current version 10 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3404876

RESEARCH ARTICLE

User-Defined Interactions for Visual Data Exploration With the Combination of Smartwatch and Large Display

YIQI XIAO AND LU LIU
College of Communication and Art Design, University of Shanghai for Science and Technology, Shanghai 200093, China
Corresponding author: Lu Liu (liu228281@gmail.com)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Academic Ethics and Discipline Supervision Committee of University of Shanghai for Science and Technology.

ABSTRACT The process of visual analytics is composed of the visual data exploration tasks supporting analytical reasoning. When performing analytical tasks with the interactive visual interfaces displayed by the large screen, physical discomforts such as gorilla-arm effect can be easily caused. To enrich the input space for analysts, there has been some researches concerning the cross-device analysis combining mobile devices with the large display. Although the effectiveness of expert-level designs has been demonstrated, little is known of the ordinary users' preferences for using a mobile device to issue commands, especially the small one like smartwatch. We implement a three-stage study to investigate and validate these preferences. A total of 181 distinctive gestural inputs and 52 interface designs for 21 tasks were collected from analysts. Expert designers selected the best practices from these user-defined interactions. A performance test was subsequently developed to assess the selected interactions in terms of quantitative statistics and subjective ratings. Our work provides empirical support and proposes a set of design guidelines for optimizing watch-based interactions aimed at remote control of visual data on the large display. Through this research, we hope to advance the development of smartwatches as visual analytics tools and provide visual analysts with a better usage experience.

INDEX TERMS User-defined interaction, human-computer interaction, visual analytics, smartwatch, large display.

I. INTRODUCTION

With the explosive growth of data in the current era, it is crucial to effectively extract useful information quickly from the massive and complex data through visual analytics. For this purpose, the design of visual analytics has received much attention, especially in terms of the representation and interaction of visual elements. Representation refers to the mapping of data to graphics, which is used to represent data in a more intuitive and understandable way. Researches on the interaction for visual analytics center on methods and techniques that facilitate the exploration and interpretation of data [1]. Better interactions not only facilitate uncover insights [2] from data to support user decision making, but also improve the efficiency of analysis while reducing the cognitive load and physical effort for manipulation.

Visual analytic interaction has traditionally employed the Windows, Icons, Mouse and Pointer (WIMP) [3] paradigm of user interfaces to perform analytic tasks and manage data content stacked on a large display. As vision-based gesture interfaces (VBGI) have become more popular, researchers begin to explore how WIMP can be replaced with interaction primitives, e.g. touch [4] or freehand [5] gestures that are conducive to the expression of the analyst's intention to enable direct manipulations from their embodied experience. However, the use of interactions based on visual gesture recognition yields several challenges, including "vocabulary disagreement" [6], "midas touch" [7], and "gorilla-arm

The associate editor coordinating the review of this manuscript and approving it for publication was Walter Didimo.

effect" [8]. These challenges have a significant negative effect on the experience of mid-air gesture interactions orienting to the large display. For data analysts, the larger size of the standing screen and more information volume can require increased upper limb movement, as they need to frequently acquire, edit, and compare data only with gestural interactions. In this situation, due to the not very ergonomic positioning that is required, users are more likely to suffer from fatigue and their arm may start to hurt over time [9], [10]. In addition, some subconscious, casual actions during freehand interaction may be misrecognized by the system as task commands. Furthermore, visual analytics involves a large number of single tasks [11]. Mapping a distinctive gesture-based interactive input to each task would undoubtedly burden user's memory of the gestural repertoire. With these concerns, this paper concentrates on an attempt to develop a set of interactions for visual analytics with the assistance of wearable devices, within the framework of a brief account of the suitable tasks.

The idea of making sense of visual data on a large display through commanding a smartwatch exploits the dimensions of wearable devices, thereby enabling micro-interactions [12] and proxemic interactions [13]. To our knowledge, the study of using the smartwatch as an input field of interactions for visual analytics is as yet in its infancy, although some research has been made, see the work of Horak et al. [14]. Unlike interactions with smartwatch only and cross-device interactions involving the watch, a smartwatch used for visual analytics interactions should be the miniature version of large screen displaying the visualizations. This requires not only an interaction style that is more appropriate for mini-screen, but also the corresponding visual feedback for a specific analytic task to be visible on the watch. For this reason, the researcher should have a good understanding of how the interaction with smartwatch plays its role in the visual analytics process. As there was a dearth of relevant design practices, we intended to investigate the interaction preferences of analysts as end-users with the user-eliciting approach. The findings may provide insights for researchers to design interactions for visual data exploration involving wearable devices.

This paper is structured as follow. We first define visual analytic tasks to be performed combining with the watch and the large display. The selected tasks are representatives of task categories of visual analytics. We perform the Wizard-of-Oz test to elicit user proposals as it is a user-research method of low cost where a user interacts with a mock interface controlled by a person. Using a smartwatch prototype and a smart blackboard displaying pre-designed visualizations, the mock interface can simulate a watch-based interaction input and its visual output. After the user-eliciting experiment, a set of user-preferred smartwatch interactions for visual data exploration of the large display is obtained. Each interaction includes the user-preferred gesture command to the smartwatch for analyzing the on-large-display visualization and the preferred user interfaces (UIs) of smartwatch in response to the command. The above work constitutes the user elicitation study of the research. We invited expert designers to select the user-defined interactions using the choice-based elicitation approach, and then statistically compared the preferences between two user groups. Finally, we tested memorability of the selected gestures and user experience of the interactions with smartwatch, thus to assess the feasibility of users' designs. The test results can indicate the complexity of the tasks for visual data exploration and contribute to design suggestions.

II. RELATED WORK

A. TASKS AND INTERACTIONS FOR VISUAL ANALYTICS

What are the interaction behaviors that make sense for interpreting data in a visual analytics system? This question has been of interest to researchers for a long time. However, different data content and visualization techniques call for different interaction tasks. The more specific interaction intentions for close-up observation of data call for categories of task with higher level of granularity. To date, there has been a handful of works focusing on the taxonomy of commonly used interaction tasks for visual analytics. An earlier attempt to provide a comprehensive overview of interactions in information visualization was done by Yi et al. [2]. This work proposed seven basic classifications including select, explore, reconfigure, encode, abstract/elaborate, filter and connect, which provided a new perspective for later studies to understand visualization interactions.

In fact, a compound or abstract visualization interaction task is probably composed of a series of low-level tasks performed on graph-specific objects [15]. A multi-level typology of abstract visualization tasks was proposed by Brehmer and Munzner [11] to classify all tasks in terms of two dimensions of interaction: ends and means. In this framework, *present, discover, enjoy, produce objects, search and query* were considered as the intents that explain why a task is performed (i.e. the ends). The framework also inventoried two main categories of user intentions regarding task execution (i.e. the means): *manipulate and introduce.* Later studies have been more specific to a particular category of visualization, giving a detailed discussion on the specificity of the interaction task engendered by the structure and key features of the visualization. For example, Ahn et al. [16] proposed a task taxonomy of temporal features based on the identification of the entities, the properties to be visualized, and the hierarchy of temporal network evolution analysis. Kerracher et al. [17] argued that the user behaviors of analyzing temporal graph visualization—lookup, comparison, and relation seeking—could be either attribute-based tasks or structural tasks. They defined corresponding interactions according to the needs of analyzing the element and structure of temporal information. For group-level graph visualization, Saket et al.'s classification [18] included group-only, group-node, group-link and group-network tasks. Adopting the multi-level typology of abstract visualization tasks, the examples of group-level tasks were thus described. Similarly, Gladisch et al. [19] introduced several tasks specially for graph editing, including *create, insert, delete, select, update and navigate.*

To accomplish such miscellaneous tasks, an integrated system of visual analysis requires a rich interaction vocabulary [19] that takes full advantage of human cognitive and expressive habits. This rich interaction vocabulary is often supported by multimodal interaction technologies, combining touch- or gesture-based interface, tangible user interface and voice control. A typical multimodal interaction case is Orko, a network visualization of European soccer players, which was developed to facilitate both natural language and direct manipulation input for highlighting the connections between players [20]. Compared to direct interaction on the screen based on multi-touch [21], [22], [23], the use of physical objects of generic shapes acting as tokens was considered a more effective strategy for interactive data visualization on tabletop displays [24]. As for mid-air gestural interaction, with the advantage of free movement, it can express a variety of concrete and abstract interaction tasks. Therefore, gestures have been used in VR [25], medical volume data visualization [26] and scientific visualization. Ntoa et al. [27] designed a big data visualization application supporting directional mid-air gesture interaction with a large display. It allowed users to navigate in a data center room by adjusting the camera and the view. Other researchers have developed interactive prototypes that involved handheld or wearable devices for visual analytics of multidimensional data on the large display, and we will provide a detailed introduction to them in the following part.

B. THE SMALL DEVICES FOR CROSS-DEVICE INTERACTION AND DATA EXPLORATION

The growing interest to designing cross-device interaction accompanies with the popularization of personal computing devices on the consumer side. Cross-device interactions can be classified into two categories from the temporal dimension: synchronous and asynchronous [28]. *Synchronous* defines the situation that simultaneously employs two or more devices to accomplish the same task. in this situation, the behavior on the secondary device affects the primary device due to a distributed or mirrored user interface. *Asynchronous* mainly refers to the design of adaptive user interface which allows the digital content to be transferred from one device to another. The second devices, especially those that are small and portable, have the additional role of acting as remote control for larger displays [29].

There are different methods for performing interactions with small devices for cross-device communication. On-screen interface is one of them, and it has been applied in prototypes such as Conductor [30], WatchConnect [31] and CurationSpace [28]. On-screen interface, in the narrow sense, is multi-touch on the display of the device. In the broad sense, however, the around-device gestures such as waving above [32], *dragging, and touching a visual proxy* are also applicable to the display of small devices, and are favored for transferring and managing the visual data [28]. As wearable or graspable artifacts, the use of small devices can easily be linked to people's behavior of using tools. Such metaphorical associations were seen in some tangible interactions, for instance, tiliting the watch-like device to change [33] or synchronizing content [34] to the target device; and using pick-up gesture [35] or knocking the target display with the phone [36] to exchange a file; or specifying device motions in 3D space [37], [38] as interaction inputs. Additionally, the mechanical movement of the device is also a form of tangible interaction. For example, Xiao et al. [39] developed a watch prototype that allows for pan, twist and tilt of the dial. This mechanical interface with multi-degree of freedom can contribute to the sophistication and accuracy of interactions we perform.

According to Brudy et al. [28], data exploration has been an increasingly popular application domain of cross-device interaction. Among the small devices, mobile phone and tablet have become the main areas of concern to assist the presentation of data on the big screen or enable co-located data analysis [40]. Relevant projects include Thaddeus [41], VisTiles [42], Surface Constellations [40], Slice WIM [43], GraSp [44] and Photo4Action [25]. Some of the above projects were equipped with control panels that acted as toolbars, for example, VisTiles placed such a panel called sidebar on the edge of the minipad screen. Photo4Action represented other projects using widely-accepted touch gestures. Users can take photo to select the subgraph on the wall display and then pan, zoom or rotate the shown box representing the selected area. They can further check more details of nodes by tapping, drawing a lasso or long-press. Similar smartphone-based interactions are also seen in Sollich et al.'s [45] prototype for exploring volumetric microscopy movies, where two scroll bars are used to change the viewpoint and create key-frame bookmarks. GraSp was a project applying tangible interactions to visual analytics scenarios. It supported spatially horizontal, vertical and axial motion of the device to select the range of displayed data and provide a comparative view of data.

The previous study closest to our research was carried out by Horak et al. [14], as mentioned above, which pioneered a new smartwatch-based interaction system for data exploration on the large display. The system is built on two basic design concepts: *item sets and connective areas*. A set shows the configuration property or the simplified version of a data item containing multiple entities; and *connective areas* are four zones representing ''components of a visualization that have a specific interaction''. This system supported three types of inputs: touch-based swiping, rotating a physical control and spatially moving the arm. Analysts can issue commands to a focused connective area through these actions, and the change of data items will be shown on the set interface. Given that the screen size of the watch precludes many on-screen direct manipulations, some studies presented the fisheye distortion technique to facilitate data selection [46].

III. STUDY DESIGN

A. ORGANIZATION OF STUDIES

The paper describes a three-stage study (Figure 1). First, the experienced analysts are recruited to initiate the user elicitation study. The results can demonstrate user's prospects of relying on the watch interface to issue commands concerning data processing and presentation, even though they may also have usability and technical feasibility problems. In the second stage, interaction designers will select the elicited proposals from the perspective of design experts. Dim et al. [47] first introduced such a choice-based elicitation method as a further step to more reliable user-elicited gestures, in which subjects were asked to choose the gesture they thought was the best from all the user-defined options. One purpose of choice-based elicitation is to compare the judgments of different user groups, and the other is to test the rationality of gestural inputs in the prior user elicitation study by observing the difference in the top gesture between two studies. Next, we evaluate the practicality of designers' choices. This is performed to measure the interaction performance and inform the designers of the limitations of using the watch as a controller for visual data exploration.

B. DEFINING INTERACTION TASKS

We conducted the top-down coding to summarize the interaction tasks commonly used in visual analytics. According to this coding method, the basic categories of tasks should be defined first, and all the single commands could be then clustered into the categories. Nine task categories were identified by mainly referring to the classification framework proposed by Yi et al. [2], and then by adding a few other categories mentioned in related literature [11], [16], [17]. We described each category in detail with a short name and a phrase that interprets users' intent of performing the interaction tasks.

- Select: highlighting or tagging items of interest.
- Explore: examining a subset of the data or browsing its full picture.
- Edit: altering the visual appearance of data element or modifying the data values.
- Abstract: adjusting the level of abstraction of data from an overview to a very detail representation.
- Filter: presenting the set of data items based on a specified range or condition.
- Connect: creating relationships between data sets or items.
- Position: changing the spatial arrangement of representations.
- Add/Remove: adding or removing data representations.
- Others: other tasks not included in the above classifications.

Referring to the existing studies on visual analytics is an effective way to extract the necessary interaction tasks. By reviewing several of the key works [15], [16], [17], [19], a total of 80 tasks were collected. To enumerate the tasks not mentioned in existing literature, we invited five experts to participate in the generation of the task set, all of whom were familiar with the use of large displays and had at least five years of experience with visualization design. We prepared a blackboard, signature pens, sticky notes and large sheets of white paper. First, the experts were asked to write the tasks they considered to be higher frequently used in visual analytics on the blackboard, according to our taxonomy of interaction tasks. They needed to present at least two tasks per category. They could also take the brainstorming method to add tasks that had not been presented by other experts to one category. Then, the experts were asked to briefly annotate the proposed tasks and transcribed them on the sticky notes. Through these procedures, a total of 89 expert-elicited tasks were obtained. We transcribed the 80 tasks collected from literature on the sticky notes, resulting to a total of 89+80=169 tasks. If an expert-elicited task coincided with one collected from previous studies, these two would not be merged but all written on the notes for subsequent classification work.

Next, all the tasks were encoded deductively into the nine categories using the affinity diagram method [48]. To start with, the theme cards with category names were listed. After that, we mixed the 169 cards on a large sheet of white paper, and invited the experts to discuss how to group them. Experts pasted the cards they believed to be of the same category together, before merging and removing the tasks in each category according to the following three principles: (i) tasks with similar users' intent of interaction were merged into one; (ii) tasks with low frequency of use were removed; and (iii) tasks with overly complex operations were not considered.

There was a procedure by which experts decided how to merge or delete tasks. One expert firstly picked out two or more tasks with a similar meaning from the same category. For example, "change the hue of points in a scatter plot" and "change the transparency of those points" mentioned in Ahn et al. [16] is both to edit the chromatic value, which can be merged into "adjust the color of the graph". Before these tasks were finally merged, a collective discussion was usually made when experts argued that some of them bear different meanings. After reaching a consensus, this new and representative task would be put in a certain category. This process was repeated until each category contained several tasks which are the summary of the relevant elicited tasks. The experts also had to determine whether the merged tasks could be retained or not based on principle 2 and 3. At least three of the five experts must agree in order for this task to be retained. Regarding the removed tasks, one described as "find data cases with extreme values in the range of an attribute in the dataset" [15] was believed to be less frequent, and consequently it was removed from the task set. In the end, we defined 21 representative interaction tasks out of the 169 candidate tasks. Table 1 explains these tasks and indicates the categories within which they fall.

C. THE EXPERIMENTAL MATERIAL: A VISUALIZATION DESIGN

As the experimental material, the visualization design should facilitate the elicitation of general interaction design

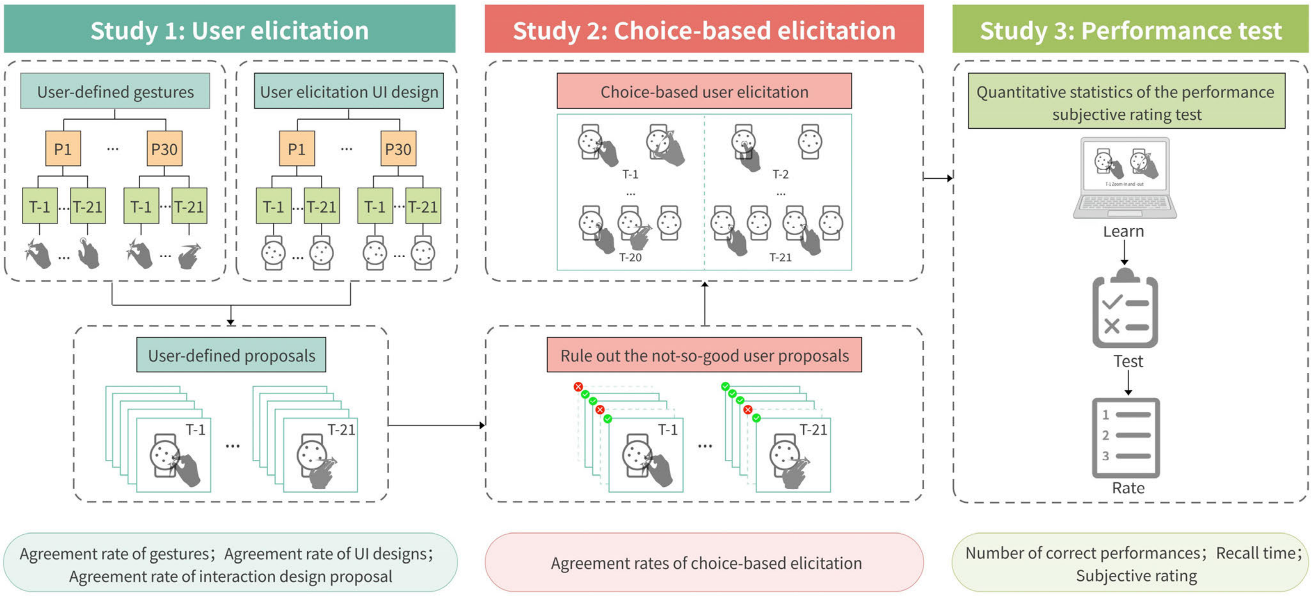


FIGURE 1. **Procedural details of the three stages of the study. At the bottom the statistical indicators for each stage are presented.**

proposals. Otherwise, the particular way of presenting data in some charts can induce the proposals that are less reasonable for other charts. Therefore, choosing the right graphical representations for the visualization is critical. First, these representations should be familiar to everyone. Second, the target objects (e.g., dots or squares representing data item) in the initial state of the chart should be displayed clearly and with a certain touchable area for the analyst to manipulate. Third, the form of the chart should be basically fit the layout of the dial, so that the analyst does not feel that he or she has to spend more operational steps in searching the target object. Fourth, the visualization should permit simulation of all the interaction tasks in real-world usage scenarios by linking them together through a theme of data analysis.

Based on the above principles, we designed a visualization depicting the trend of online shopping of Chinese college students as shown in Figure 2a. This topic was chosen because all the subjects in the elicitation study were college students. Histogram was applied in order to follow the first and second principles, while others like scatter plots and line plots were excluded. Chord diagram is circular, and can represent multiple categories of data that are connected to each other (Figure 2b). The two charts were integrated to make the entire visualization be a circular form by changing the histogram to a coxcomb diagram surrounding the chord diagram. The area of the slices of the coxcomb diagram can be scaled, reducing it to an arc-shaped histogram (Figure 2d). It highlights the amount of online purchases for different categories of goods and the number of purchasers in different grades. These two dimensions are represented by a pair of vertical bar charts with two different hues (Figure 2c). The color gradients within the bars represent the magnitude of the values. The interpretations of the amount of online purchases are hidden in the initial state before analysis (Figure 2e).

The visualization we designed allows the implementation of all 21 interaction tasks. Subjects will be presented with the initial state of the visualization for each analysis task and its state after the execution of that task on the large display. These two states are illustrated into two separate pictures, each for one slide of PowerPoint (Figure 3). According to Table 1, we define the users' intent of performing the tasks, so that the execution of them one after the other constitutes a story line of the complete analysis process. The story begins with the two tasks: Zoom-in and -out and Select, because the objects must be found and selected before subsequent tasks can be performed.

IV. STUDY 1: USER ELICITATION

A. SUBJECTS

A total of 30 volunteered subjects (16 males and 14 females) with more than one-year experience in visual analytics were recruited from a Chinese university. The subjects had a variety of majors, including design, journalism, engineering, and data science. The number of subjects from different majors was approximately equal. The age of the subjects ranged from 18 to 23 years old, with a mean of 19.6 years old (SD=1.7). These subjects were recruited because (i) college students of the same age range can represent the user groups who accept the same digital technology, and (ii) younger subjects usually have better comprehension and expression skills than the elder, making them easier for the researchers to communicate with. All the subjects did not have any experience in gesture design, and their dominant hand was the right hand. At the end of the user elicitation study, each

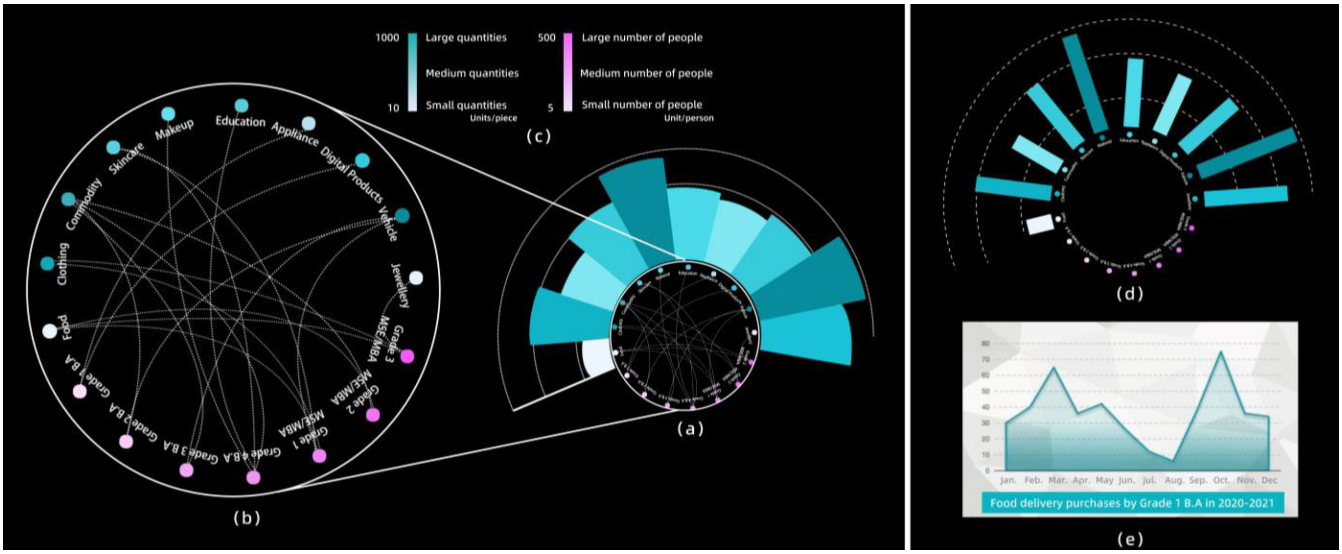


FIGURE 2. The visualization design for this study: (a) the integration of chord diagram and coxcomb diagram; (b) the enlarged chord diagram; (c) two bar charts indicating the magnitude of values; (d) scalable slices of the coxcomb diagram; (e) the interpretations of the amount of online purchases.

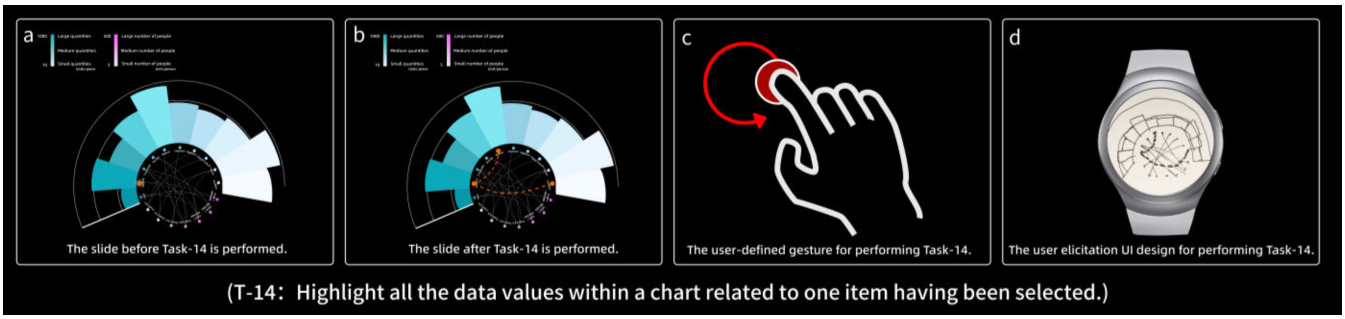


FIGURE 3. Presentation of the visualization: (a) showing the initial state of visualization on the large display before interaction with one slide; (b) showing the effect of interaction; (c) user defines the gesture; (d) user designs the UI of smartwatch in response to the gesture.

subject would receive a small gift to the value of 50 RMB as a reward.

B. APPARATUS

The user elicitation study was conducted in a usability laboratory with the space covering up to approximately 55 square meters. The 86-inch display (approx. 190 cm$\times$110 cm side length, 3840 $\times$ 2160 resolution) of a smart blackboard was used to present the slideshow. The subjects were provided with a smartwatch model copying the appearance of the Samsung Gear S2 watch to demonstrate gesture commands. The dial of this model was rotatable that it could be used as a mechanical input modality. A set of table and chair was placed in the lab, with several sheets of A3 white paper on the table. A high-definition camera was fixed on the table to record the subjects' performance for subsequent extraction of the gestures from the captured videos.

C. PROCEDURE

Prior to the initiation of the study, subjects were briefed about the purpose and requirements of the study and signed an informed consent form. The 21 tasks appeared in the same order as the numbering in Table 1, in accordance to a pre-written story line of visual analytics. At the beginning of each task, the PowerPoint software showed the initial state of the visualization with the name of that task and its description for the subjects to read and understand. Once the subject executed a gesture command, the researcher switched the slideshow to the next page to show the interaction effect of the visualization resulting from the task. The time limit between the start of the task and the completion of the gesture execution should not exceed one minute, forcing the subjects to intuitively suggest their most preferred interaction input. After performing the gesture, the subject was asked to follow the think-aloud protocol to give an account of (i) the number of fingers required to perform the gesture, (ii) the target area clicked by the fingers, (iii) the duration of each atomic gesture, and (iv) the reason why this certain gesture was recommended.

After the think-aloud session for explaining the gestural input, the subjects were asked to draw the UI of the smartwatch in response to the gesture command on a white sheet

TABLE 1. The twenty-one tasks defined for this study. T-1 is the abbreviation of Task-1, and so forth.

No.	Task	Category
T-1	Zoom-in and -out	Explore
T-2	Select	Select
T-3	Multi-Select	Select
T-4	Show more details of the selected graph or data	Others
T-5	Hide the selected graph or data	Others
T-6	Change the color of a selected data graph	Edit
T-7	Change the shape of a selected data graph	Edit
T-8	Adjust the font size	Edit
T-9	Modify the value of a data graph	Edit
T-10	Present data items based on one specific condition	Filter
T-11	Split a data set and present the data items of different categories separately	Filter
T-12	Move the data graph to a new location	Explore
T-13	Merge two groups of data	Connect
T-14	Highlight all the data values within a chart related to one item having been selected	Connect
T-15	Annotate the selected data item	Abstract
T-16	Juxtapose two or more data graphs	Position
T-17	Sequence data items according to a certain rule	Position
T-18	Add a visualized data item to the data set	Add/Remove
T-19	Delete a data item from the date set	Add/Remove
T-20	Restore to previous state	Others
T-21	Perform a redo operation	Others

of paper using a needle pen (Figure 4). As with the previous session, they were asked to verbally explain why the UI was designed in this way. If more than one UIs depicted by the subjects differed in layout, widget, and visual element, but the functions of the current page were essentially the same, the UIs were considered to be the same design proposal. At the end of each trial, we checked the consistency of the gestural input and the feedback of two display devices in each step of interaction. The study took about 75 minutes overall for each subject.

D. DATA ANALYSIS

A total of 651 gestures and UI design proposals for 21 tasks were collected from 30 subjects. The UI design proposals that were identical in main design characteristics were merged into one, leaving 52 proposals as a result. The gestures with the same dynamics were also merged. Taking "*T-4: Hide the selected graph or data" as an example, some subjects* "dragged the graph or data item to the edge of the screen to hide it", while some "using a long-press to drag the graph or data item out of the screen". For both gestures the graph or data item is dragged off the screen, and the direction of dragging is not important. Therefore, these two gestures can be grouped together. In the end, we identified 181 distinctive gestures. When combining the gesture and the UI as an interaction design proposal, there are 228 distinctive proposals.

We calculated the agreement rate of gestures, UIs and interaction design proposals for each task with the formula introduced by Vatavu & Wobbrock [49]. The higher the agreement rate, the more likely the subjects were to come up with the same proposal for the target task. The agreement rate (AR) was computed by AGATe. The four levels of agreement are low ($AR(r) \leq 0.100$), medium ($0.100 \leq AR(r) \leq 0.300$), high ($0.300 \leq AR(r) \leq 0.500$) and very high ($AR(r) > 0.500$) [49].

1) AGREEMENT RATE OF GESTURES

Figure 5 shows the agreement rate of the 21 tasks, and the average AR value is 0.303. Among them, the highest agreement rate was seen for T-1, which was 0.709, indicating that 26 out of 30 subjects chose the "pinch-to-zoom" gesture to execute this task. High agreement also appeared in T-6 (0.488) and 8 (0.400). Two subjects (P_3, P_{15}) reported that these were tasks for editing images or data value and were easier to perform through clicking a button from the menu. In contrast, the ARs for "*T-20: Restore to previous state*" and "*T-21: Perform a redo operation*" were both relatively the lowest (AR=0.107). Popular gestures for these tasks were "two-finger tap", "two-finger left/right swipe", "clockwise/counterclockwise rotation", etc. Some subjects indicated that the gesture they defined for this task were extracted from their daily use of smartphone and digital products. The difference in prior knowledge of using device

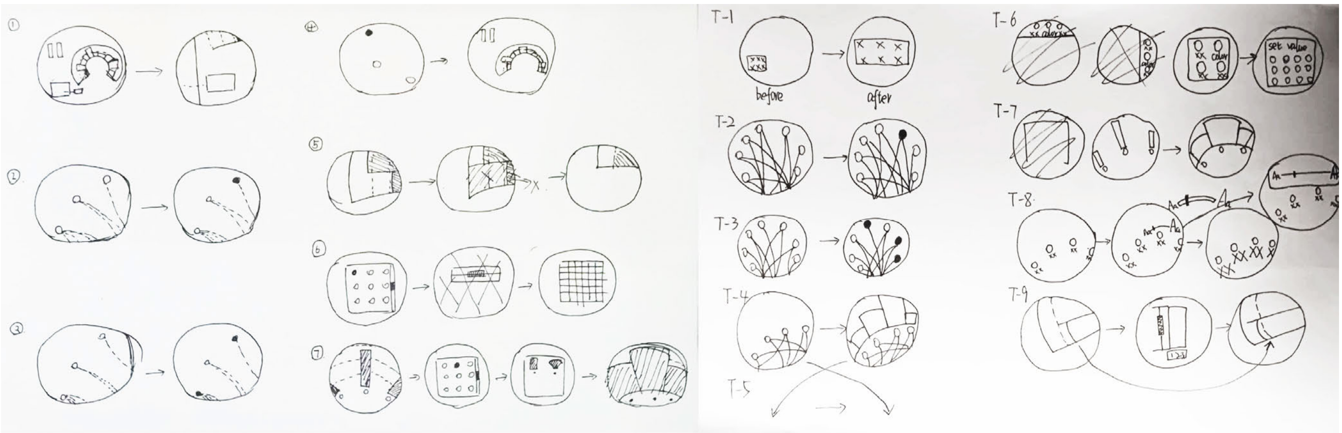


FIGURE 4. Two examples of subject's designs for the UI of smartwatch in elicitation study.

resulted the diversity in eliciting gestures for these two tasks. Medium agreement was observed for T-10 and T-17. Five subjects stated that the two tasks occurred less frequently on data analysis, and three of them even had never handled them before, resulting in arbitrary mappings of their personal gestures to the target task.

2) AGREEMENT RATE OF UI DESIGNS

The agreement rates of subjects' UI design for 21 tasks are also shown in Figure 5, with a mean value of 0.493. For "*T-2: Select*" and "*T-3: Multi-Select*", the agreement was very high (0.821). The majority of subjects preferred an enlarged view of the visualization which only displayed the focused area of the target to be selected. They felt that such a view made it easier to identify the desired item, improving the speed and accuracy of selecting it. T-9 had the lowest agreement rate of 0.201 among all the tasks, and the agreement rate of gestures for it was also low (0.142). The UI design was generally based on the gesture design, and so the high level of agreement on the UI often went with a consensus on the gestural input for one task. However, for T-1, there was a second-lowest AR of 0.263, in contrast to its high level of agreement in gesture. Some subjects (e.g., P_5 and P_{19}) wanted only a text or symbol shown on the watch to indicate the zoom-in operation was in progress, while others (e.g., P_{25}) hoped that the watch would display a miniature version of the zoom-in effect on the large screen. A considerable proportion of subjects suggested that the watch should only display the zoom-in percentage, allowing the analyst to know to what extent the object could be enlarged more.

3) AGREEMENT RATE OF INTERACTION DESIGN PROPOSAL

An interaction design proposal is composed of the gestural input and the UI design. Figure 5 shows the agreement rates of interaction design proposal for 21 tasks, with a mean value of 0.257. We observed 7 out of 21 tasks for which the most popular proposal received more than fifteen votes. They are: T-1, T-2, T-3, T-12, T-13, T-16, and T-19. Except for "*T-13: Merge two groups of data*", the other six tasks are more frequently used by the subjects and highly common across digital devices. The related use experience much contributed to the popularity of certain human-machine interaction. To execute T-13, 23 out of 30 subjects defined "press the object and swipe it to the target object" as its gestural input. When pressing the object to be merged, subjects suggested that there should be an object list on the watch to show every target the selected object would be swiped to. One subject stated that his first reaction to "merge data" was a dragging gesture for establishing a connection between two objects. In addition, a low level of agreement appeared in T-10 (0.093) and T-11 (0.088), with most subjects indicating that these tasks were not usually performed and therefore they would improvise some random gestures during the elicitation study. There was no clear explanation from these subjects as to why the particular gestures were presented.

4) GESTURE TAXONOMY

We examined the characteristics of all user-defined design proposals by building a classification schema based on gesture taxonomy. Deriving from the pioneering studies of categorizing gestures conducted by Kendon [50] and McNeill [51], a variety of classification schemata for gestures in the domain of interaction design has been proposed [52], [53], [54], [55]. To give an exact description of the elicited proposals, we specifically established a taxonomy of watch-based gestural interactions differing from the previous ones. This taxonomy included four dimensions: *form, semantics, input area and UI design*, each was divided into multiple categories, shown in Table 2. The frequencies of the proposals encoded as these categories are illustrated in Figure 6.

V. STUDY 2: CHOICE-BASED ELICITATION

A. SUBJECTS

Thirty subjects with experience in visualization design and interaction design were recruited, from universities,

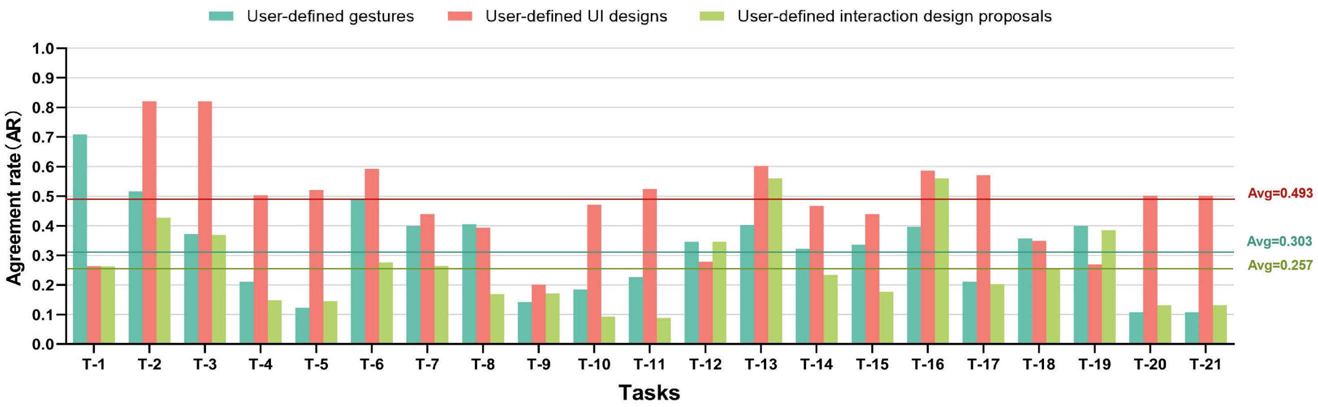


FIGURE 5. Agreement rates of user-defined gestures, UI designs and interaction design proposals.

TABLE 2. The taxonomy of proposals based on 651 user-defined designs.

Dimensions	Categories	Description
Form	one-move gesture	A consecutively performed gesture
	two atomic gestures	A compound gesture containing two individual gestures
	three atomic gestures	A compound gesture containing three individual gestures
	four atomic gestures	A compound gesture containing four individual gestures
Semantics	direct manipulation	Touch-based gestures like click, tap or drag
	deictic	Gesture indicates a position or direction
	metaphorical	Gesture indicates a metaphor
	abstract	Gesture-task mapping is arbitrary
Input area	finger(s) touch the watch's display	The input area is the watch's display
	manipulation of the physical user interface	The input area is the bezel of the watch or the skin
	other in-air movements	The input area is the space above or around the watch
UI design	a thumbnail view of the currently focused part of the visualization	The analyst interacts with the thumbnail view to execute the task
	pop-up menu	The analyst uses this menu to execute the task
	widget icon	The analyst executes the task with the help of icon
	functional pop-ups	The analyst uses the pop-ups to execute the task

companies, and independent design firms, to participate in Study 2. These subjects were considered experts because they had at least 5 years of study and practice in both of the two relevant fields. Their votes on the user-defined proposals in Study 1 were able to suggest the difference between experts and non-tech savvy users in understanding "better interaction". Eighteen of the thirty subjects are females and the rest are males with the age range from 22 to 32 ($\overline{X} = 26.78$, SD $=$ 2.62). All the subjects participated voluntarily and signed an informed consent form.

B. APPARATUS AND EXPERIMENTAL MATERIALS

This study was conducted in the same lab with the user-elicitation study. We first used the slideshow to show the experts the visualization interaction effect for the corresponding task on the large display to help them become familiar with the task list. To reduce the experiment time, each video recording of a user-defined gesture was reformatted into a GIF that would be played on an iPad. The middle of the GIF is the gesture action taken from a top view angle. On the left side, it shows the enlarged UI of smartwatch in response to the gestural manipulation; while the visualization interaction effect on the large screen resulting from that manipulation is displayed on the right side of the GIF. The current task is indicated at the top of the GIF, along with a textual description of the gestural manipulation and the UI, as shown in Figure 7. To maintain consistency in the experimental equipment, we provided the experts with the same model of watch as the subjects in Study 1. The rest of the hardware devices were also the same as in Study 1. In the watch we embedded a set of figures presenting the interface switches for every task, each of which was called up immediately upon subject's

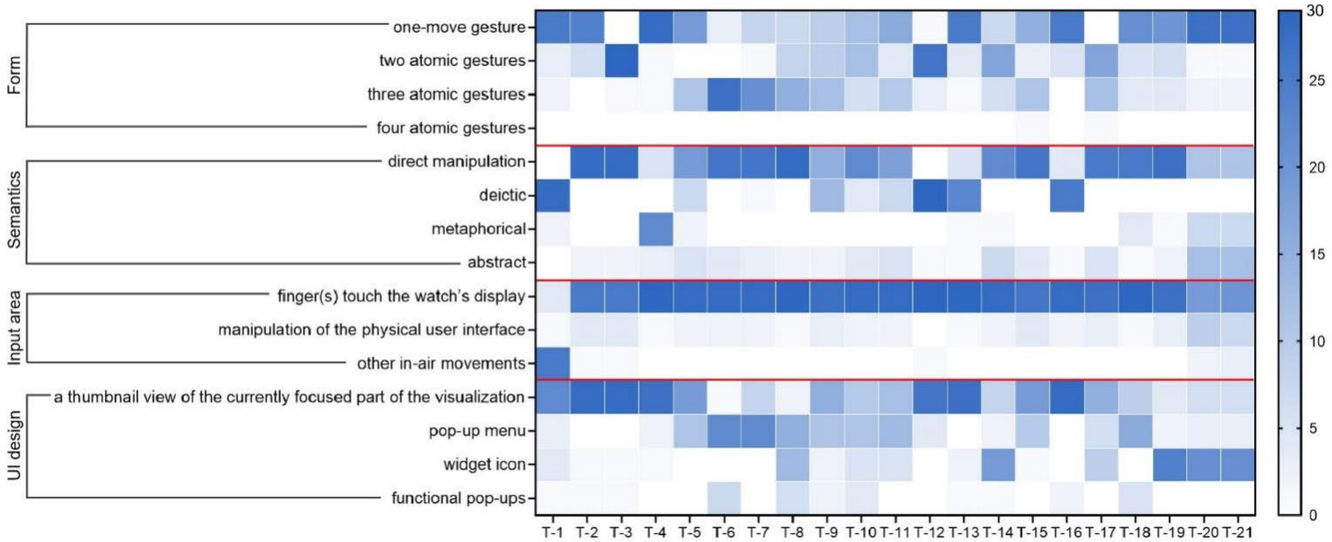


FIGURE 6. Distribution of user-defined gestures and UI designs across the taxonomy.

performance of the gesture. This was an experimental setting to perform the Wizard-of-Oz test.

C. PROCEDURE

Before initiating the choice-based elicitation, we considered some measures to ensure the validity of this approach. First, there should not be too many or too few options to make the subjects difficult to choose. Second, the very personalized options that seem clearly unreasonable should be ruled out. In order to reduce the number of options, we decided no more than six candidates of user-defined interaction for each task, according to the conventional maximum number of options that should be set. We formulated four rules of eliminating the not-so-good user proposals. First, all the most elicited proposals were selected; Second, the complex manipulation that required more than three atomic gestures were excluded; Third, gestures that existed in the current watch products and in previous studies were retained; Fourth, options that received only one vote had to meet both rule 2 and 3. Ultimately, a range of three to six options was used for all tasks. It is important to note that the options in the choice-based elicitation are the complete proposals of interaction combining gestures as the input and the UIs as the output.

During the process of elicitation, expert subjects were asked to demonstrate each gesture to better weigh how comfortable the manipulation was. They watched the GIF of each proposal firstly, and then immediately follow the animation to perform the gesture action. The time limit of this session could not exceed one minute, but the performance can be repeated. After all the options for each task had been experienced, the subjects needed to choose the one they thought was the best. To prevent subjects from forgetting the details of the gestural manipulation, they would make their choice while looking at the GIF again. The order of all tasks and the order of proposals for each task were randomized across subjects to prevent order effects. Afterwards, the subjects reviewed their choices in the sequence shown in Table 1 to check if there were any contradictions within the list of selected proposals. When the best interaction for each task was selected, the subject had to state its advantages. Once the proposal chosen by the expert was not the one by the user group in Study 1, the subject should explain the reason for his or her choice. The whole study lasted approximately for two hours per person.

D. RESULTS

Through Study 2, the expert subjects' attitudes to the highly-agreed interaction design proposals that user preferred were explored. According to the "winner-take-all" [6] methodology, the option with the highest number of votes was identified as the expert's preferred design. As a result, a set of expert-chosen interaction design proposals was generated, shown in Figure 8.

We calculated the agreement rates of the expert-preferred proposals using the same formula as in Study 1. As depicted in Figure 9, the average AR value for the 21 tasks is 0.458. This high level of AR may be due to the fact that fewer options in Study 2 cause a slighter chance of disagreements. "*T-15: Annotate the selected data graph*" had the lowest agreement rate, and the second lowest was "*T-9: Modify the value of a histogram*". The top three interaction design proposals for T-15 adopted three separate gestures: "long press the graph", "long press the center of the watch screen to open the function menu", and "make a flick gesture on the graph". The design received the most votes from experts was "make a flick gesture on the graph". In Study 1, it received only the third highest number of votes among all user-defined gestures. One expert (E18) argued that all the three gestures were applicable to this task, but the one-to-one mapping of the gesture to

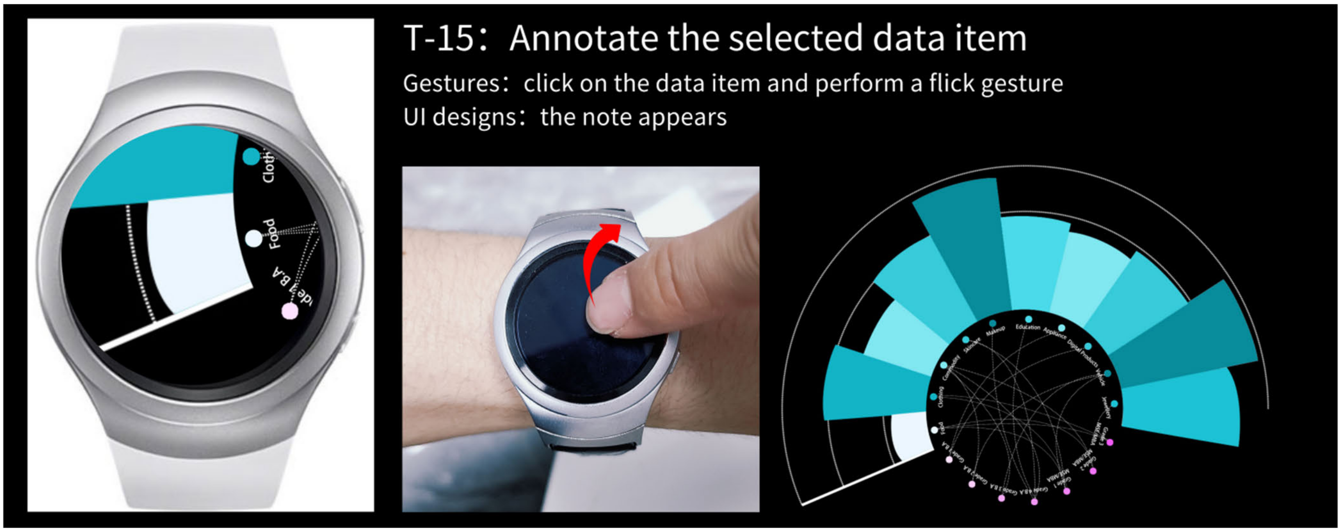


FIGURE 7. An example of the GIF pictures for choice-based elicitation.

a specific effect should be considered. Since "long press" was also assigned to T-3, the applicable category of task of this gesture should be clarified to avoid cognitive confusion between gestures. Comparing to the other tasks, T-1, T-4 and T-16 had higher agreement rates. There were six tasks with very high level of agreement rates and the rest are with high level.

Statistical analysis showed that the number of votes of some design proposals selected in Study 2 as the options for experts differed from that in Study 1. This difference was significant for T-16 ($\chi^2 = 33.77$, $p < 0.001$, $\varphi = 0.78$), T-2 ($\chi^2 = 13.27$, $p < 0.01$, $\varphi = 0.49$) and T-3 ($\chi^2 = 12.90$, $p < 0.01$, $\varphi = 0.52$). For the rest of the tasks, we detected a considerable difference for T-10 ($\chi^2 = 7.99$, $p = 0.10$, $\varphi = 0.41$) and T-4 ($\chi^2 = 4.56$, $p = 0.16$, $\varphi = 0.32$). Overall, the 2×N Chi-square tests found no fundamental inconsistency between the expert-preferred and user-preferred proposal for most tasks (Table 3). In particular, for T-1 and T-20, the difference between the two groups of subjects was assumed to be negligible ($p = 1$). In total, there were seven tasks in which the most elicited proposal from experts differed from the one from normal users, namely the analysts. In addition to T-2, T-3, T-4, T-10 and T-16, the other two are T-14 and T-15 (Figure 10). The most elicited proposals from the experts were designated as the final selected ones. By reviewing the experts' explanations for their choices, we analyzed the causes for these differences in preference as follows.

(1) *Select:* The most-frequently chosen proposal by the experts was "click the target object on the watch screen, then the selected object is highlighted, and a small icon appears at the edge of the screen after a few seconds". This design proposal was suggested by only two subjects in Study 1, because most of the subjects did not mention the icon but only the highlighting effect. Expert subjects, however, were more favorable to the design of icon. One expert (E25) pointed out that the floating icon could effectively handle with the situation when the selected object was moved off the screen before further operations. This icon design helps the analyst to quickly find the selected object so as to perform the analysis task.

(2) *Multi-Select:* The most preferred gestural input and UI of watch were inconsistent for this task between experts and normal users. Fourteen expert subjects chose "long press the first target object, then a pop-up appears with the other objects of the data set having been lined up for the user to click one by one". In Study 1, only one subject suggested this design. Most subjects would like to click on the targets one after another in the usual way. In Study 2, however, only eleven experts chose this option. Some experts (E8, E14) expressed that inserting data items into the pop-up enabled a focused view on a specific data set. This design insured the analysts against misclick and repeated operations.

(3) *Hide the selected graph or data:* The subjects in Study 1 preferred the "long press on the center of the watch screen to enter the menu with related functions" option. From the perspective of expert subjects, this solution not only requires more steps, but will also consume more time to find the corresponding button in the menu. By comparison, "drag the target object to the edge of screen" is a design of which the experts were more in favor.

(4) *Present data items based on one specific condition:* The majority of the experts (17 subjects) disagreed with the design of placing a function button in the menu. The representative opinion from E15 is that all the filter tasks should be evoked by a distinctive gesture. Amongst the limited number of options, expert subjects advocated "double-tap on the data graph". However, some of the subjects were concerned that this gesture was not strongly associated with the semantics of filtering.

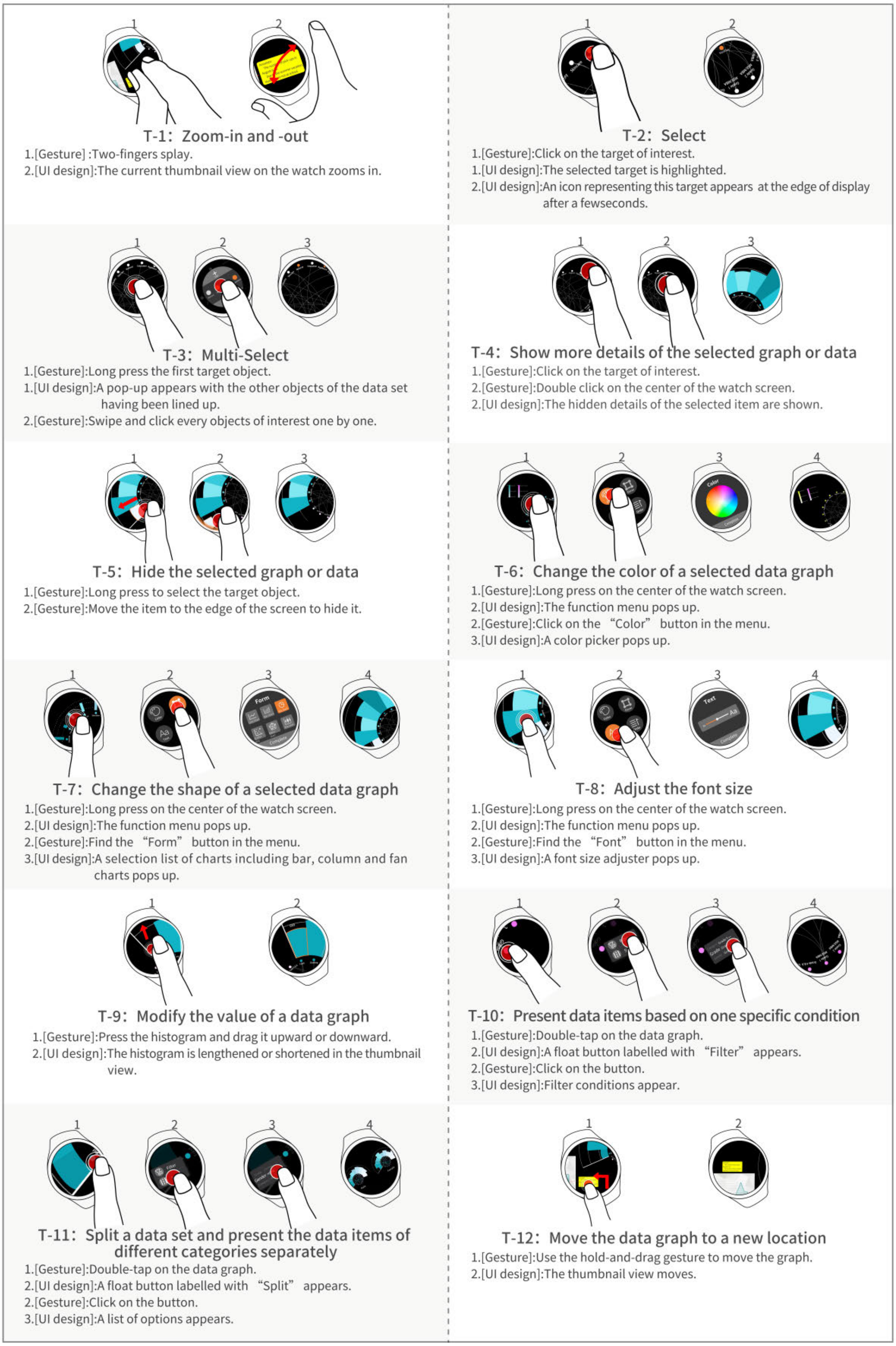


FIGURE 8. Selected design proposals chosen by expert subjects.

FIGURE 8. ***(Continued.)*** **Selected design proposals chosen by expert subjects.**

(5) *Highlight all the data values within a chart related to one item having been selected:* To show data items differently with some criteria, the most-frequently opted proposal in Study 1 was "when the target object is selected, the edges of relevant objects will be outlined for the user to select". Most experts preferred to express such a single command with a one-move gesture. The number of experts who chose "draw a circle around the watch screen" was the highest.

(6) *Annotate the selected data item:* For this task, the top gesture in Study 1 was "long press on the data graphic", while Study 2 was "click on the data item and perform a flick gesture". Some expert subjects who specified long press as the trigger for the pop-up menu argued that this gesture could

TABLE 3. Comparison of the subjects' preferences in Study 1 and 2.

Task No.	Frequency		identical selected proposal	χ^2	p	φ
	the most elicited proposal in Study 1	the most elicited proposal in Study 2				
T-1	22	25	Yes	0.436	1.000	0.09
T-2	20	12	No	13.272	0.002	0.49
T-3	17	14	No	12.902	0.001	0.52
T-4	9	26	Yes	4.556	0.164	0.32
T-5	9	18	No	4.928	0.285	0.32
T-6	13	13	Yes	4.307	0.372	0.27
T-7	15	15	Yes	5.279	0.236	0.32
T-8	10	17	Yes	1.264	0.978	0.16
T-9	8	13	Yes	1.426	0.933	0.17
T-10	5	17	No	7.988	0.104	0.41
T-11	5	14	Yes	2.888	0.623	0.25
T-12	16	15	Yes	0.562	0.809	0.18
T-13	23	24	Yes	1.892	0.835	0.18
T-14	12	15	No	4.864	0.298	0.30
T-15	9	13	No	5.217	0.380	0.31
T-16	23	26	No	33.766	<0.001	0.78
T-17	8	18	Yes	4.546	0.510	0.30
T-18	13	16	Yes	2.895	0.409	0.23
T-19	19	19	Yes	3.207	0.737	0.24
T-20	9	20	Yes	1.279	1.000	0.17
T-21	9	21	Yes	2.016	0.354	0.22

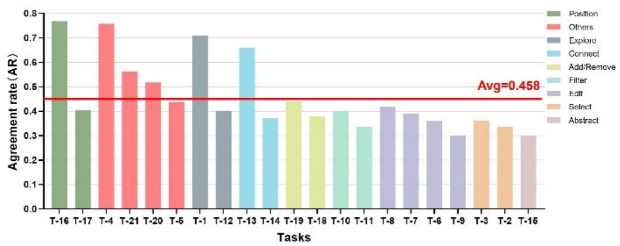


FIGURE 9. Agreement rates of choice-based elicitation.

not be mapped to another task, otherwise it would be error-prone.

(7) *Juxtapose two or more data graphics:* For this task, dragging was the gesture that most subjects agreed on. Expert subjects preferred dragging the graphics in a pop-up window, using left and right swipe to choose the position where the data item was placed. E3 strongly recommended adopting this interactive design, as the horizontally swiping would be an easy method for users to linearly explore objects. By comparison, it would be more cumbersome to constantly pan the dial display to find distant objects to be juxtaposed.

In summary, the experts' understanding of good interaction led to different results of Study 2 from Study 1. T-2 and T-3, as the two with the highest frequency of execution in practice among all tasks, were highly valued by both experts and subjects, while their opinions on the optimal solution were quite divergent. For some tasks only one or two experts' votes reversed the results of the elicitation study. This especially happened to T-2, for which 11 experts agreed the subjects' preference in Study 1 for "highlighting the selected object in the thumbnail view", while 12 chose the interaction proposal shown in Figure 9. The rough equality of scores indicated that each of these two proposals has its rationality. Many of the experts would give consideration of both operational experience and design creativity. They would not only design one command, but rather the interaction for the context of tasks, placing greater emphasis on the smoothness of the entire analysis process. Most of the expert subjects mentioned that scrolling of the function menu was to be considered. One design proposal recommended by them is using the back-and-forth rotation of the wrist to scroll up and down. Figure 11 illustrates the breakdown of the proposals chosen by expert subjects using our taxonomy. When comparing the results of Figure 7 and 12, we found that the distribution of proposals of normal subjects and expert designers are roughly the same (Figure 12).

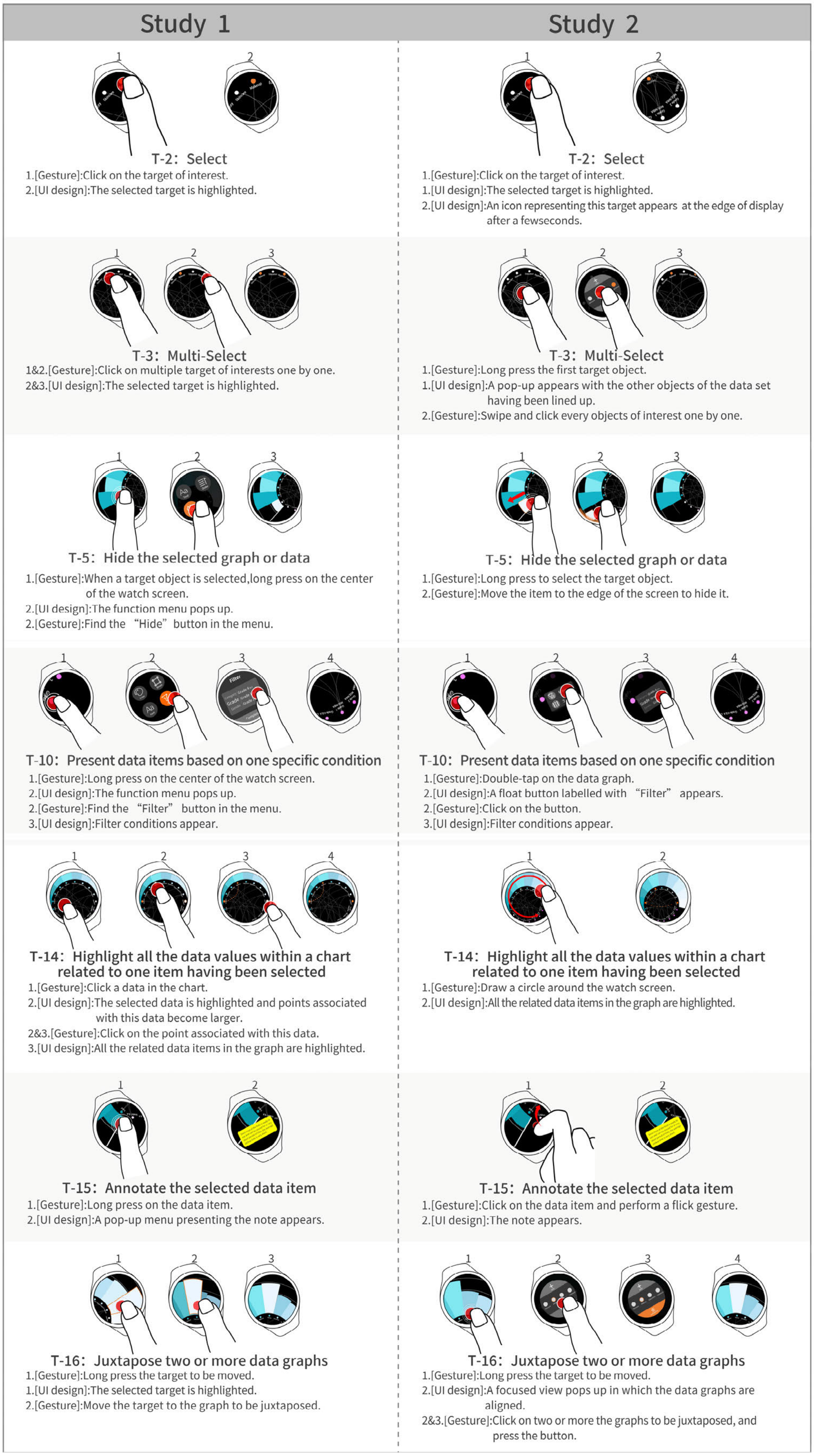


FIGURE 10. **The tasks with different most elicited proposals in Study 1 and 2.**

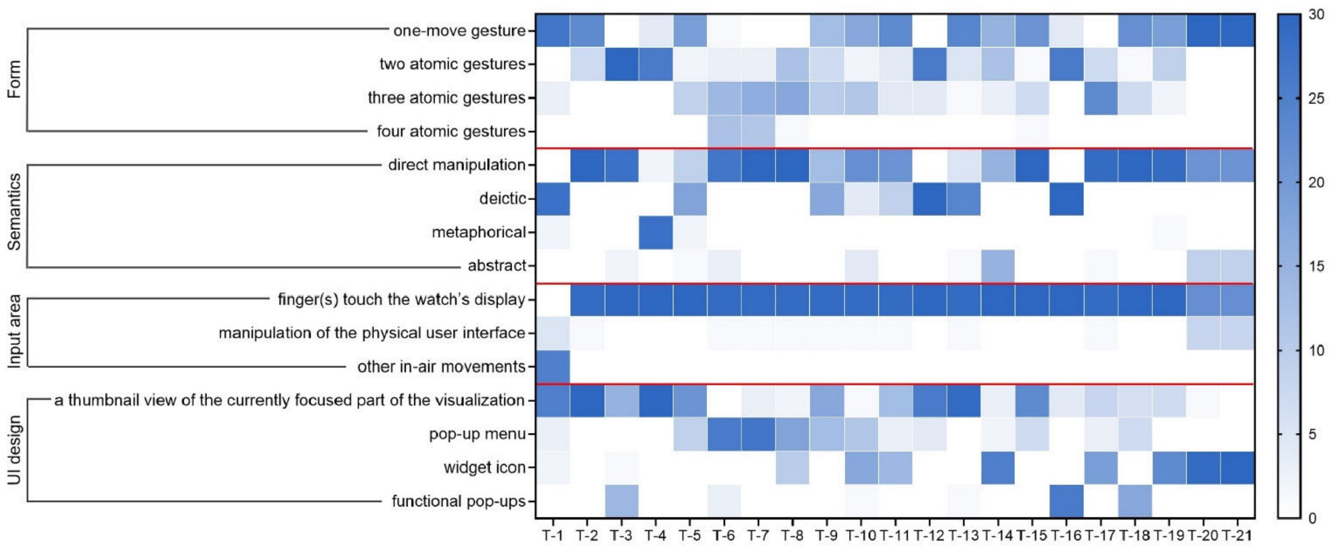


FIGURE 11. Distribution of expert-chosen gestures and UI designs across the taxonomy.

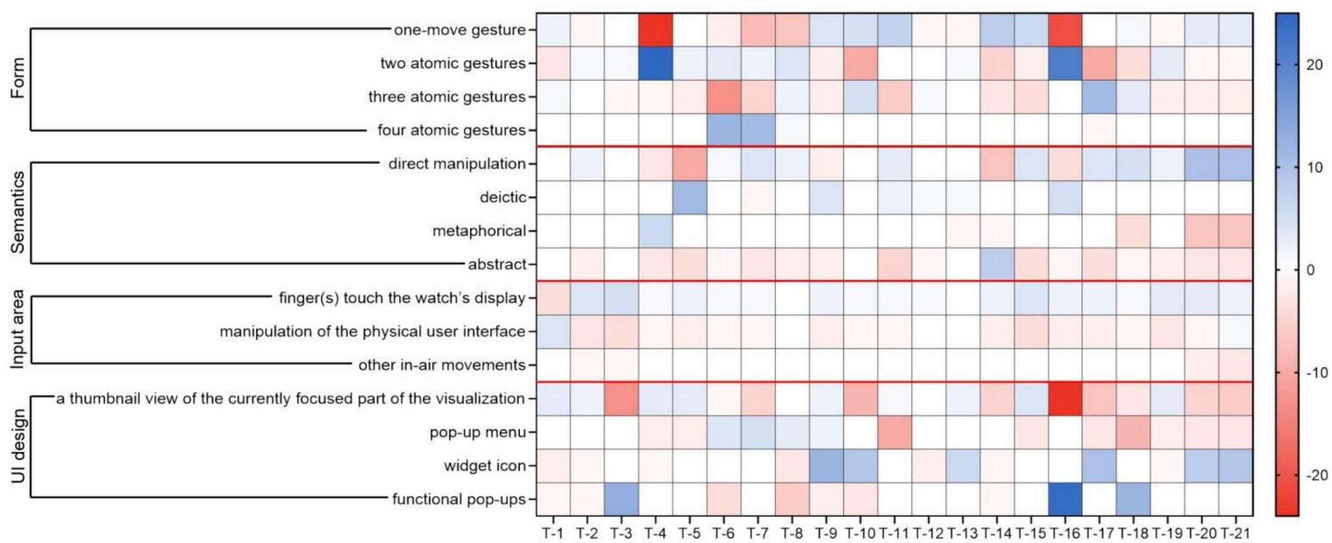


FIGURE 12. Difference of two distributions of gestures and UI designs across the taxonomy in Study 1 and 2.

VI. STUDY 3: PERFORMANCE TEST

Through Study 1 and 2, we derived a set of interaction design proposals that were more accepted by the subjects for controlling the on-screen visualization with a smartwatch. In this section, these proposals are tested through the Wizard-of-Oz method to verify their performance in practice, with respect to both memorability and usability.

A. SUBJECTS AND APPARATUS

Twenty-seven subjects participated in the test with a mean age of 22.15 years (SD = 1.21), including 11 females and 16 males. The subjects' demographic information and professional background are the same as the user group in Study 1, but neither have they participated the first two studies nor been experienced in gesture design. The subjects are all right-handers. Each subject was remunerated with 50 RMB for participation.

The setting of laboratory for the test was consistent with that in Study 1 and 2. GIF pictures acted as learning materials for the subjects. We used a static camera to capture the subjects' hand movements from which to analyze the accuracy of the gesture action and the time point at which the movement can be considered as the initiation of making gesture.

It also helped to record the subjects' self-report on the experience of their performance.

B. PROCEDURE

The experimental procedure consisted of three phases: learning, testing, and rating. In the learning phase, the subjects were presented with 21 tasks and their interaction effects on the large display. The presentations were sequenced according to the task frequency and priority in visual analytics, in line with the story line as described in Table 1. Subjects memorized the gestures by practicing them in this order until they confirmed that they could accurately express the execution process of every tasks. When this work was down, only then could the subject move on to the testing phase.

We arranged a pretest before the testing phase. The tasks in the pretest were randomly ordered for the subjects to demonstrate the correct gestures. Only if the subject successfully performed at least 80% of the gestures would he or she be allowed to start the formal test. The order of tasks in the formal test was also randomized but was different from the pretest. Using the watch prototype in Study 2, subjects were instructed to intuitively execute the gestural manipulation as soon as the prompt of a target task had been completely read out. Subjects would be not informed of any errors during the test. For each gesture, the available time for its performance was 30 seconds. Each task was repeated five times, so that the overall number of trials was 27 participants $\times$ 21 tasks $\times$ 5 repetitions = 2835.

C. MEASURES

The study carried out evaluation in two dimensions: quantitative statistics and subjective ratings. The quantitative data are about the efficiency and effect of learning the interaction design proposals. We measured these two aspects with two metrics: number of correct performances and recall time. If the subject performed the gesture exactly as the motion in GIF, it was regarded as correct performance. Recall time refers to the time interval from the moment the prompt finished to the moment the subject moved the right hand. The criteria for a correct performance included the accuracy in gesture dynamics, the correctness of the sequence of multi-step manipulations, and the exactness of input area. A gesture performance that did not meet any of these criteria was considered as incorrect. The recall time of an incorrect performance would not be statistically processed.

When the testing phase was finished, subjects were required to rate the design of gestural input and UIs for each task with a seven-point Likert scale. The scoring items were in the following four areas:

1. Comfort: Do you feel easy to physically perform the gesture? (1 = very hard, 7 = very easy)
2. Matching: Do you think there is a one-to-one match between the gesture and its target task? (1 = very low, 7 = very strong)
3. Promptness: Do you think the UI of smartwatch is prompting for your gestural manipulation? (1 = totally not, 7 = very useful prompt)
4. Learnability: Do you think the watch-based interaction can be easily remembered and learned? (1 = very hard, 7 = very easy)

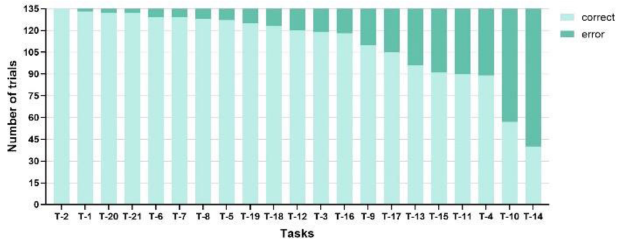


FIGURE 13. Number of correct performances and errors of 21 tasks.

D. RESULTS

1) NUMBER OF CORRECT PERFORMANCES

The number of correct trials for each of the 21 tasks correct is shown in Figure 13. The statistics showed that almost all the trials for T-1, T-2, T-20 and T-21 were correct. Relatively higher error rates were observed for T-4, T-10, T-11, T-13, T-14, and T-15. Among these tasks, T-10 and T-14 had unexpectedly higher than 50 percent error rates. We performed a 2×2 Chi-square test to execute pairwise comparisons of the 21 tasks in the number of correct trials. For T-14, it was found that the number of correct trials was significantly smaller than that for the other 20 tasks (T-14 vs. T-10: $\chi^2 = 1.90$, $p < 0.05$, $\varphi = 0.131$). Subjects (e.g., P6) suggested that it would be appealing but also unfamiliar to users to perform a metaphorical gesture on the small screen of the smartwatch to highlight all the data items of the same category. High-frequently used gestures such as tapping or pressing can be learned with ease, but not suitable for characterizing tasks of low frequency such as "highlighting". The number of correct trials for T-10 was significantly smaller than for T-4 ($\chi^2 = 15.27$, $p < 0.01$, $\varphi = 0.238$). Subjects (e.g., P17) reported that the gestures were easy to perform, but they needed more time to associate the "double-tap" gesture with presenting the filter condition. In contrast to the tasks mentioned above, there were fewer recall errors for T-19 and for tasks that required the pop-up of function menu. Subjects (P2, P7) said that they learned the gesture for T-19 quickly. The icon denoting delete instructed the subjects the area they can naturally hold and drag the target object to.

2) RECALL TIME

The recall time of gesture for each of the 21 tasks are illustrated in Figure 14. A Friedman test showed that there was a significant difference in the recall time for all the tasks ($p < 0.01$). According to the homogeneous subsets, the tasks were divided into 12 groups. The tasks with the longest recall time were T-14, secondly T-11 and then followed by T-10. The task that required the shortest recall time is T-12. The recall time for "*T-12: Move the data graph to a new location*" was extremely short, even slightly less than Select. A Wilcoxon signed-rank test revealed that the recall time for T-14 was significantly longer than any of the other 20 tasks (T-14 vs. T-11: $Z = -4.80$, $p < 0.01$). The results of subjective rating can further validate the quantitative data and explain the performance of the design proposals.

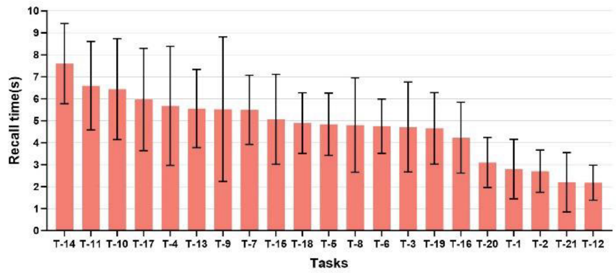


FIGURE 14. **Recall times of 21 tasks. Error bars denote the between-subjects standard deviation (SD).**

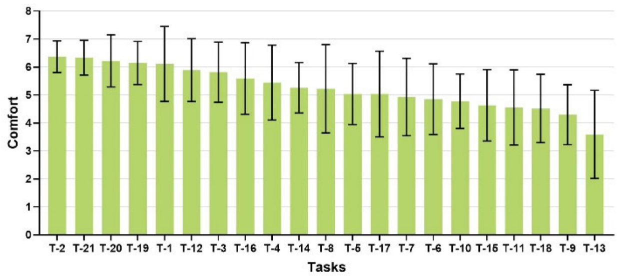


FIGURE 15. **Subjective ratings of Comfort. Error bars denote the between-subjects standard deviation (SD).**

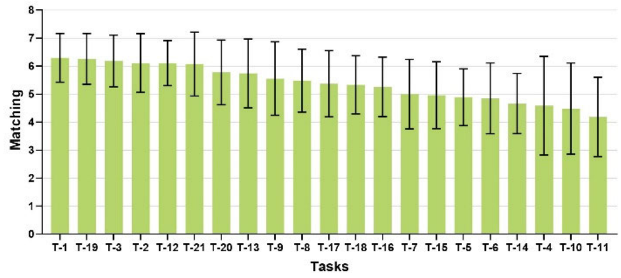


FIGURE 16. **Subjective ratings of Matching. Error bars denote the between-subjects standard deviation (SD).**

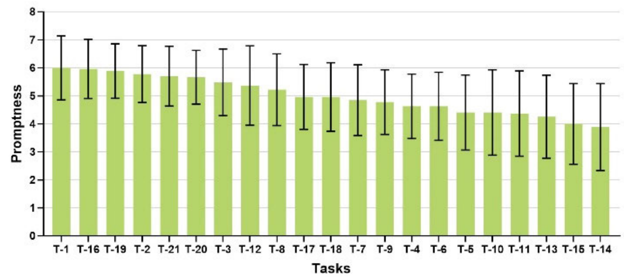


FIGURE 17. **Subjective ratings of Promptness. Error bars denote the between-subjects standard deviation (SD).**

3) SUBJECTIVE RATING

Figures 15-18 depict the four areas of the subjective ratings. We used the Wilcoxon signed-rank tests to execute statistical analysis. The proposals for T-20 and T-21 were highest rated in regard to *comfort* of gestural manipulation. Subjects stated they felt more comfortable when performing the swipe, tap, and drag gestures. For the gestures that require fine-grained adjustments, they felt much less comfortable. This is reflected by the low rating scores of the proposals for T-9 (T9 vs. T18: $Z = -1.97$, $p < 0.05$) and T-13 (T13 vs. T18: $Z = -4.84$, $p < 0.001$). Several repetitions of a simple gesture also increase the difficulty in performance, as is the case for gestural inputs for T-5, T-6, T-7 and T-10 (Figure 15).

Regarding the *matching score*, proposals for T-1, T-2, T-3, T-12 and T-19 were higher. For these tasks, subjects defined click or drag gestures borrowed from existing interactions. Surprisingly, ‘‘double tapping on the center of the watch screen’’ for T-4 had a very high agreement rate in Study 2, yet the average match score obtained in Study 3 was extremely low. When testing this design, the subjects made more errors in executing gestures than most of the proposals. They were prone to perform gestures such as long press (21 of 135 trials) or tap (14 of 135 trials) to show the details of data. Some of the subjects who made the long press gesture stated they could not differentiate the situation of using it from double-tapping. For T-10, T-11 and T-14, the matching scores of gesture were also relatively low (Figure 16).

Regarding the *promptness* of UI design for the watch, the use of a scroll menu for T-10 had the lowest rating, with a mean score of 3.89. The general attitude of the subjects to this design is that the system does not provide the user with a text box allowing them to submit the word they want. The rating of UI design for T-6 was significantly lower than T-8 ($Z = -5.89$, $p < 0.01$) and T-17 ($Z = -3.03$, $p < 0.01$), even though the pop-up menu was applied to all these tasks. Comparing to the design for T-8, the precise selection of color via a color picker displayed on the watch screen is more difficult. As to the presentation of options in the function menu, the design for T-6 is not so self-evident as those for T-8 and T-17 (Figure 17).

Regarding the item *learnability*, the interaction design for T-11 is ‘‘double tapping on the data graph to trigger a button which gives access to selecting the category of data’’. This proposal was rated as the most difficult to remember in its entirety. The double-tap gesture was also seen in the design for T-10. But as with that design proposal, the subjects (e.g., P10 and P20) stated that the double-tap was not easily discoverable. Moreover, the entire interaction demands more steps to complete. Furthermore, the process of finding keywords from the scroll menu is more complex due to the limited size of the watch screen. The issue of multiple steps for interaction was also raised by the subjects for every proposal using pop-up menu. Statistical analysis showed a significantly lower rating of T-10’s design in learnability than T-14’s ($Z = -3.88$, $p < 0.01$), though which was viewed as a design adopting an arbitrary mapping (Figure 18).

VII. DISCUSSION

A. EVALUATION OF THE USER-DEFINED INTERACTIONS AND TASKS

In the current research, a set of interaction designs connecting the large screen displaying the data visualizations to the smartwatch serving as a controller were elicited by

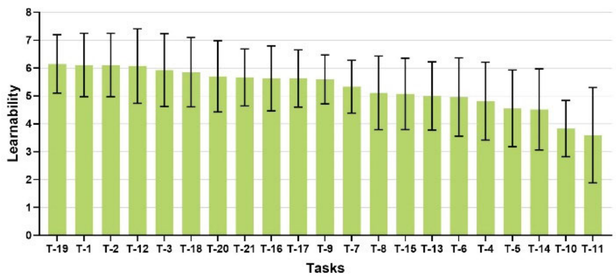


FIGURE 18. Subjective ratings of Learnability. Error bars denote the between-subjects standard deviation (SD).

Study 1 and 2. Study 3 was then launched to evaluate the user-defined design proposals. We use quartiles of ranking to assess the level of proposals as high, above average, lower-middle and low. Better proposals are reflected in relatively higher agreement rate, less difference between the preferences of design experts and non-experts, smaller number of errors, less recall time, and higher rating score. In line with these criteria, the proposals for T-19 and T-1 rank high on all metrics. These two proposals adopt the dragging and *splay* gesture that are commonly used in screen-based interactions. By checking the consistency between agreement rate and performance data of the final proposal (Figure 19), all tasks can be roughly divided into four categories. This classification scheme lacks support from statistical evidence, but can provide an overview of the evaluation results of user-defined interactions and suggest the key points of interaction design for visual analysis tasks.

T-13 and T-16 top the rankings jointly in respect of the agreement rate, but for some of the evaluation metrics these two tasks rank below average among all tasks. Overall, the performance of the proposal for T-16 is higher than that of T-13, but its ranking for subjective rating on matching is only 13. The expert-preferred proposal for T-16 is not the one the analysts mostly agreed in Study 1, so this kind of subjects seemed not to accept it very much in Study 3. For T-13, most subjects and experts agreed that the merging of data items demands an operation sequence which is composed of several fundamental subtasks. The inherent complexity of this task hinders users from learning it quickly, but does not cause highly diversified eliciting gestures.

T-2, 20, and 21 are another category of tasks, with lower overall agreement rates but higher rankings in performance. These three tasks are single commands and required in different scenarios as basic interactive tasks. Therefore, there will be more alternatives proposed by the subjects based on their own habits than other complex tasks. Nevertheless, the consensus of the subjects' preference for these tasks can be made to obtain the most elicited proposal. This proposal is user-friendly and easy-to-perform for the beginner users. The design of reverse gestures for the opposite referents [56] are certainly controversial, as different directional metaphors determine the association between gesture direction and semantics. The statistics show that the learnability of the proposals for T-1, T-2, T-16 and T-19 outperform the others. The proposals for T-20 and T-21 receive learnability scores lower than these four, probably because the subjects often perform the swiping gesture in its contrary direction.

We noticed some user-defined interactions for specific tasks that were more difficult to learn and prone to more mistakes, Moreover, the ranking of the agreement rate of these tasks is generally low. Representatives of such tasks are T-9, T-10, T-11, T-14, T-15 and T-18. They entail multi-step interactions resulted from their conceptual and procedural complexity. Based on the overview of the measurements, there are a couple of user proposals that show especially poor practical effects on the corresponding tasks. This means that the analysts without design experience are not very capable of giving mature proposals for complicated tasks in visual analytics. Representatives of such tasks are T-14, T-11 and T-10. It is necessary to have some user-preferred designs modified by design experts rather than just selecting them by the frequency of being proposed.

Except for T-1 and T-19, we did not find any other tasks with high agreement rate and ranking in performance metrics both, indicating that there is no agreement on the watch-based interaction for most visual analytics tasks, and the low precision of on-watch gestures also limits users' favor of expressive interactions.

B. DESIGN GUILDLINES

By summarizing and reviewing all the above results, we suggest following design guidelines for optimizing the watch-based interactions that enable seamless remote control of visual data displayed on the large screen.

1) ASSIGNING GESTURES OF THE SAME DYNAMIC CHARACTERISTICS TO THE TASKS OF THE SAME CATEGORY [57]

Some visual analytics tasks involve multi-step operations such as selecting multiple objects and conditional filtering. The gestural input of collected proposals for related tasks are comprised of two or more atomic gestures, which may represent operational steps that exist across heterogeneous interaction tasks. For example, a long press on the data graph makes the focused view pops up, and for other tasks the long press gesture can also but only trigger the focused view. Interactions for one certain category of task should be encoded in the same way. The triggering conditions, semantics of gesture and interface feedbacks of the inputs should also be unified to reduce learning and memory costs for users.

An evidence supporting this design suggestion is the elicited gestures exhibit certain biases in gesture taxonomy, indicating that the results of user elicitation may be influenced by legacy bias. In the *semantics* respect, direct manipulations and deictic gestures account for the higher portion. A very small percentage of abstract gestures is present for most tasks. As for the *input area*, gestures in which the finger(s) touch the watch's display are the dominant preference

	T-1		T-2		T-3		T-4		T-5		T-6		T-7		T-8		T-9		T-10		T-11	
	Average	Rank	Average	Rank	Average	Rank	Average	Rank	Average	Rank	Average	Rank	Average	Rank	Average	Rank	Average	Rank	Average	Rank	Average	Rank
Agreement rates of user-defined interaction design proposals ↑	0.262	9	0.427	3	0.369	5	0.148	16	0.145	17	0.276	7	0.264	8	0.169	15	0.171	14	0.093	20	0.088	21
Agreement rates of choice-based elicitation ↑	0.709	3	0.336	19	0.362	16	0.758	2	0.438	8	0.360	17	0.391	13	0.420	9	0.301	20	0.400	12	0.337	18
Number of errors ↓	2	20	0	21	16	10	46	3	8	14	6	16	6	16	7	15	25	8	78	2	45	4
Recall times ↓	2.80	18	2.70	19	4.72	14	5.68	5	4.84	11	4.76	13	5.50	8	4.80	12	5.56	6	6.44	3	6.60	2
Comfort ↑	6.11	5	6.37	1	5.81	7	5.44	9	5.04	12	4.85	15	4.93	14	5.22	11	4.30	20	4.78	16	4.56	18
Matching ↑	6.30	1	6.11	4	6.19	3	4.59	19	4.89	16	4.85	17	5.00	14	5.48	10	5.56	9	4.48	20	4.19	21
Promptness ↑	6.00	1	5.78	4	5.48	7	4.63	14	4.41	16	4.63	14	4.85	12	5.22	9	4.78	13	4.41	16	4.37	18
Learnability ↑	6.11	2	6.11	2	5.93	5	4.81	17	4.56	18	4.96	16	5.33	12	5.11	13	5.59	11	4.15	20	3.59	21

	T-12		T-13		T-14		T-15		T-16		T-17		T-18		T-19		T-20		T-21	
	Average	Rank	Average	Rank	Average	Rank	Average	Rank	Average	Rank	Average	Rank	Average	Rank	Average	Rank	Average	Rank	Average	Rank
Agreement rates of user-defined interaction design proposals ↑	0.346	6	0.56	1	0.234	11	0.177	13	0.560	1	0.202	12	0.255	10	0.385	4	0.131	18	0.131	18
Agreement rates of choice-based elicitation ↑	0.402	11	0.660	4	0.371	15	0.300	21	0.769	1	0.404	10	0.380	14	0.440	7	0.518	6	0.562	5
Number of errors ↓	15	11	39	6	95	1	44	5	17	9	30	7	12	12	10	13	3	18	3	18
Recall times ↓	2.18	21	5.53	7	7.60	1	5.07	9	4.23	16	5.97	4	4.90	10	4.65	15	3.10	17	2.20	20
Comfort ↑	5.89	6	3.59	21	5.26	10	4.63	17	5.59	8	5.04	12	4.52	19	6.15	4	6.22	3	6.33	2
Matching ↑	6.11	4	5.74	8	4.67	18	4.96	15	5.26	13	5.37	11	5.33	12	6.26	2	5.78	7	6.07	6
Promptness ↑	5.37	8	4..26	19	3.89	21	4.00	20	5.96	2	4.96	10	4.96	10	5.89	3	5.67	6	5.70	5
Learnability ↑	6.07	4	5.00	15	4.52	19	5.07	14	5.63	9	5.63	9	5.85	6	6.15	1	5.70	7	5.67	8

FIGURE 19. **The means and rankings of the selected proposal for each task in respect of all the evaluation metrics. Indicates the smaller the rank, the better the proposal is, while indicates the opposite.**

for all tasks. The analysts generally treat the smartwatch as a scaled-down version of the smartphone or tablet, persisting in the preference for touch interaction paradigm. Many of the atomic gestures, such as the dragging and *splay* gesture for T-19 and T-1, are taken from the screen-based interactions. As demonstrated by a fair number of existing studies, legacy bias is a contributing factor that users generate personally satisfying designs [58], [59], [60]. Due to the conscious use of previous knowledge [61], the designs that successfully leverage legacy bias can have a positive impact on their performance. When there is a spatially limited input area and a traditional task, interactions with high degree of freedom are less likely to be proposed. This makes the subjects more fixated to touch gestures, and believe that the gestural input can be more guessable and learnable if the watch and other mobile devices share the same interaction vocabulary. Designers are encouraged to leverage legacy bias with no outstanding issues to represent specific tasks or operational steps.

2) SIMPLIFYING UNNECESSARY OPERATIONS FOR SELECTING, BROWSING AND FINE-TUNING WITH INNOVATIVE DESIGNS

The choice-based elicitation results show that for seven of the tasks the most elicited proposal is not the one in user elicitation study. The expert designers explained that the proposals the analysts most preferred were either too unconstrained or too clumsy. An unconstrained input means it lacks design requirements for precise control and error prevention, resulting in repetition of basic operations. Such is the case with the final proposal in Study 1 for T-2, T-3 and T-18. As E3 said, the small size of watches "definitely limits precise clicks and frequent pan gestures to be performed, so selecting objects located in different areas of the large display only with one-by-one clicks can be particularly troublesome." A clumsy input refers to a design treating the smartwatch as "mini TV remote control equipped with many buttons" (E7's comment). With this feature, the user must find the functional button for the target task, such as the way the user proposals for T-5, T-10, T-14 and T-15 work. Too many times of click on the buttons within the menu could prolong performance of the task. Some of the expert designers advised that iconic or metaphorical gestures [62] could be used to represent single commands such as hiding, highlighting or displaying data items, as long as these gestures did not increase user's memory burden. The analysts should be inspired to define physical controls such as rotating the ring of a circle-like watch, making mid-air swipe gestures moving within a wide spatial range [63] [64] above the watch, and raising and lowering the wrist.

The key reason for the disagreement of preferred proposal between analysts and expert designers is the negative effects of non-designers' reliance on prior experience. Firstly, if the task requires multiple steps to complete, subjects tended to follow the interactions for multi-layered page and multi-touch technology. This made some designs not well adapted to the physical size of the smartwatch. For example, the final design proposal for T-13 was considered to be more difficult to be accurately dragged to the target object because of the fat finger effect [65] [66]. Secondly, if the subjects lacked experience of the interaction for executing a certain task, they would refer to the proposals they had already defined. For example, the second most elicited gesture for T-15 and T-18 in Study 1 was "long press the center of the watch screen to open the menu". It was also frequently proposed for and assigned as the final proposal for T-6, T-7, T-8 and T-17. Such practices might hinder the subjects from producing more creative designs.

3) IT IS ADVISABLE TO UTILIZE THE THUMBNAIL VIEW AND POP-UPS TO CATER TO USER PREFERENCES

First, the watch should display the thumbnail view of each independent visual data graph [67] rather than a partial area of the entire visualization. In the elicitation study, the subjects designed the interface of watch as the result of gesture operation. Using the gesture taxonomy, we see the design presenting the objects of interest on the large screen with a thumbnail view was most-frequently proposed in terms of *UI design.* Subjects did not mention the need to constantly move this view to display the target object on the watch in practice, which is critical for smooth experience of data analysis. Nevertheless, one focused view should be easily switched to another via some kind of action. A new page displaying simplified data graph can be invoked and popped up under this view. Simply performing the task on this page allows for more efficient and accurate interaction with the visualizations.

In addition to thumbnail view, the results of Study 1 and Study 2 indicated that the design proposal with higher agreement rate mostly used a pop-up menu in response to gesture operation. In particular, the design of pop-up menu considerably helped subjects to recall the interaction. Many subjects said that the operations on the menu page basically do not need to be remembered but just to search the buttons. However, the discoverability of pop-ups is a problem in addition to the so-called clumsy designs, as few subjects designed the visual hint or instructions on how to invoke the menu. Some subjects in Study 3 mentioned the long presses on the center of watch would accidentally touch data graphs located near the center point. These comments not only explain to some extent the medium level of user ratings on the related design proposal, but also suggested further improvement of the pop-up designs.

The interface design proposals for T-1, T-2, T-16, T-19, T-20 and T-21 were rated higher in terms of promptness. Except the one for T-1, the other five design proposals contained the widget icon or functional pop-ups. These designs enlarged the interface elements, as compared to treating the watch's display as a miniature view of the large screen. Moreover, they can directly inform the users of the next operation.

C. LIMITATIONS

Due to some limitations, our study needs to be further developed. First, we did not use a real smartwatch with real-time feedback to recognize gestural commands. An interactive prototype should be equipped with to improve the reliability and validity of the test results. Without this prototype, subjects would be not so cognizant of physical constraint of the dial on the expressivity of touch gesture. Therefore, some elicited proposals were not defined with strict triggering conditions of contents, nor did they consider the difficulty of performing small precise movements and the details of using continuous gestures [68]. As to the sampling method, the subjects in our studies are all habitual right-handers for daily activities, so it remains to be seen whether the dexterity of hand use can have an impact on the preference of gestural manipulation habits or not.

In contrast to classical user elicitation research, the effect of interacting with visualizations is not constant. It varies with different representations of the visual data, which accordingly may call for the gesture input specifically mapped to it. Although there is no direct evidence that our visualization design causes particular research findings, this effect can be somewhat detected in the user interpretations in Study 1 and 2. A typical example is the interpretation of the selected design proposal for T-14. Some expert designers disagreed with drawing a circle around the watch screen, arguing that this gesture ''just make use of the circularity of the chord diagram''. One expert (E6) stated if a Sankey diagram is used to link items under different categories, a circular swipe cannot be a good representation of highlighting all the relevant items. Given that thumbnails are the dominant user preference for the UIs of smartwatch, it is possible that their styles could elicit corresponding gestures. More work is need to detect the strength of this effect. Fortunately, we did not gather any negative comments from subjects about the visualization design itself.

VIII. CONCLUSION

This study can provide insight into the possibility of applying wearable devices as the remote control in visual analytics. In the current study, we explore watch-based input and output designs that are more adapted to user habits and cognitions from the perspective of user preferences. Through a classical elicitation study, we collect a set of watch-based gestures for representative visual analytics tasks and the user interface designs for the gestures. Expert users evaluate these interaction design proposals by choosing them again. Finally, the selected proposals are tested in terms of memorability and use experience, giving us design guidelines for better interaction in watch-based visual analytics.

In the first two stages of this study, we contrived to integrate users' and experts' opinions on the best interaction designs. Users' preference was double-edged that their proposals made reference to the existing products for good use but tended to be somewhat conservative. For designers, more attention should be paid to balancing natural interaction and user's habits. Similar interaction intentions should be expressed through a universal mapping between the interface and gestures, forming a unified interaction language for menus and functional gestures in order to reduce memory burden. These all require higher requirements for interaction design. In future work, we will focus on exploring other freehand inputs for data visualization interaction, which can break through the limitations of watch-based interaction in physical size while exploiting the advantages of wearable devices.

REFERENCES

[1] R. Maciejewski, S. Rudolph, R. Hafen, A. Abusalah, M. Yakout, M. Ouzzani, W. S. Cleveland, S. J. Grannis, and D. S. Ebert, ''A visual analytics approach to understanding spatiotemporal hotspots,'' *IEEE Trans. Vis. Comput. Graphics*, vol. 16, no. 2, pp. 205–220, Mar. 2010, doi: 10.1109/TVCG.2009.100.

[2] J. S. Yi, Y. A. Kang, J. Stasko, and J. A. Jacko, ''Toward a deeper understanding of the role of interaction in information visualization,'' *IEEE Trans. Vis. Comput. Graphics*, vol. 13, no. 6, pp. 1224–1231, Nov. 2007, doi: 10.1109/tvcg.2007.70515.

[3] M. Beaudouin-Lafon, "Instrumental interaction: An interaction model for designing post-WIMP user interfaces," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, The Hague, The Netherlands, 2000, pp. 446–453.

[4] M. Agarwal, A. Srinivasan, and J. Stasko, "VisWall: Visual data exploration using direct combination on large touch displays," in *Proc. IEEE Visualizat. Conf. (VIS)*, Vancouver, BC, Canada, Oct. 2019, pp. 26–30.

[5] D. Vogel and R. Balakrishnan, "Distant freehand pointing and clicking on very large, high resolution displays," presented at the *Proc. 18th Annu. ACM Symp. User Interface Softw. Technol.*, Seattle, WA, USA, Oct. 2005, doi: 10.1145/1095034.1095041.

[6] H. Wu, S. Zhang, J. Liu, J. Qiu, and X. Zhang, "The gesture disagreement problem in free-hand gesture interaction," *Int. J. Human–Computer Interact.*, vol. 35, no. 12, pp. 1102–1114, Jul. 2019, doi: 10.1080/10447318.2018.1510607.

[7] B. B. Velichkovsky, M. A. Rumyantsev, and M. A. Morozov, "New solution to the Midas touch problem: Identification of visual commands via extraction of focal fixations," in *Proc. 6th Int. Conf. Intell. Human Comput. Interact. (IHCI)*, Evry, France, 2014, p. 12.

[8] 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. SIGCHI Conf. Human Factors Comput. Syst.*, Toronto, ON, Canada, 2014, pp. 1063–1072.

[9] G. Arm. (2023). *What Does Gorilla Arm Mean?*. Accessed: Jan. 21, 2023. [Online]. Available: https://www.techopedia.com/definition/31480/gorilla-arm

[10] J. T. Hansberger, "Dispelling the gorilla arm syndrome: The viability of prolonged gesture interactions," in *Proc. 9th Int. Conf. Virtual, Augmented Mixed Reality (VAMR) Held 19th Int. Conf. Hum.-Comput. Interact. (HCI International)*, Vancouver, ON, Canada, 2017, pp. 505–520.

[11] M. Brehmer and T. Munzner, "A multi-level typology of abstract visualization tasks," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 2376–2385, Dec. 2013, doi: 10.1109/TVCG.2013.124.

[12] V. G. Motti and K. Caine, "Micro interactions and multi dimensional graphical user interfaces in the design of Wrist Worn wearables," in *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, 2015, vol. 59, no. 1, pp. 1712–1716.

[13] Y. Xiao, "Exploring user-preferred gestures for interaction tasks of data visualization on the large display," *Human Factors Syst. Interact.*, vol. 52, pp. 380–386, Jun. 2022.

[14] T. Horak, S. K. Badam, N. Elmqvist, and R. Dachselt, "When David meets goliath: Combining smartwatches with a large vertical display for visual data exploration," in *Proc. CHI Conf. Human Factors Comput. Syst.*, Montreal, QC, Canada, 2018, pp. 19–31.

[15] B. Lee, C. Plaisant, C. S. Parr, J.-D. Fekete, and N. Henry, "Task taxonomy for graph visualization," presented at the *Proc. AVI workshop BEyond time Errors: Novel Eval. methods Inf. Visualizat.*, Venice, Italy, May 2006, doi: 10.1145/1168149.1168168.

[16] J.-w. Ahn, C. Plaisant, and B. Shneiderman, "A task taxonomy for network evolution analysis," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 3, pp. 365–376, Mar. 2014, doi: 10.1109/TVCG.2013.238.

[17] N. Kerracher, J. Kennedy, and K. Chalmers, "A task taxonomy for temporal graph visualisation," *IEEE Trans. Vis. Comput. Graphics*, vol. 21, no. 10, pp. 1160–1172, Oct. 2015, doi: 10.1109/TVCG.2015.2424889.

[18] B. Saket, P. Simonetto, and S. Kobourov, "Group-level graph visualization taxonomy," 2014, *arXiv:1403.7421*.

[19] S. Gladisch, U. Kister, C. Tominski, R. Dachselt, and H. Schumann, "Mapping tasks to interactions for graph exploration and graph editing on interactive surfaces," 2015, *arXiv:1504.07844*.

[20] A. Srinivasan and J. Stasko, "Orko: Facilitating multimodal interaction for visual exploration and analysis of networks," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 1, pp. 511–521, Jan. 2018, doi: 10.1109/TVCG.2017.2745219.

[21] M. Tobiasz, P. Isenberg, and S. Carpendale, "Lark: Coordinating co-located collaboration with information visualization," *IEEE Trans. Vis. Comput. Graphics*, vol. 15, no. 6, pp. 1065–1072, Nov. 2009, doi: 10.1109/TVCG.2009.162.

[22] S. M. Drucker, D. Fisher, R. Sadana, J. Herron, and M. C. Schraefel, "TouchViz: A case study comparing two interfaces for data analytics on tablets," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Paris, France, 2013, pp. 2301–2310, doi: 10.1145/2470654.2481318.

[23] R. Sadana and J. Stasko, "Designing and implementing an interactive scatterplot visualization for a tablet computer," in *Proc. Int. Working Conf. Adv. Visual Interfaces*, May 2014, pp. 265–272.

[24] S. Al-Megren and R. A. Ruddle, "Comparing tangible and multi-touch interaction for interactive data visualization tasks," in *Proc. 10th Int. Conf. Tangible, Embedded, Embodied Interact.*, Eindhoven, The Netherlands, 2016, pp. 279–286.

[25] T. Chen, L. Xu, X. Xu, and K. Zhu, "GestOnHMD: Enabling gesture-based interaction on low-cost VR head-mounted display," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 5, pp. 2597–2607, May 2021, doi: 10.1109/TVCG.2021.3067689.

[26] C. Kirmizibayrak, N. Radeva, M. Wakid, J. Philbeck, J. Sibert, and J. Hahn, "Evaluation of gesture based interfaces for medical volume visualization tasks," presented at the *Proc. 10th Int. Conf. Virtual Reality Continuum Appl. Ind.*, Hong Kong, Dec. 2011, doi: 10.1145/2087756.2087764.

[27] S. Ntoa, C. Birliraki, G. Drossis, G. Margetis, I. Adami, and C. Stephanidis, "UX design of a big data visualization application supporting gesture-based interaction with a large display," in *Proc. 19th Int. Conf. Human-Comput. Interact. (HCI International)*, Vancouver, BC, Canada, 2017, pp. 248–265.

[28] F. Brudy, C. Holz, R. Rädle, C.-J. Wu, S. Houben, C. N. Klokmose, and N. Marquardt, "Cross-device taxonomy: Survey, opportunities and challenges of interactions spanning across multiple devices," in *Proc. CHI Conf. Human Factors Comput. Syst.*, Glasgow, U.K., 2019, pp. 562–590.

[29] T. Neate, M. Jones, and M. Evans, "Cross-device media: A review of second screening and multi-device television," *Pers. Ubiquitous Comput.*, vol. 21, no. 2, pp. 391–405, Apr. 2017, doi: 10.1007/s00779-017-1016-2.

[30] P. Hamilton and D. J. Wigdor, "Conductor: Enabling and understanding cross-device interaction," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, Toronto, ON, Canada, 2014, pp. 2773–2782.

[31] S. Houben and N. Marquardt, "WatchConnect: A toolkit for prototyping smartwatch-centric cross-device applications," in *Proc. 33rd Annu. ACM Conf. Human Factors Comput. Syst.*, Seoul, Republic of Korea, 2015, pp. 1247–1256.

[32] C. Kray, D. Nesbitt, J. Dawson, and M. Rohs, "User-defined gestures for connecting mobile phones, public displays, and tabletops," presented at the *Proc. 12th Int. Conf. Hum. Comput. Interact. Mobile Devices Services*, Lisbon, Portugal, 2010, doi: 10.1145/1851600.1851640.

[33] Y. Xu, L. Wang, Y. Xu, S. Qiu, M. Xu, and X. Meng, "Cross-device task interaction framework between the smart watch and the smart phone," *Pers. Ubiquitous Comput.*, vol. 25, no. 6, pp. 1039–1049, Dec. 2021, doi: 10.1007/s00779-019-01280-7.

[34] N. Marquardt, K. Hinckley, and S. Greenberg, "Cross-device interaction via micro-mobility and F-formations," in *Proc. 25th Annu. ACM Symp. User Interface Softw. Technol.*, Cambridge, MA, USA, 2012, pp. 13–22.

[35] A. Stylianidis, J. Vermeulen, S. Houben, L. MacDonald, and R. Beale, "SenseBelt: A belt-Worn sensor to support cross-device interaction," in *Proc. CHI Conf. Extended Abstr. Human Factors Comput. Syst.*, Denver, CO, USA, 2017, pp. 2123–2131.

[36] D. Schmidt, J. Seifert, E. Rukzio, and H. Gellersen, "A cross-device interaction style for mobiles and surfaces," in *Proc. Designing Interact. Syst. Conf.*, Newcastle, U.K., 2012, pp. 318–327.

[37] S. McGuckin, S. Chowdhury, and L. Mackenzie, "Tap 'n' shake: Gesture-based smartwatch-smartphone communications system," in *Proc. 28th Austral. Conf. Comput.-Hum. Interact. OzCHI*, Launceston, TAS, Australia, 2016, pp. 442–446.

[38] M. M. Luna, T. P. Carvalho, F. A. A. M. N. Soares, H. A. D. Nascimento, and R. M. Costa, "Wrist player: A smartwatch gesture controller for smart TVs," in *Proc. IEEE 41st Annu. Comput. Softw. Appl. Conf. (COMPSAC)*, Torino, vol. 2, Jul. 2017, pp. 336–341.

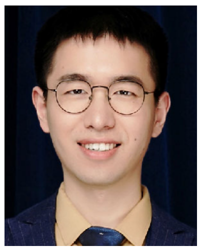
[39] R. Xiao, G. Laput, and C. Harrison, "Expanding the input expressivity of smartwatches with mechanical pan, twist, tilt and click," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, Toronto, ON, Canada, 2014, pp. 193–196.

[40] N. Marquardt, F. Brudy, C. Liu, B. Bengler, and C. Holz, "SurfaceConstellations: A modular hardware platform for ad-hoc reconfigurable cross-device workspaces," in *Proc. CHI Conf. Human Factors Comput. Syst.*, Montreal QC, Canada, 2018, pp. 354–368.

[41] P. Woźniak, L. Lischke, B. Schmidt, S. Zhao, and M. Fjeld, "Thaddeus: A dual device interaction space for exploring information visualisation," in *Proc. 8th Nordic Conf. Human-Computer Interaction: Fun, Fast, Foundational*, Helsinki, Finland, 2014, pp. 41–50.

[42] R. Langner, T. Horak, and R. Dachselt, "VisTiles: Coordinating and combining co-located mobile devices for visual data exploration," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 1, pp. 626–636, Jan. 2018, doi: 10.1109/TVCG.2017.2744019.

[43] D. Coffey, N. Malbraaten, T. Le, I. Borazjani, F. Sotiropoulos, and D. F. Keefe, "Slice WIM: A multi-surface, multi-touch interface for overview+detail exploration of volume datasets in virtual reality," in *Proc. Symp. Interact. 3D Graph. Games*, FSan Francisco, CA, USA, 2011, pp. 191–198.
[44] 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. Graph. Forum*, vol. 36, no. 3, pp. 503–514, Jun. 2017.
[45] H. Sollich, U. von Zadow, T. Pietzsch, P. Tomancak, and R. Dachselt, "Exploring time-dependent scientific data using spatially aware mobiles and large displays," in *Proc. ACM Int. Conf. Interact. Surf. Spaces*, Niagara Falls, ON, Canada, 2016, pp. 349–354.
[46] C. Yang, "Visualizing large time-series data on very small screens," in *Proc. IEEE VGTC/Eurographics Conf. Vis.*, 2017, pp. 37–41, doi: 10.2312/EUROVISSHORT.20171130.
[47] N. K. Dim, C. Silpasuwanchai, S. Sarcar, and X. Ren, "Designing mid-air TV gestures for blind people using user- and choice-based elicitation approaches," in *Proc. ACM Conf. Designing Interact. Syst.*, Brisbane, QLD, Australia, 2016, pp. 204–214.
[48] A. Lucero, "Using affinity diagrams to evaluate interactive prototypes," in *Proc. 15th IFIP TC.13 Int. Conf. Hum.-Comput. Interact. (INTERACT)*, Bamberg, Germany, 2015, pp. 231–248.
[49] R.-D. Vatavu and J. O. Wobbrock, "Formalizing agreement analysis for elicitation studies: New measures, significance test, and toolkit," in *Proc. 33rd Annu. ACM Conf. Human Factors Comput. Syst.*, Seoul, Republic of Korea, 2015, pp. 1325–1334.
[50] A. Kendon, "How gestures can become like words," in *Cross-Cultural Perspectives in Nonverbal Communication*. Ashland, OH, USA: Hogrefe & Huber Publishers, 1988, pp. 131–141.
[51] D. McNeill, *Hand and Mind: What Gestures Reveal About Thought*. Boston, MA, USA:: Univ. of Chicago Press, 1992.
[52] J. O. Wobbrock, M. R. Morris, and A. D. Wilson, "User-defined gestures for surface computing," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, Boston, MA, USA, 2009, pp. 1083–1092.
[53] J. Ruiz, Y. Li, and E. Lank, "User-defined motion gestures for mobile interaction," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, Vancouver, BC, Canada, 2011, pp. 197–206.
[54] S. S. Arefin Shimon, C. Lutton, Z. Xu, S. Morrison-Smith, C. Boucher, and J. Ruiz, "Exploring non-touchscreen gestures for smartwatches," in *Proc. CHI Conf. Human Factors Comput. Syst.*, San Jose, CA, USA, 2016, pp. 3822–3833.
[55] H. Wu and L. Yang, "User-defined gestures for dual-screen mobile interaction," *Int. J. Human–Computer Interact.*, vol. 36, no. 10, pp. 978–992, Jun. 2020, doi: 10.1080/10447318.2019.1706331.
[56] S. Villarreal-Narvaez, J. L. Perez-Medina, and J. Vanderdonckt, "Exploring user-defined gestures for lingual and palatal interaction," *J. Multimodal User Interfaces*, vol. 17, no. 3, pp. 167–185, Sep. 2023, doi: 10.1007/s12193-023-00408-7.
[57] 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 Technol. Interact.*, vol. 4, no. 4, p. 88, Dec. 2020.
[58] I.-A. Zaiţi, Ş.-G. Pentiuc, and R.-D. Vatavu, "On free-hand TV control: Experimental results on user-elicited gestures with leap motion," *Pers. Ubiquitous Comput.*, vol. 19, nos. 5–6, pp. 821–838, Aug. 2015, doi: 10.1007/s00779-015-0863-y.
[59] 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. ACM Designing Interact. Syst. Conf.*, Eindhoven, The Netherlands, 2020, pp. 855–872.
[60] 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.
[61] A. Macaranas, A. N. Antle, and B. E. Riecke, "What is intuitive interaction? Balancing users' performance and satisfaction with natural user interfaces," *Interacting Comput.*, vol. 27, no. 3, pp. 357–370, May 2015, doi: 10.1093/iwc/iwv003.
[62] D. McNeill, *Gesture and Thought*. Chicago, IL, USA: University of Chicago Press, 2005.
[63] A. Colley, J. Vayrynen, and J. Hakkila, "Exploring the use of virtual environments in an industrial site design process," in *Proc. 15th IFIP TC.13 Int. Conf. Hum.-Comput. Interact. (INTERACT)*, Bamberg, Germany, 2015, pp. 363–380.
[64] T. Moscovich, "Multi-touch interaction," in *Proc. CHI Extended Abstr. Human Factors Comput. Syst.*, Montréal, QC, Canada, 2006, pp. 1775–1778.
[65] K. A. Siek, Y. Rogers, and K. H. Connelly, "Fat finger worries: How older and younger users physically interact with PDAs," in *Human-Computer Interaction–Interact 2005* (Lecture Notes in Computer Science), Rome, Italy, 2005, pp. 267–280.
[66] S. Voida, M. Tobiasz, J. Stromer, P. Isenberg, and S. Carpendale, "Getting practical with interactive tabletop displays: Designing for dense data, 'fat fingers,' diverse interactions, and face-to-face collaboration," in *Proc. ACM Int. Conf. Interact. Tabletops Surf.*, Nov. 2009, Paper 3.
[67] W. Willett, Q. Lan, and P. Isenberg, "Eliciting multi-touch selection gestures for interactive data graphics," in *Proc. Eurographics Conf. Vis.*, 2014, pp. 79–83.
[68] N. Marquardt, R. Jota, S. Greenberg, and J. A. Jorge, "The continuous interaction space: Interaction techniques unifying touch and gesture on and above a digital surface," in *Proc. 13th IFIP TC 13 Int. Conf. Hum.-Comput. Interact. (INTERACT)*, Lisbon, Portugal, 2011, pp. 461–476.



YIQI XIAO received the Ph.D. degree in design from Hunan University, China, in 2018. He is currently an Assistant Professor and a Master's Supervisor with the University of Shanghai for Science and Technology, China. His research interests include multi-model interaction design, design methodology, and the cross-domain between user interface and cognitive linguistics.

LU LIU received the B.S. degree in industrial design from Zhongyuan University of Technology, Zhengzhou, China, in 2022. She is currently pursuing the M.S. degree with the College of Communication and Art Design, University of Shanghai for Science and Technology, China. Her research interests include human–machine interaction and gesture elicitation study.