# BotMap: Non-Visual Panning and Zooming with an Actuated Tabletop Tangible Interface

JULIE DUCASSE, MARC MACÉ, and BERNARD ORIOLA, CNRS IRIT, Université Toulouse 3
CHRISTOPHE JOUFFRAIS, CNRS IRIT, Université Toulouse 3, CNRS IPAL

The development of novel shape-changing or actuated tabletop tangible interfaces opens new perspectives for the design of physical and dynamic maps, especially for visually impaired (VI) users. Such maps would allow non-visual haptic exploration with advanced functions, such as panning and zooming. In this study, we designed an actuated tangible tabletop interface, called BotMap, allowing the exploration of geographic data through non-visual panning and zooming. In BotMap, small robots represent landmarks and move to their correct position whenever the map is refreshed. Users can interact with the robots to retrieve the names of the landmarks they represent. We designed two interfaces, named Keyboard and Sliders, which enable users to pan and zoom. Two evaluations were conducted with, respectively, ten blindfolded and eight VI participants. Results show that both interfaces were usable, with a slight advantage for the Keyboard interface in terms of navigation performance and map comprehension, and that, even when many panning and zooming operations were required, VI participants were able to understand the maps. Most participants managed to accurately reconstruct maps after exploration. Finally, we observed three VI people using the system and performing a classical task consisting in finding the more appropriate itinerary for a journey.

CCS Concepts: • **Human-centered computing → Interaction techniques**;

Additional Key Words and Phrases: Visual impairment, non-visual interaction, tangible interaction, actuated interface, tangible user interface, pan, zoom, tactile map, interactive map

**ACM Reference format:**
Julie Ducasse, Marc Macé, Bernard Oriola, and Christophe Jouffrais. 2018. BotMap: Non-Visual Panning and Zooming with an Actuated Tabletop Tangible Interface. *ACM Trans. Comput.-Hum. Interact.* 25, 4, Article 24 (September 2018), 42 pages.
https://doi.org/10.1145/3204460



## 1 INTRODUCTION

The ability to pan and zoom is mandatory when one wants to explore a map. For sighted users, pinch-to-zoom and drag-to-pan gestures have become basic gestures to navigate a map. For visually impaired (VI) users, however, these functionalities have very rarely been implemented and studied. Up to now, research has either focused on digital maps or hybrid maps (see [10] for a review). Maps that are purely digital can be displayed on a flat screen or projected onto a surface. Even though they can be easily refreshed, their exploration is usually constrained to a single point

Authors' addresses: J. Ducasse, M. Macé, B. Oriola, and C. Jouffrais, IRIT-ELIPSE, Université P. Sabatier, 118 route de Narbonne, F-31062 Toulouse Cedex 09, France; emails: {Julie.Ducasse, Marc.Mace, Bernard.Oriola, Christophe.Jouffrais}@irit.fr.

1073-0516/2018/09-ART24 $15.00
https://doi.org/10.1145/3204460

of contact, such as a finger or a virtual cursor. This makes exploration highly sequential (i.e., the user can only perceive the pieces of information one after the other) and challenging from perceptual and cognitive perspectives [38]. Interactive raised-line maps, the most common type of hybrid maps, are both digital and physical: users can interact with them using their two hands [5], which makes the exploration less sequential and less cognitively demanding. However, their content cannot be easily updated, making it impossible to implement panning and zooming.

The development of shape-changing or actuated interfaces opens new avenues for the design of physical maps that could support panning and zooming. For example, we can mention two prototypes: one relies on a refreshable tactile display [77], the other, named Linespace, on a 3D-printer [67]. Even though promising, the former is prohibitively expensive and the latter is currently too slow to be usable. On the contrary, actuated tabletop tangible interfaces have not been investigated for displaying non-visual maps (i.e., maps that are not displayed with visual elements only, such as tactile maps), but they are of particular interest, in terms of cost, physicality, and dynamicity. They are composed of several actuated tangible objects that can serve as tools to interact with the display or represent pieces of information, such as landmarks [16], and can support panning and zooming.

On the other hand, even though a few prototypes implemented non-visual panning and zooming [49, 65], the design of suitable interaction techniques for panning and zooming has very rarely been considered. In addition, the effect of panning and zooming on mental representations has never been investigated with VI users.

To address these questions, we developed BotMap, an affordable and actuated tangible tabletop interface that allows the dynamic display of tangible maps. The digital map is displayed on the screen of an interactive tabletop, and small robots are placed over the main landmarks (e.g., main cities). The robots move to their new position whenever the map is rescaled or moved. One of the user's fingers is tracked by a camera, and auditory feedback is provided when this finger passes over a robot. To investigate the effect of panning and zooming on mental representations, we developed two interfaces that allow for *discrete* vs. *continuous* panning and zooming. The Keyboard interface (*discrete* control) relies on a numeric keypad, whereas the Sliders interface (*continuous* control) relies on three tangible sliders.

A first experiment was conducted with 10 blindfolded participants to evaluate the usability of the two interfaces. A second experiment was conducted with eight VI participants to evaluate the effect of the interfaces and of panning and zooming on map comprehension. The results showed that both interfaces are usable, with the Keyboard interface having a slight advantage over the Sliders interface. More interestingly, the second study showed that the majority of VI users were able to understand the maps and make inferences about the topology of different landmarks, even though these landmarks had never been displayed simultaneously. Finally, we conducted an additional study with three blindfolded and three VI participants. We added four basic functionalities (e.g., "Go to that landmark") and assessed that BotMap enables users to independently explore a "pan & zoom" map in order to plan a journey.

## 2 RELATED WORK

### 2.1 Maps for Visually Impaired Users

Maps are essential in our everyday life, whether it is when learning an itinerary, visualizing geo-statistical data, such as temperatures, or acquiring knowledge about the geography of a country. Maps are also widely used at school. However, map accessibility for VI users is limited. Traditionally, visual maps are made accessible by creating a tactile version of the map. One of the most common methods to create a tactile map is to print the map on a microcapsule paper and then to heat it in a special oven. The printed areas rise above the surface of the paper and this tactile map can be explored with the fingers. Because these tactile maps must be prepared and printed by a tactile graphics specialist, they are rarely available outside specialized education centers. More importantly, they can only present a limited amount of information. Therefore, to enable VI people to access the same amount of spatial data as sighted users do, it is necessary to spread the map content over many tactile documents, which may raise cognitive issues [25] and/or storage issues [34]. An alternative, which we investigate in this article, is to enable users to perform panning and zooming operations on accessible interactive maps. A key benefit of interactive maps is that they can provide auditory and/or haptic feedback. Different types of accessible interactive maps exist; the classification proposed in [10] distinguishes between digital maps and hybrid maps.

*Digital maps* can be displayed on a screen or projected over a flat surface. The user can explore them by moving a cursor, and auditory feedback is provided according to what is under the cursor. The cursor can be moved indirectly using regular 2D-pointing devices, such as a keyboard, a mouse, or a joystick. 2D-pointing devices with additional feedback can also be used, such as force-feedback devices [12] or mice augmented with two Braille cells [26]. Digital maps can also be directly explored with the finger when they are displayed on a touch-screen device [53]. To make up for the absence of tactile feedback and facilitate the exploration of digital maps (and in particular touch-screen maps), a number of techniques have been proposed. For example, Kane et al. [31] designed innovative interaction techniques called *access overlays* that enable VI users to locate a landmark on a map (*edge projection),* retrieve the name of the nearest landmark (*neighborhood browsing*) or initiate guided directions to find a landmark. Other techniques involve, for example, the use of a virtual grid [1, 78] or of tactile overlays [32, 61]. The main advantage of digital maps is that they can be instantly refreshed and can even be used in mobility when displayed on a tablet. However, both the lack of physicality and the fact that only one point of the map can be explored at a time are important issues: the exploration of the map is sequential and the user must integrate both the cursor position and the corresponding auditory or vibratory feedback over time, which is cognitively demanding [34, 38, 47].

Hybrid maps rely on both a digital and a physical representation, therefore allowing VI users to physically interact with the map, which improves many cognitive processes involved in map exploration. By sweeping their hands over the map, users can quickly locate the map elements. They can also quickly relocate specific elements in the graphic [75]. In addition, they can touch several elements simultaneously, making it possible to estimate distances between them as well as their relative position [35, 44, 45, 76], for example, using "back-and-forth" movements [15, 73]. These behaviors, as well as the cognitive processes that they underlie, are particularly interesting for "pan & zoom" maps. Indeed, in the case of "pan & zoom" maps, users must frequently (re)explore different parts of the map. Three types of hybrid maps exist. *Interactive tactile maps* usually consist of a tactile overlay (e.g., a raised-line map) placed over a touch-screen device [5]. Users can explore the map with both hands but also interact with it, for example, to retrieve the name of an element or to compute distances. *Tangible maps* are composed of a set of objects that represent the map elements. For example, in the prototype of Ducasse et al. [11], users were guided by audio instructions to progressively construct and explore a map. The objects were used to make digital points (e.g., cities) and lines (e.g., borders) tangible using retractable reels. Finally, *refreshable tactile displays* consist of a matrix of pins that can be raised or lowered. The major limitation of hybrid maps concerns the updating of the physical representation. Indeed, interactive tactile maps cannot be updated without changing the overlay. Maps composed of a set of non-actuated tangible objects can be reconfigured, but the procedure is too slow to be usable. As for refreshable tactile display, they can be almost instantly updated but they are of limited size and very expensive (around $40,000 USD for a $120 \times 60$ matrix).

In conclusion, in order to implement accessible hybrid "pan & zoom" maps, it is mandatory to investigate alternative technical solutions that must be more dynamic than existing interactive tactile and tangible maps, and less expensive than refreshable tactile displays. In that sense, actuated tabletop Tangible User Interfaces (TUIs) are very promising.

### 2.2 Actuated Tabletop TUIs

Since the seminal work of Ishii and Ullmer on Tangible Bits [23], a large number of tabletop TUIs have been developed and several usage scenarios were related to the exploration of tangible maps. In their article, Ishii and Ullmer introduced the Tangible Geospace application, based on the metaDESK system, which enabled users to interact with a map of the MIT campus by manipulating tangible objects representing buildings. Urp [70] allowed sighted urban planners to simulate wind flow and sunlight by moving physical architectural models. With the MouseHous Table [22], users could simulate several arrangements of urban elements, such as streets and buildings. Maquil et al. [42] proposed the ColorTable, a tool to aid urban planners and stakeholders when discussing urban changes. The potential of TUIs for geospatial applications has also been discussed in a number of publications [28, 57].

According to Holmquist et al.'s taxonomy [20], tangible objects can be associated with any type of digital content (*containers*), can represent a piece of information (*tokens*) or can be used to manipulate the data (*tools*). In the scope of the current study, we focus on tangible objects that represent points of interest on a geographical map, and then act as tokens, and on tangible objects that can be used to interact with the map and then act as tools. When tangible objects are used as tokens, issues arise when the physical representation must be updated to reflect a change that occurred in the digital model, for example, when an element is repositioned. In fact, most tabletop TUIs rely on tangible objects that are passive and can only be manipulated by the user. Although users can move the tangible objects to their new position, it takes some time when several objects must be replaced, especially in absence of vision [11].

This issue can be tackled with actuated interfaces, defined as "*interfaces in which physical components move in a way that can be detected by the user*" [54]. Actuation can be used to update the physical representation of a TUI by altering, for example, the shape or the position of the tangible objects. Actuated tabletop TUIs have been developed for various application domains, including remote collaboration [60] and simulation [51]. Interestingly, Riedenklau et al. [59] used actuated tangibles in combination with sonification techniques to make scatterplots accessible to VI users; tangible objects moved to the center of data clusters to help users locate them.

Two main approaches exist to move an object over a surface [52, 58]: either the objects are passive and the surface itself is moving the objects, or the objects are motorized and they can move by themselves. With the first approach, the surface is usually composed of an array of electromagnets (e.g., in [50, 51]). These interfaces, referred as electromagnetic surfaces [52], can move passive objects but are expensive and complex to build. With the second approach, the objects are motorized and their wheels can be controlled either wirelessly or by displaying a particular pattern on the screen. These interfaces are usually easier and cheaper to build than magnetic surfaces, but require the tangibles to be self-powered. In both cases, tangibles differ in size, motion speed and motion abilities, and interfaces are limited by the number of tangibles that can be used simultaneously. A more detailed comparison of these two approaches can be found in [52, 58].

With the ongoing development of robotics, it becomes easier to buy or build small, affordable, and fast robots, which opens new perspectives for the design of actuated tabletop TUIs. For example, Le Goc et al. [16] recently proposed a new platform composed by many Zooids, which are small (2.6cm), affordable (50 $), and high-speed (44cm/s) custom-made robots. Possible scenarios with Zooids included drawing of Bézier curves, interactive visualizations, such as scatterplots as well as ambient displays. In this study, we used robots to display physical and dynamic maps: each robot represents a landmark and can move to a new position whenever the digital map is updated.

### 2.3 Panning and Zooming Interfaces

According to Hornbaek et al. [21], "panning changes the area of the information space that is visible, and zooming changes the scale at which the information space is viewed." The visible area of the *canvas* is often referred to as the *view* and it is displayed inside the *viewport* [27]. For panning, two conceptual models can be used [2]: users can either move the canvas directly (e.g., using "grab/touch and drag" techniques) or move the viewport over the canvas (by using navigation buttons or by moving a field-of-view box). Panning can be *continuous* (in which case the user can move the canvas or the viewport in any direction), or *constrained/discrete* (in which case the user can move the canvas or the viewport a predefined distance to a predefined set of directions). For zooming, sighted users can usually select a scale [27] by moving a slider along a vertical or horizontal axis, by pressing zoom-in or zoom-out buttons, or using the mouse wheel. Two main types of zoom exist [3, 21]: *geometric* (all elements are always displayed, whatever the scale, but the size of the icons depends on the chosen scale) or *semantic* (different elements are displayed at different scales, for example, the name of buildings or rivers only appear beyond a certain scale). Semantic zooming is more often used in online maps. As an example, OpenStreetMap provides both discrete (with on-screen buttons and key presses) and continuous (with the mouse) panning and zooming functions, and relies on semantic zooming. Users can zoom in and out in order to explore continents, countries, wide areas, or villages. In OpenStreetMap, there are 20 zoom levels, corresponding to 20 predefined scales.[1] When panning and zooming, users can experience "desert fog" if the part of the map displayed does not contain any elements, and users may feel lost or disorientated [29].

For visual panning and zooming, the effect of various factors on users' performances, satisfaction, and spatial memory have been investigated: presence or absence of overviews, comparison of input techniques, use of animation when zooming, display sizes, and so on. In addition, frameworks, toolkits, novel input techniques, and navigation aids (e.g., to avoid *desert fog*) have been proposed (see [3] for examples).

The literature on non-visual panning and zooming is much more restricted. These functionalities have mainly been implemented with refreshable tactile displays [24, 65, 77]. Input techniques included buttons [65] or drag-to-pan and pinch-to-zoom gestures [77], but were not evaluated. With Linespace [67], users can ask the system to 3-D print another part of the map currently displayed (similar to panning) or a detailed view of the area (similar to zooming), but the process is currently very slow. With iSonic [78], a tool for the exploration of geostatistical data, a numeric keypad enabled users to zoom in or out using a $3 \times 3$ recursive grid, a technique first proposed by Kamel et al. [30]. A number of zooming algorithms have also been developed to take into account the limited amount of information that can be accessed non-visually. For example, Rastogi et al. [56] developed an algorithm that determines intuitive zoom levels for the exploration of detailed diagrams, based on a tree hierarchy of the diagram elements. Palani et al. [49] proposed two types of zoom for touch-screen devices: with Fixed zoom, the elements displayed at Level 1 are displayed at Level 2 alongside new elements; with Functional zoom, different elements are displayed at different zoom levels (e.g., walls at Level 1 and corridors at Level 2). These two zooming modes were compared to a no-zoom condition. Even though the evaluation was conducted with sighted and blindfolded participants, results showed that users managed to build an accurate cognitive map regardless of the conditions.

[1] http://wiki.openstreetmap.org/wiki/Zoom_levels.

### 2.4 Summary of Related Work and Research Questions

To sum up, it appears that despite being very promising, actuated tabletop TUIs have not been used to display physical and dynamic maps accessible to VI users. A few prototypes have implemented non-visual panning and zooming but their usability has not been studied and it is unclear whether they enable VI users to efficiently navigate and understand the maps. Although the work of Palani et al. [49] addressed similar questions, their study was limited to simple and non-physical maps and was conducted with blindfolded participants only. In this work, we tackled the issue of non-visual panning and zooming by addressing the following research questions: (RQ1) how to design usable non-visual interaction techniques for panning and zooming on an actuated tabletop TUI? (RQ2) Which panning and zooming modes (*discrete* vs. *continuous*) should be used for VI users? (RQ3) Can VI users understand large tangible maps whose exploration requires panning and zooming? (RQ4) Is comprehension influenced by the interaction technique? To answer these questions, we designed and evaluated BotMap, an actuated tabletop TUI that allows for the display of dynamic maps, and supports panning and zooming.

## 3 BOTMAP: SYSTEM OVERVIEW AND INTERFACES

The design of BotMap was based on an iterative user-centered process. We included one VI person and six blindfolded participants (including three HCI experts and one ergonomist) in many brainstorming and pilot sessions. During the brainstorming sessions, we addressed many questions related to the type of viewport, the design of interactions (whether keyboard or slider based), the number of zoom levels, the number of elements in the map, and so on. We aimed to design a system that was answering user needs and that was based on their skills (e.g., moving viewport instead of moving map, see Section 3.3.1). During the pilot test sessions, we evaluated the usability of individual components, such as sliders behavior or zooming function. In this section, we describe the main aspects of BotMap as well as how panning and zooming can be performed using the Keyboard and the Sliders interfaces.

### 3.1 System Overview

In BotMap, landmarks are represented by robots that are tracked by a camera placed above an interactive table and that can move freely on the surface whenever the map is updated (Figure 1(A)). Users can explore the map using both hands and they can interact with the robots in order to retrieve the names of the corresponding landmarks (Figure 1(B)). They can also ask for other pieces of information using voice commands. When panning and zooming, feedback concerning the viewport position and the scale (i.e., how many kilometers the viewport represents) is provided. When the user confirms the ongoing operation, the robots move to their new position.

Maps are composed of several landmarks (in the current version of BotMap, lines and areas are not rendered). Similar to Google Maps or OpenStreetMaps, we defined different zoom levels, and each landmark appears at a certain level depending on its type (semantic zooming). For example, for the first two studies, we used the following zoom levels: city, town, and village. At the City level, the whole map can be displayed within the viewport but only the cities are displayed; at the Town level, the cities and the towns of a selected area are displayed; at the Village level, the cities, towns, and villages of a smaller area are displayed. Users can switch between these levels by zooming in or out.

### 3.2 Identification of Landmarks

Users can retrieve the names of the landmarks (Figure 1(B)). A marker is attached to the index finger and is tracked by the camera placed above the tabletop. To retrieve the name of a landmark, users must place the index on top of the robot. The name is given via Text-To-Speech (TTS),

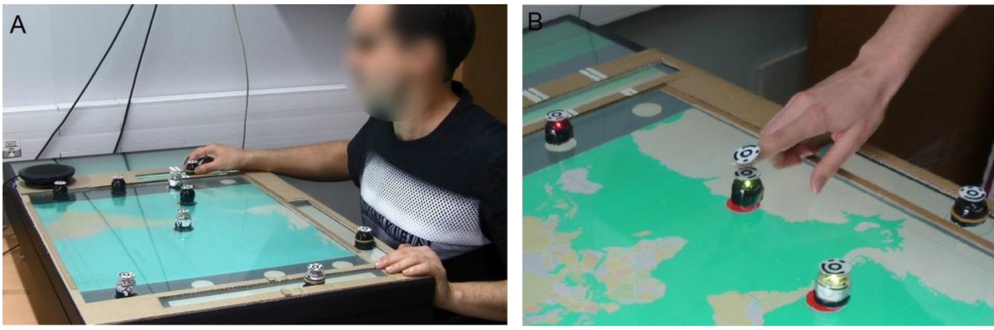


Fig. 1. (A) BotMap is composed of an interactive tabletop and several robots, tracked by a camera. Each robot represents a landmark and moves to its new position whenever the map is refreshed. On the picture, a blind user is zooming with the Sliders interface. (B) Users can select a robot to retrieve the name of the corresponding landmark.

followed by the type of the landmark (town, city, or village). When there are more robots than landmarks to be displayed, robots move to parking areas located on each side of the viewport (three on the left, three on the right). If the user selects a robot that is on a parking space, the message "parking" is played.

## 3.3 Interaction Techniques for Panning and Zooming

In this section, we first describe the design rationale of BotMap in terms of input modalities and panning and zooming implementations, and then describe the two interfaces.

### *3.3.1 Design Rationale.*

*3.3.1.1 Input Modalities.* The design space of interaction techniques above or around an interactive table is large and mainly includes touch interaction, tangible interaction, mid-air interaction, or the use of regular devices, such as a keyboard/mouse or voice commands. Throughout our iterative design process, we found that touch-based techniques were not suitable for panning and zooming, as unintentional inputs were often triggered, making it difficult to correctly perform a gesture. Mid-air interaction techniques have been successfully used to enable VI users to interact with a map [1]. However, the detection of gestures is usually performed with a motion-capture system and therefore requires additional and expensive hardware. In addition, in [1], users reported that mid-air gestures induced fatigue. On the other hand, keyboards and tangible objects are cheap and provide tactile and/or kinesthetic feedback. In particular, keyboards are the most common input devices for VI users to interact with computers. Keyboard and tangible objects have been successfully used in various prototypes developed for VI users, notably to navigate a map (e.g., keyboard in iSonic [78]) or a graph (e.g., a tangible slider in the Tangible Graph Builder [43]). Building on these successful prototypes, we used a numeric keyboard for *discrete* panning and zooming (Keyboard interface) and tangible sliders for *continuous* panning and zooming (Sliders interface). Users can also interact with the system using voice commands.

*3.3.1.2 Implementations of Panning and Zooming.* In this section, we briefly explain the decisions concerning the choice of particular implementations for panning and zooming:

- *Constrained vs. continuous panning and zooming.* Two models exist for panning and zooming [2]: discrete vs. continuous. In order to evaluate whether one model could be more usable than the other, we designed two interfaces: the Keyboard interface allows constrained or discrete panning and zooming, while the Sliders interface allows continuous panning and zooming.

| | Activation | Action | Confirmation |
|---|---|---|---|
| Input | "Pan" or "Zoom" | Keyboard: Pressing the keys<br>Sliders: Moving the sliders | "OK" |
| Output | "Pan activated" or "Zoom activated" | Pan : "3 h, 30 km"<br>Zoom : "Town Level, 100km" | List of landmarks + robots updated |

Fig. 2. Before panning and zooming, users must activate the corresponding mode. They must confirm (or cancel) the action. Audio feedback is the same for the two interfaces.

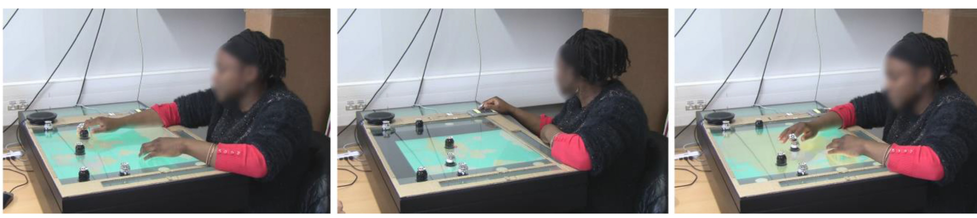

Fig. 3. Zooming with the keyboard interface: A participant selects a landmark in order to center it (left), then zooms in by pressing a key (middle), and finally explores the map once the robots are in place (right).

- *Clutch-free panning and zooming.* During the design process, various techniques were implemented and evaluated. Based on these preliminary tests, we decided not to implement any interaction technique that would require "clutching" operations (when users must lift their finger or the object they are manipulating and reposition it [46]). These operations are cognitively demanding because they disrupt the panning and zooming processes. We therefore discarded "drag-to-pan" and "pinch-to-zoom" interaction techniques, as they might require clutching.
- *Moving the map vs. moving the viewport.* To our knowledge, the question of which panning implementation should be used for VI users has never been addressed. However, VI users do extensively use panning when using a screen reader or a refreshable Braille display. On these devices, pressing "down" results in listening or displaying the line below the current line and is therefore similar to moving down the viewport. According to this observation, during the design process, we decided to implement panning that requires users to move the viewport instead of the map.

*3.3.2 Description of the Keyboard and Slider Interfaces.* The voice commands and feedback for panning and zooming are the same for the two interfaces (Figures 2, 3, and 4). However, the input techniques differ between the two interfaces (Figures 5 and 6). In this section, we first describe the voice commands and feedback for activation and confirmation, and then describe the input techniques.

*3.3.2.1 Activation/Confirmation.* Users can activate the pan or zoom modes using the voice commands "Pan" and "Zoom." The message "pan activated" or "zoom activated" is then played. When panning, feedback concerning the viewport's position is given, with respect to its initial position (i.e., at the time of activation). Direction is given using an analogy of the 12-hour clock[2]

[2]Providing directions with the clock-metaphor is commonly used in specialized education centers for visually impaired users and this method is particularly appreciated by visually impaired people [19]. In addition, it has been successfully used in several prototypes [11], as it is an effective way to provide precise directions [62].

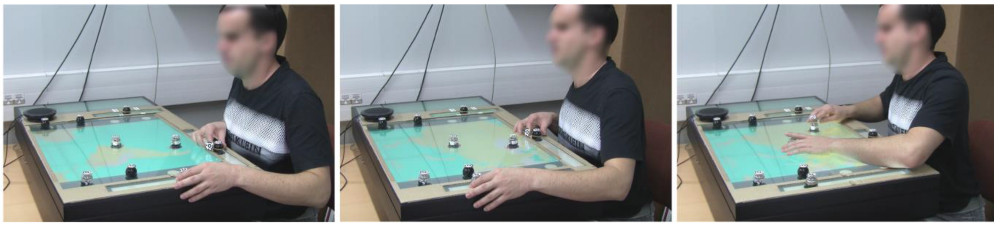

Fig. 4. Panning with the sliders interface: A participant moves the two sliders for panning (left and middle) and explores the map once the robots are in place (right).

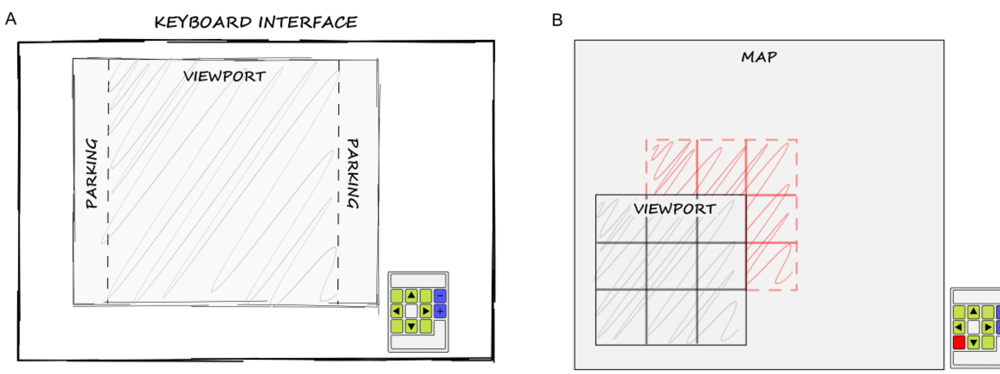


Fig. 5. (A) Keyboard interface. For panning, users can move the viewport in any of eight directions by pressing the corresponding key (in green). For zooming (in blue), they must press the "plus" key (zoom in) or the "minus" key. Only three predefined scales can be selected (one per zoom level). Unused robots move to the parking areas. (B) Illustration of the viewport with its nine virtual cells, as well as two successive viewport's positions over the map (before and after the red key has been pressed).

(3 hours means that the viewport has been moved to the right); distance is given in kilometers. The frequency at which feedback is given depends on the interface and is described in the following sections. When zooming in or out, feedback concerning the zoom level and scale is provided (e.g., "Cities level. The viewport represents 100 km.").

When the user confirms the current operation (voice command "ok"), the names of the landmarks that have disappeared and appeared are given (e.g., "Previous towns: *Bordeaux*. New city: *London*. New town: *Bristol*."). At the same time, the robots move to their new position. When the robots are correctly positioned, the message "ok" is played. Users may also cancel the ongoing operation.

*3.3.2.2 Keyboard Interface.* For this interface, the viewport is divided into a grid of 3 × 3 square cells (Figure 5). A numeric keyboard is placed on the right side of the viewport. Tactile cues were added to the keyboard to help users find the central key.

For panning, users must press one of the eight direction arrows in order to move the viewport one cell in the corresponding direction. Whenever a key is pressed, feedback concerning the distance and direction of the viewport are provided with respect to its initial position (i.e., at the time of activation). For example, if the user presses the right key at the Town level, the message "3 hours, 30 kilometers" is played; if the user then presses the right key another time, the message "3 hours, 60 kilometers" is played.

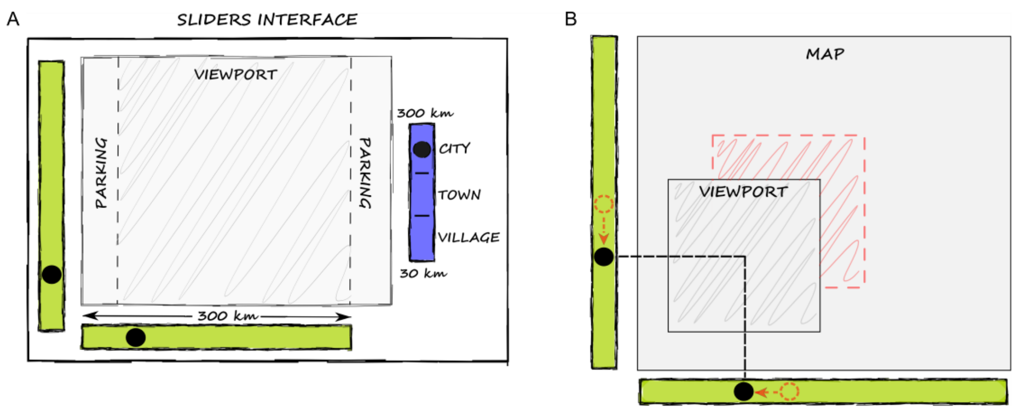


Fig. 6. (A) Sliders interface. For panning, users can move the viewport horizontally and vertically by moving two robots (black dots on the green areas) that represent the position of the viewport's center. Users can move the third robot in the zooming area (in blue) to select any scale between 30km and 300km. (B) Illustration of two successive sliders' and viewport's positions over the map.

For zooming, users have to press the "plus" or "minus" keys. The "plus" key allows users to zoom in: the part of the map displayed in the central cell expands to fill in the entire viewport. The "minus" key allows users to zoom out: the part of the map displayed within the entire viewport shrinks to fill only the central cell. Whenever the "plus" or "minus" key is pressed feedback concerning the current zoom level is provided. The predefined scales are 300km (City zoom level), 100km (Town zoom level), and 30km (Village zoom level). If the user cannot move the viewport (because one of the edges of the map has been reached) or cannot zoom in or out, feedback is provided accordingly (e.g., "Impossible to go to the right," and "Impossible to zoom in").

The Keyboard interface only provides the *relative* position of the viewport with respect to its initial position (i.e., at time of activation). Users can infer the absolute position of the viewport on the map but cannot retrieve it directly from the interface.

*3.3.2.3 Sliders Interface.* For this interface, three additional robots (referred to as sliders) can be moved by the user inside three rectangular areas delimited by cardboard (Figure 6). Two tactile cues (rubber-bands) were added on each slider to identify them.

The vertical and horizontal areas, respectively, to the left of and below the viewport, are used for panning. Users can move the viewport by moving the sliders. The position of the sliders within these areas corresponds to the vertical/horizontal position of the viewport over the map. When panning, feedback concerning the current position of the viewport is provided with respect to its initial one every 2 seconds (the delay of 2 seconds between messages has been chosen during the design sessions, after several iterations). For example, if the user moves the slider in the horizontal area to the right, the messages "3 hours, 50km," "3 hours, 70km," "3 hours, 100km" are successively played. Feedback is interrupted when the user validates or cancels the current operation.

For zooming, a vertical slider area is placed to the right of the viewport. The position of the slider placed within this area represents the current scale of the map, in kilometers. The highest position of the slider corresponds to a viewport representing 300km × 300km, while the lowest position corresponds to a viewport of 30km × 30km. Any scale between 30 and 300km can be selected. Feedback concerning the current zoom level and scale is provided every second or as soon as a new zoom level is selected. Tactile cues were placed on both sides of this slider area to indicate the

Table 1. Summary of Voice Commands

| Context | Command | Goal |
|---|---|---|
| **Exploration** | List | List of landmarks |
| | **Scale** | Current zoom level / current length (in km) of one side of the map |
| | Repeat | Repeat the last message |
| **Manipulation** | Center | Center the last selected landmark |
| | Pan | Activate “pan” mode |
| | Zoom | Activate “zoom” mode |
| | Ok | Validate the current panning/zooming operation |
| | Cancel | Cancel the current panning/zooming operation |

limit between the three zoom levels (City level: between 300km and 150km; Town level: between 150km and 80km; Village level: between 80km and 30km).

By touching the Sliders, users can retrieve the absolute position of the viewport on the map (see Figure 6(B)), as well as the current zoom level (City level in Figure 6(A)) and the approximate scale (around 200km in Figure 6(A)).

## 3.4 Additional Commands

*3.4.1 Voice Commands during Exploration.* Three voice commands allow users to retrieve additional pieces of information:

— “List”: The system lists all the landmarks currently displayed, according to their type (e.g., “Cities: London. Towns: Cambridge”). The message “no landmarks” is given when there are no landmarks displayed.
— “Scale”: The system indicates the zoom level as well as how many kilometers the length of the viewport represents (e.g., “Towns Level. The window represents 100km”).
— “Repeat”: The system replays the last message.

*3.4.2 Centering on a Landmark for Zooming In and Out.* To display more details around a landmark, it is necessary to place this landmark at the center of the viewport before zooming in, otherwise it may move out of the viewport when zooming in. We therefore implemented a “centering” feature that enables users to place any landmark currently displayed at the center of the viewport. Users first need to select the landmark and then use the voice command “Center.” The map is then updated accordingly.

During pilot tests, we observed that when zooming out, participants were sometimes disorientated because a landmark that was displayed at Village level could not be displayed at Town/City level. This was particularly an issue when users were asked to find the nearest town/city to a village: when zooming out, the village would disappear. In order to enable users to have a fixed reference between different zoom levels, we modified the zooming algorithm. When the user centers the viewport on a landmark before zooming out, this landmark is preserved and displayed at inferior zoom levels until the user centers the viewport on another landmark (it is therefore possible to display one village at Town and City levels and one town at City level).

## 3.5 Summary

Table 1 summarizes the voice commands available on BotMap.

Fig. 7. Photographs illustrating how the robots move to reach their position by following lines displayed on the underlying screen. Arrows indicate the trajectories computed for each robot.

## 4 IMPLEMENTATION

The table was a MultiTaction interactive table (MT420S, MultiTouch Ltd., Helsinki, Finland), running Windows 7. The display area was 93cm × 52cm for a diagonal of 42in. The different slider areas were surrounded by thick cardboard (0.5cm) and bordered the viewport (see Figure 1(A)). Two laminated strips were placed at the top and bottom of the viewport to help users distinguish between the viewport and the parking spaces.

Although the table can track fiducial markers, these markers were too large for our application (4cm × 4cm). We therefore used an additional webcam placed above the tabletop to track the robots upon which we attached a small circular marker (2.5cm diameter). The user's index finger was identified with a marker and tracked by the same camera. A USB speaker and a numeric keyboard (for the Keyboard interface only) were connected to the table.

For the robots, Ozobots Bits (Ozobot & Evollve, Inc.) were used, which are small and light toy robots (2.5cm diameter × 2.5cm high, 9 grams). The robots have two wheels and can move at speeds up to 44mm/s. Ozobots Bits are equipped with a color sensor, and their behavior can be programmed in advance, using the OzoBlockly Editor that is based on the Blockly graphical programming language. Possible behaviors include following a colored line, rotating and changing the LED color. The autonomy of the Ozobots Bits is around 1 hour when moving. Three additional Ozobots Bits were used as sliders for the Sliders interface, so that the sliders can be repositioned if the user cancels a panning or zooming operation. Each robot costs 50€. For the purpose of our studies, we used nine robots, yielding to a total of 450€ (although we used a Multitaction table, a regular 42in screen can be used to reduce the overall cost of the system).

Audio instructions were provided with a SAPI4 compliant TTS engine distributed as part of the CloudGarden TalkingJava SDK 1.7.0. To avoid any issue with voice recognition, vocal commands were triggered by the evaluator using a keyboard.

The BotMap application was developed using the MultiTouch4Java library (MT4J, [37]). Besides receiving TUIO messages sent by the interactive table, the library provides basic methods for panning and zooming. However, the library requires the use of tiled web maps. To work offline, we generated offline map tiles using TileMill, an open source map design studio. The application also managed the robots.

Robots' markers were tracked using the TopCode library, which provides their position and orientation. To improve the precision of the detection, the markers' coordinates were refined using a homography. Whenever the map is rescaled or the viewport repositioned, each robot is assigned a landmark or a parking space, depending on the number of landmarks to be displayed. A Java application controls the robots in order to avoid collisions and lock-ups by displaying lines and circles of different colors that make the robots move, pause or rotate (Figure 7). In the current version of BotMap, six robots can be used simultaneously, in addition to the three sliders for the Slider Interface. All robots move simultaneously, and the average time required for all the robots to be correctly positioned was 9 seconds. In addition, the landmarks must not be too close to each other to ensure the reliability of the algorithm (we found that a distance of approximately 6cm between two landmarks was appropriate), and the number of landmarks to be displayed must not exceed the number of available robots. For the user studies, we designed maps that fulfilled these criteria.

## 5 STUDY 1: COMPARISON OF THE USABILITY OF THE TWO INTERFACES

The aim of this study was to evaluate whether the two interfaces enabled users to perform panning and zooming operations of various complexities (panning various distances and directions, with or without zooming). We also aimed to investigate whether one interaction technique was more usable than the other. To do so, we used a basic target reaching task that did not require high cognitive processes but rather basic interaction processes (pressing the keys or moving the sliders). As the aim of this study was to assess usability and not mental representations, we did not expect any significant differences between blindfolded and VI participants.

### 5.1 Material and Methods

*5.1.1 Participants.* In this first study, the purpose was to assess that the two interfaces were usable to perform panning and zooming operations without vision. Recruiting VI participants is difficult and time consuming; therefore, conducting a first study with sighted users is an accepted procedure [43]. It is a way to assess basic features of a system before conducting a more comprehensive study where differences between VI and sighted participants could be expected. We recruited 10 sighted participants (2 females and 8 males) from the research laboratory. Participants were aged between 24 and 29 years ($M = 26.4$, $SD = 1.6$) and were all right-handed.

*5.1.2 Material.* We used the setup described in the previous section with six robots (in addition to the three sliders for the Sliders interface). In order to decrease the duration of the experiment, the robots were not moving alone but were manually placed at their correct position after each panning/zooming command. Two randomly generated maps were used: 8 landmarks were selected in the first map for the training, and 24 landmarks were selected in the second map for the test. Landmark names were extracted from a list of existing village names and were randomly assigned at the beginning of each trial.

*5.1.3 Task.* The task was to find and select a landmark called "Target" as quickly as possible. At the beginning of each trial, a message was played indicating the current zoom level, the zoom level at which the target was present and its direction and distance with respect to the center of the viewport (e.g., "City level. Target village located at 3 hours and 150 kilometers"). At any time (i.e., even when panning), users could ask the system to give the updated direction and distance with the voice command "Info." Distance and direction were given with respect to the current position of the viewport.

*5.1.4 Experimental Design and Conditions.* The experiment used a within-subject design with four independent variables and two factors for each independent variable. Therefore, there were $2 \times 2 \times 2 \times 2 = 16$ conditions, i.e., 8 for each interface:

- *Interface*. The Keyboard and the Sliders interfaces were compared.
- *Direction*. Targets were either located next to the vertical or horizontal axes (*Vertical/Horizontal*), i.e., North, South, West, East, or next to the diagonal axes (*Diagonal*), i.e., NE, NW, SE, SW. We ensured that all targets were within an angle of 15° around the axis.
- *Distance*. Targets were either located within 40–50km from the current position (*Small*) or within 120–130km from the current position (*Large*).

Table 2. Procedure for Each Session of Study 1

| **Procedure – Study 1** | |
|---|---|
| Introduction | Explanations concerning the experiment's goal and organization. Consent form + photo/video authorization form |
| Basic features | Selection, voice commands (*List, Scale, and Repeat), centering* |
| Panning | Explanations using a tactile map and frame + 4 training trials with BotMap |
| Zooming | Explanations using a tactile map and frame + 4 training trials with BotMap |
| Test | Three trials per condition (24 trials). The same set of 24 trials was used for both sessions and for all participants, but trials were presented in a random order and names changed |
| Questionnaires | SUS and ranking questionnaires |

There was one session for each interface.

— *Zoom level.* At the beginning of the trial, the zoom level of the map was either City or Village but all the targets were villages. Therefore, the initial zoom level and the target zoom level were either identical (*Identical*) or different (*Different*), in which case users had to zoom in in order to display the target.

*5.1.5 Variables.* To assess the usability of the interfaces, we measured the time required to display and select the target. For each trial, we also logged the successive positions of the viewport and scales. Finally, participants had to fill out the System Usability Scale (SUS) questionnaire [6], and to indicate which interface was the more efficient, which one was the more pleasant, and which one they would choose if they had only one choice, and for what reasons. All sessions were video recorded.

*5.1.6 Data Analysis.* In accordance with the recent recommendations from the APA organization [14], we report effect sizes with 95% confidence intervals (CI) instead of p-Value statistics based on the Null Hypothesis Significance Testing paradigm, the use of which has been criticized (see [7, 9] for instance).

For completion times, we first log-transformed all completion times to correct for right skewness. We then computed means and 95% exact CIs on the transformed data and report the results anti-logged, i.e., we report geometric means instead of arithmetic means, as recommended in [63]. For pairwise comparisons, we first computed differences between means on log-transformed data (Sliders minus Keyboard), for each participant and each factor. We then computed means and 95% exact CIs and report the results anti-logged: differences between mean completion times are therefore expressed as ratios [9, 13].

For distances, we computed means and 95% bootstrap CIs. For pairwise comparisons, we first computed ratios between interfaces (Sliders/Keyboard), for each participant and each factor.[3] We then computed means and 95% bootstrap CIs.

*5.1.7 Procedure.* The experiment was composed of two sessions (one per interface), conducted on separate days but during the same week. Interface order was counterbalanced. Each session lasted approximately 1 hour and 30minutes. The procedure is described in Table 2.

[3]As the distance that users had to pan depended on the trials (*Small vs. Large*), computing mean differences in distances across different types of trials would not have been relevant, and we therefore computed ratios.

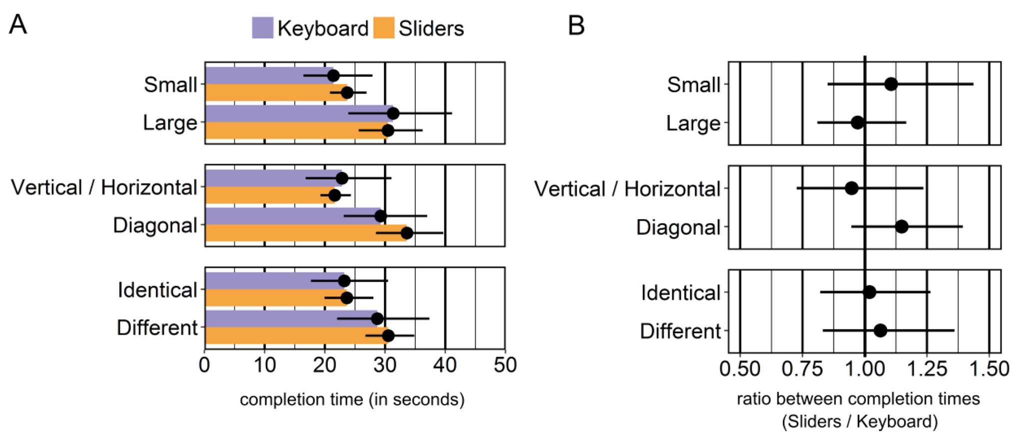


Fig. 8. (A) Mean completion times per condition, in seconds ($N = 8$). (B) Mean ratios between completion times (Sliders/Keyboard, $N = 8$). Values superior to 1 indicate larger times for the Sliders than for the Keyboard. Error bars show 95% exact CIs.

## 5.2 Results

*5.2.1 Completion Times.* Because different strategies were used to select the landmark once it was displayed (such as asking for the system to provide distance and direction with respect to the viewport's center), we only report times required to display the target. As a reminder, we report geometric means, and differences between completion times (Sliders minus Keyboard) are expressed as ratios. Values superior to 1 indicate that participants took longer with the Sliders than with the Keyboard.

All participants managed to display and select the target within the allowed time (3 minutes). Times were similar between the two interfaces (Keyboard: *mean completion time* = 25.8 seconds, 95% CI [19.8, 33.5]; Sliders: *mean completion time* = 26.8 seconds, 95% CI [23.3, 30.7]). However, with the Sliders, participants took consistently longer to perform *Diagonal* than *Vertical/Horizontal* trials (the two orange bars in the central row of Figure 8(A)).

The mean ratio between completion times, computed across all conditions, was 1.0, 95% CI [0.8, 1.2], meaning that overall, participants took as much time to perform the task with the Sliders as with the Keyboard. Figure 8(B) shows pairwise comparisons between the two interfaces for each factor: it shows that participants tended to take longer with the Sliders than with the Keyboard for *Small* and *Diagonal* trials.

*5.2.2 Distances.* We computed the distance (in kilometers) panned by the participants to reach the target in each condition. Results show that participants panned larger distances for *Diagonal* than for *Vertical/Horizontal* trials with the Sliders (the two orange bars in the central row of Figure 9(A)). With the Keyboard, participants tended to pan smaller distance when the zoom level was *Identical* than when it was *Different* (the two purple bars in the lower row of Figure 9(A)). This suggests that when the zoom level was *Different*, participants performed unnecessary panning actions at the City level.

The mean ratio between distances panned (Sliders/Keyboard), computed across all conditions, was 1.3 (95% CI [1.2, 1.6]). Figure 6(B) shows pairwise comparisons for each factor. Given the fact that interval endpoints are about seven times less plausible than the point estimate [7], participants tended to pan larger distances with the Sliders in all conditions, and in particular for trials where

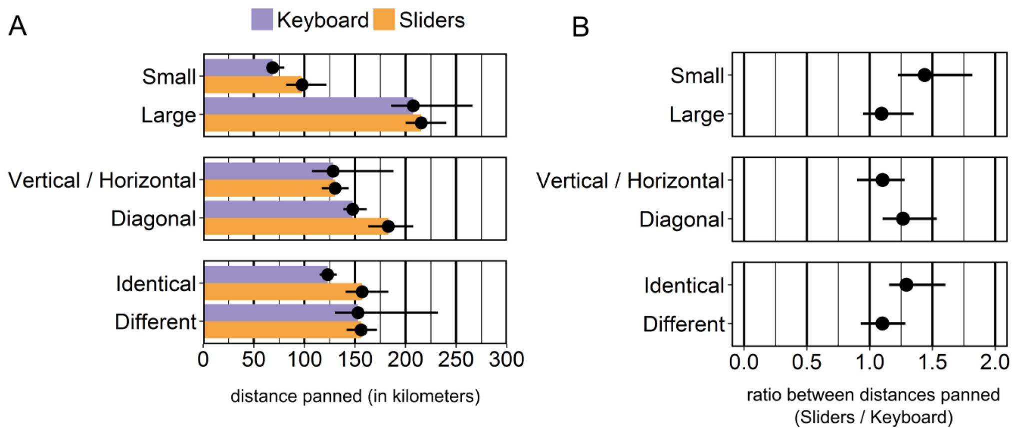


Fig. 9. (A) Mean distance panned (in kilometers) per condition ($N = 8$). (B) Mean ratio between distances panned (Sliders/Keyboard, $N = 8$). Values superior to 1 indicate that participants panned larger distances with the Sliders than with the Keyboard. Error bars show 95% bootstrap CIs.

the distance was *Small* (*ratio* = 1.4, 95% CI [1.2, 1.8]), the direction *Diagonal* (*ratio* = 1.3, 95% CI [1.1, 1.5]), and the zoom level *Identical* (*ratio* = 1.3, 95% CI [1.2, 1.5]).

*5.2.3 Navigation Strategies.* In this section, we report different strategies that we identified, based on log analysis. As a reminder, three zoom levels were defined (Village, Town, and City). With the Sliders, users could select any scale between 30km and 300km (Village level: between 30km and 80km; Town level: between 80km and 150km; City level: between 150km and 300km). With the Keyboard, there was only one predefined scale per zoom level (Village level: 30km; Town level: 100km; City level: 300km).

With the Sliders, two participants (P5 and P10) systematically moved to 80km (the limit between Village and City level), regardless of the initial zoom level. Other participants (P1, P2, P3, P4, P6, and P9) used a similar strategy, but only when the initial zoom level was City. In this condition, the other two participants (P7 and P8) systematically moved to 30km (the lowest position of the slider).

With the Keyboard, when the initial zoom level was City level, four participants (P1, P3, P5, and P7) almost systematically zoomed in to the Village level before panning, regardless of the distance to the target. In contrast, other participants tended to zoom out to the City and/or Town levels, then pan, then zoom in to the Village level.

We also observed that the final position of the target within the viewport varied between the two interfaces: with the Sliders, 45.8% of the targets were located in the center of the viewport at the end of the trials, vs. 26.2% for the Keyboard.

*5.2.4 Satisfaction and Feedback.* The overall score for the SUS questionnaire was 85.5 (out of 100) for the Keyboard (95% CI [77.0, 91.2]) and 86.0 for the Sliders (95% CI [80.7, 90.7]). Four participants out of ten found the Keyboard interface more efficient than the Sliders, and five found it more enjoyable than the Sliders. Overall, four participants preferred the Keyboard and six the Sliders.

Concerning the Sliders interface, three participants appreciated the fact that they could pan large distances without having to zoom out (P2, P7, and P9) and several participants mentioned the usefulness of the tactile cues in the zooming areas (P5, P6, and P8). Participants preferred the

Sliders because they were easier to use (P6 and P10), more precise (P7) and intuitive (P8), but also because they could be used to move the viewport in any direction (P2 and P6) and that, compared to the Keyboard, they were based on large movements with both hands (P8 and P9). However, several participants reported that practice and time were necessary in order to learn how to use both sliders simultaneously (P5) or to correctly position the sliders (P5, P9, and P6).

Concerning the Keyboard interface, half of the participants found that it was sometimes difficult to distinguish between the directional keys (for panning) and the plus and minus keys (for zooming) and suggested that the two sets of keys should be more clearly separated. They also reported that the Keyboard interface enabled them to move the viewport to a certain distance by "counting" the number of key presses. This feature was the main reason why the Keyboard was found more efficient than the Sliders (P1, P3, P4, and P5) or less efficient (P10) and less enjoyable (P6 and P7).

### 5.3 Discussion and Summary of Study 1

The results show that all participants managed to successfully perform the task with the two interfaces, each in less than 30 seconds. Completion times indicate that overall, participants took as much time with the Keyboard as with the Sliders interface. However, participants tended to pan larger distances with the Sliders than with the Keyboard, especially for *Small, Diagonal,* and *Identical* trials. This is in line with the fact that, with the Sliders, participants took longer and panned larger distances to perform *Diagonal* than *Vertical/Horizontal* trials, whereas this difference was not as clear with the Keyboard. These observations suggest that the Sliders are less easy to control than the Keyboard when the viewport must be moved diagonally or for small distances. As suggested by some participants, additional practice with the Sliders may improve performance.

There were no clear preferences in subjective ranking for one interface or the other. The fact that the Keyboard interface allowed for *discrete* panning was seen by some as an advantage (users counted how many times they had to press each key to reach the target) and by others as a disadvantage (it was not always possible to place the target at the center of the viewport). It is interesting to note that using BotMap, participants managed to develop strategies similar to those that they might have developed using a visual interface, such as zooming out to pan larger distances.

To sum up, this first study showed that the designed input techniques and feedback allowed users to perform panning and zooming actions of various complexities without vision. Both interfaces led to similar performances, with the Keyboard having a slight advantage over the Sliders in terms of efficiency (smaller distances panned). However, blindfolded participants were already familiar with the concepts of panning and zooming, which might have helped them determine how to use the interfaces. In the following study, we specifically investigated whether VI users were also able to efficiently navigate maps through panning and zooming and, above all, to understand them.

## 6 STUDY 2: USABILITY AND MENTAL REPRESENTATIONS WITH VISUALLY IMPAIRED USERS

The main aim of this study was to evaluate whether BotMap enables VI users to understand a map whose exploration requires panning and zooming: users had to explore a map with several landmarks and memorize the overall configuration in order to answer comprehension questions and eventually reconstruct the map. We also wanted to investigate whether one interface would be more usable than the other for VI users.

### 6.1 Material and Methods

*6.1.1 Rationale.* In this study, users had to explore a map with 10 landmarks and memorize the overall configuration in order to answer comprehension questions and eventually reconstruct the

Table 3. Participants' Main Characteristics

| | Gender | Age | Age at onset of blindness | Current activity | Usage of maps during school | Self-reported spatial ability |
|---|---|---|---|---|---|---|
| 1 | Male | 66 | 5 | Retired | Occasionally | Excellent |
| 2 | Female | 44 | 0 | Unemployed | Very often | Average |
| 3 | Female | 63 | 0 | Civil servant | Occasionally | Excellent |
| 4 | Male | 44 | 0 | Software developer | Occasionally | Excellent |
| 5 | Male | 35 | 4 | Unemployed | Very rarely | Excellent |
| 6 | Male | 60 | 6 months | Retired | Occasionally | Excellent |
| 7 | Female | 33 | 12 | Teacher | Occasionally | Neutral |
| 8 | Male | 26 | 16 | Unemployed | Very often | Excellent |

map. Besides providing distinct measures related to map navigation (completion time and distance panned for each trial) and map comprehension (percentage of correct answers and accuracy of the reconstructed map), such a task mimics a realistic scenario where users would first discover and learn spatial relations between items before recalling them [55]. A similar procedure was used in a number of studies dealing with sighted user's spatial memory in panning and zooming interfaces [27, 55].

*6.1.2 Participants.* We recruited eight legally blind people (five males and three females), aged between 26 and 66 years ($M$ = 46.4, $SD$ = 15.0). P2, P3, and P8 had residual perception of very bright light but could not rely on it to distinguish shapes. All participants possess a smartphone and a laptop, and use them very frequently. All participants use a screen reader. All participants also reported that, when at school, they had used maps, but at various frequencies. The following table sums up participants' main characteristics (Table 3).

We asked participants whether they knew how online maps (e.g., Google Maps) work as well as whether they were familiar with the concepts of panning and zooming. Apart from P8 who lost sight at 16, participants knew very little about Google Maps. They mainly knew that Google Maps enables users to compute an itinerary between two points of interest and that this itinerary can be displayed on the map. Other reported some knowledge about Google Street View. For example, P5 indicated that "one can enlarge the images to have more details and we can even see streets and cars." As for the concept of zooming, most participants said that it was used to enlarge a picture. None were familiar with the word "panning."

*6.1.3 Material.* We used the same apparatus as for Study 1. Four different maps were used for the Training session and the Evaluation session. During training, names of landmarks were either names of planets, musical instruments, or vegetables. In that way, we ensured that the maps used for training did not interfere with those used for the test. For the evaluation, two maps were used; the second map was symmetric to the first one. Both maps were fictitious in order to ensure that participants had no prior knowledge regarding these maps. Both maps were composed of ten landmarks: three cities, five towns, and two villages (see Figure 10 for an example). In order to help the users memorize the three cities and their configuration, we used the names of well-known cities and respected their relative locations (for example, in Map 2 the western city was labeled Madrid and the eastern was labeled Zurich). The names of the towns and villages were randomly chosen from a list of municipalities so that each landmark began with a different letter. Different names were used for the two maps.

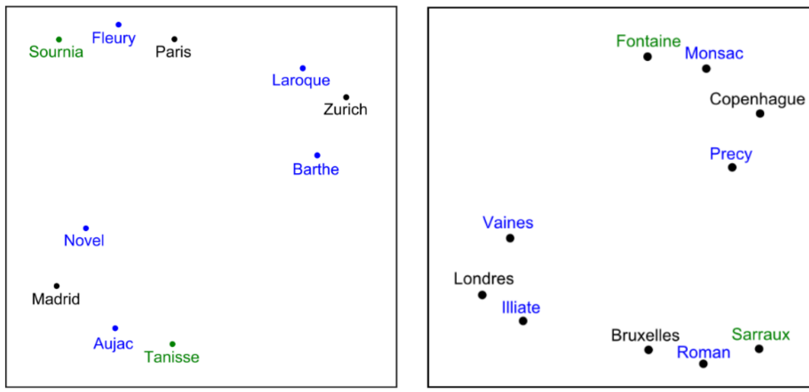


Fig. 10. The two maps used for the test. The three cities are written in black, the five towns in blue, and the two villages in green (colors for illustration purpose only).

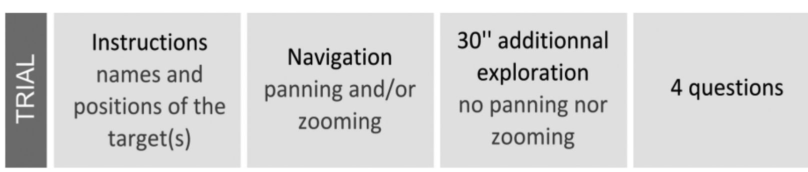


Fig. 11. Summary of a trial.

*6.1.4 Task.* At the very beginning of the test, users were asked to explore the map at the City level and to memorize the locations and names of the three cities. The task was composed of eight trials. In each trial, users had to find one or two landmarks as quickly as possible and were asked to understand and memorize their locations. In order to ensure that participants would use both panning and zooming, the order of the landmarks to be found was controlled and panning or zooming were either not allowed or mandatory, depending on the trials being performed. In addition, to keep the length of the experiment within 2.5 hours, while taking into account the fact that haptic exploration takes time, two landmarks were explored only once and six landmarks were explored twice. In order to compensate for the lack of repetitions, we asked comprehension questions immediately after each trial. The overall reconstruction of the map was made at the end of the session.

To help the subjects locate a target, they were told its type (Town or Village) as well as its approximate position with respect to a reference landmark (given as "R" in the upcoming text), which was a Town or a City already explored. Users had to pan and/or zoom to find the target(s). When they found and selected a target with their finger, its name was given, followed by its type and the message "found." At the end of each trial, users were given 30 additional seconds to explore the current view (without panning, zooming, or centering). Then, they had to answer four comprehension questions. At the beginning of the following trial, the viewport was repositioned so that all users started a given trial with the same configuration. Figure 11 summarizes how each trial was conducted.

*6.1.5 Experimental Design and Conditions.* The experiment relied on a within-subject design with two independent variables: Interface (Keyboard and Sliders) and Actions needed in each trial (*Zoom in*; *Pan*; *Zoom in & Pan*; and *Zoom out & in*), i.e., $2 \times 4 = 8$ conditions. Participants performed two trials for each condition. The trials were presented in the same order for each participant. In that way, participants were "guided" to progressively explore the whole map. For each trial, users had to perform a different set of actions:

— *Zoom in:* The initial zoom level was City level. Users had to zoom in to find two towns located within 60km of the reference city R. It was possible to simultaneously display the city R and the targets if the viewport was centered on R at the City level.
— *Pan:* The initial zoom level was Town level. Users had to pan in order to find one town located beyond 160km from the reference town R, within a range of 3 hours (e.g., 6–9 hours).
— *Zoom in & Pan:* The initial zoom level was Town level. Users had to find one village located within 60km of town R, on its left or right side. Users had to zoom in (in order to display villages) and then pan (the target and R could not be displayed simultaneously).
— *Zoom out & in:* The initial zoom level was Town level. Users had to find two towns located within 60km of city R. Therefore, users had to zoom out, pan to find R, and finally zoom in to find the two targets.

These four types of trials were chosen to mimic realistic map exploration tasks, similar to those performed by sighted users on Google Maps or Open Street Maps, for example. *Zooming in* is required to focus on a specific area of the map. *Panning* is required when one is discovering what is around a specific area or landmark. *Zooming out and in* is a basic process in multi-scale environments [17]. It must be noted that all of these tasks can be performed with the Keyboard and with the Sliders (as shown in Study 1), and did not favor one interface in particular. However, with the Sliders, users were free to place the viewport anywhere on the map and to select any possible scale. Provided that participants were able to take advantage of this feature, it could lead to a greater satisfaction (increased sense of control) and to a better understanding (some distant landmarks could not be displayed simultaneously with the Keyboard, but they could be displayed simultaneously with the Sliders by precisely adjusting the scale). Finally, it must be noted that users can easily move from one discrete level to the next with the Sliders. They can rely on the two tactile cues that were added to the zooming slider area.

#### 6.1.6 *Variables.*

*6.1.6.1 Map Comprehension.* Many methods have been used or suggested to assess mental representations of VI users (see [33] for a review). The tests often measure differences between estimated and effective distances or directions between two landmarks or several landmarks. They usually rely on a set of questions or on graphical methods, such as sketching, filling blank elements on a map, or reconstructing a model using building blocks. In this work, we used both distance and direction questions, as well as a reconstruction task.

Two main variables were used: (1) the number of correct answers given by users to the questions asked at the end of each trial; (2) the bidimensional regression coefficient, which indicates how similar two 2D configurations are [68]. Additional questions concerning subjective map comprehension were also asked (Questionnaire 2).

*Multiple Choice Questions.* Four questions were asked after each trial. They were multiple choice questions with four options. Half of the questions, referred to as *Local* questions, required users to compare landmarks that belonged to the same cluster (i.e., a city and its surrounding towns and villages). The other half, referred to as *Global* questions, required users to compare one landmark to landmarks that did not belong to the same cluster (e.g., City A with towns B1 and B2). Questions concerned either *distances* or *directions*, and required the users to compare two landmarks only (*simple*) or three or more landmarks (*complex*).

Table 4. Examples of Multiple Choice Questions Asked After Each Trial

| Question | Type | Answer A | Answer B | Answer C | Answer D |
|---|---|---|---|---|---|
| I am in A, facing North. Where is B? | Simple/Direction | 12 h–3 h | 3 h–6 h | 6 h–9 h | 9 h–12 h |
| | Simple/Distance | <80km | 80–180 | 180–280 | >280km |
| Which landmarks are located east of A? | Complex/Direction | B | B, C | C, D | B, C, D |
| What is the shortest distance? | Complex/Distance | A–B | A–B1 | A–C | A–C1 |

Only one answer is correct.

In total, 32 questions were asked: 16 were *Local* questions and 16 were *Global* questions; out of the 16 *Local*/*Global* questions, there were four questions of each type (*distance*/*direction* × *simple*/*complex*). The following table gives an example of questions asked after each trial (Table 4).

*Map Reconstruction.* At the end of the eight trials, participants were asked to reconstruct the map. To do so, a set of magnets were placed on a magnetic board, each magnet being labeled with the Braille initial of a landmark. Two cities were already placed to provide anchor points and scale, and the edges of the map were delimited with magnetic strips. Participants were read the names of the landmarks before reconstruction. At the end of the reconstruction, a photo was taken and later analyzed to retrieve the coordinates of each landmark.

The reconstructed maps were compared to the initial map using a bidimensional regression analysis based on Euclidean geometry. This statistical method, initially proposed by Tobler [68], can be used to compute how similar two 2D-configurations are (regression coefficient ranging from 0 to 1). We used the true coordinates as independent variables and the coordinates of the reconstructed map as dependent variables.

*6.1.6.2 Usability.* To assess the usability of the interfaces for VI users, we used the same dependent variables as for Study 1. In addition, because we expected the task to be cognitively demanding, participants had to answer the NASA-TLX questionnaire [18] at the end of each session. Additionally, several questions were asked concerning the usability of each interaction technique (Questionnaire 3). Users also had to indicate which interface they preferred and for what reasons (Questionnaire 4). Finally, all sessions were video-recorded.

*6.1.7 Data Analysis.* We used the methodology described in Study 1. For completion times, we computed geometric means and 95% exact CIs on log-transformed data, and pairwise comparisons of completions times are expressed as ratios. For the other variables (percentage of correct answers, regression coefficients, and NASA-TLX scores), we computed means and 95% bootstrap CIs. For pairwise comparisons, we first computed differences between interfaces (Sliders minus Keyboard), for each participant and each factor. We then computed means and 95% bootstrap CIs. For distances only, we computed ratios between interfaces (Sliders/Keyboard), instead of differences between interfaces.[4]

*6.1.8 Procedure.* The experiment was composed of two sessions, each of which lasted approximately two and a half hours. During the first session (training only), participants were explained the basic features of the application and were explained how to pan and zoom using the two interfaces. During the second session (evaluation session per se), participants had a brief training

[4] As the distance that users had to pan depended on the trials, computing mean differences in distances across different types of trials would not have been relevant, and we therefore computed ratios.

Table 5. Description of the Procedure for the First Session (Training)

| Training session | |
|---|---|
| Introduction | Explanations concerning the goal and organization of the experiment<br>Consent form + photo/video authorization form<br>Questionnaire 1 (user's profile)<br>Clock face test[5] |
| Pan & Zoom | Explanations concerning panning and zooming using three tactile maps of different sizes and one tactile frame moved over the maps |
| Basic features | Selection, voice commands (*List, Scale,* and *Repeat*), centering – 3 trials[6]. |
| For each interface | Panning – 4 trials<br>Zooming – 4 trials<br>Panning and Zooming – 4 trials |

Table 6. Description of the Procedure for the Second Session (Evaluation)

| Evaluation session | |
|---|---|
| Introduction | Reminder of the aim of the experiment |
| Pan & Zoom | Reminder of the concepts of panning and zooming |
| Commands | Selection, voice commands (*List, Scale* and *Repeat*), centering – 2 trials |
| For each interface | Training – 7 trials<br>Test – 8 trials<br>Map reconstruction<br>SUS and NASA-TLX questionnaires<br>Questionnaires 2 (subjective comprehension) and 3 (usability) |
| Debriefing | Questionnaire 4 (users' preferences and comments) |

period before performing the test. The following tables summarize how both sessions were organized (Tables 5 and 6).

The order of presentation of the interfaces was counterbalanced within and between the sessions. For the evaluation session, two different maps were used and their order was also counterbalanced (half of the participants explored Map 1 first, and half explored Map 2 first).

### 6.2 Results

*6.2.1 Navigation Performances.*

*6.2.1.1 Success.* Out of the 128 trials, 115 were performed without the help of the evaluator and considered as successful. Other trials during which the experimenter had to help the participants were considered as unsuccessful. Out of the 13 unsuccessful trials, there was 1 *Pan*, 10 *Zoom in & Pan,* and 2 *Zoom out & in* trials. *Zoom in & Pan* trials mainly failed because participants became disoriented; one participant misunderstood the instruction; one participant forgot to zoom in to

[5] *Clock face test*: we ensured participants knew how to interpret directions given using the clock metaphor. We asked them to point on a raised-line clock-face the ticks corresponding to hours given by the evaluator. Participants also had to explore a raised-line map with three landmarks and to answer questions such as "I am in A. Where is B?" using clock directions.
[6] *Training*: For Session 1 and Session 2 training trials, participants had to find a target and were then asked a series of questions concerning the position of the target with respect to other landmarks. In that way, we ensured that participants not only performed the action properly, but also understood what the actions meant in terms of map manipulation/exploration. If an incorrect answer was given, explanations were given.

Table 7. Percentage of Successful Trials for Each Type of Trial and Interface

| Type of trial | Keyboard | Sliders | Total |
|---|---|---|---|
| Zoom in ($N = 16$) | 100% | 100% | 100% |
| Pan ($N = 16$) | 93.8% | 100% | 96.9% |
| Zoom in & Pan ($N = 16$) | 62.5% | 75% | 68.7% |
| Zoom out & in ($N = 16$) | 100% | 87.5% | 93.7% |

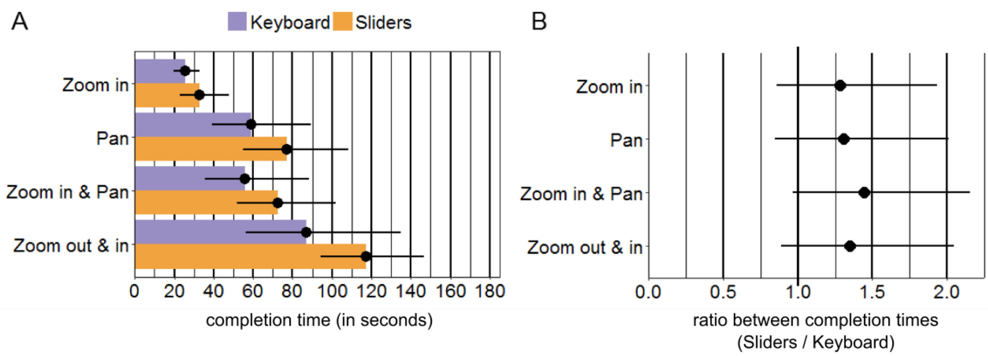


Fig. 12. (A) Mean completion times (in seconds) for each type of trial ($N = 8$). (B) Mean ratios between completion times (Sliders/Keyboard, $N = 8$). Values superior to 1 indicate larger times for the Sliders than for the Keyboard. Error bars show 95% exact CIs.

the Village level and therefore could not find the target. Two *Zoom out & in* trials failed because participants did not manage to correctly perform the required sequence of actions (zooming out, panning, centering, and zooming in). The following table gives the percentage of success according to each type of trial and interface (Table 7).

*6.2.1.2 Completion Times.* As a reminder, for completion times, we report geometric means and exact CIs, and differences between completion times are expressed as ratios. Similarly to the first study, we measured the time spent between the beginning of the trial and the moment when the last target was displayed. To compute mean completion times, we discarded unsuccessful trials.

*Zoom in* trials were performed faster than other trials, while *Zoom out & in* trials tended to take longer than other trials, especially with the Sliders (Figure 12(A)). The mean ratio between completion times (Sliders/Keyboard), computed across all types of trial, was 1.32, 95% CI [1.03, 1.71]. Figure 9(B) shows pairwise comparisons for each type of trials. Given the fact that interval endpoints are about seven times less plausible than the point estimate [7], the figure shows that participants performed faster with the Keyboard than with the Sliders. However, because the CIs are relatively long, we do not have a precise estimation of how large the effect of the interface on completion times is, but it ranges from zero effect to strong effect (up to 2 times longer).

*6.2.1.3 Distance Panned.* For each type of trial (except *Zoom* trials that did not require panning), we computed the ratio between the total distance panned (in kilometres) and the optimal distance.[7] Overall, participants panned much larger distances than required (Keyboard: *ratio* = 5.2, 95% CI

[7]Optimal distance was computed for each type of trial, based on possible paths with the Keyboard. For *Pan* trials, we used the average distance between the targets and the points of reference. For *Pan* and *Zoom in & Pan* trials, we computed the optimal distance for the worst-case scenario, i.e., in the case where the users did not choose the right direction (which had to be chosen randomly).

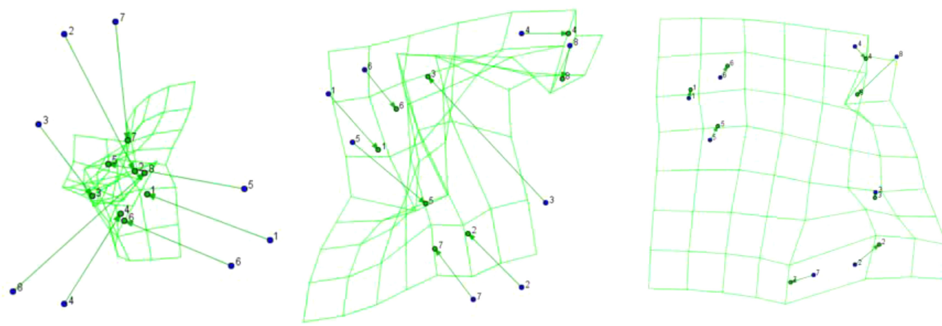


Fig. 13. Three examples of maps interpolation, based on the bidimensional regression analysis[8]. Arrows represent displacement vectors between the source map and the reconstructed map.

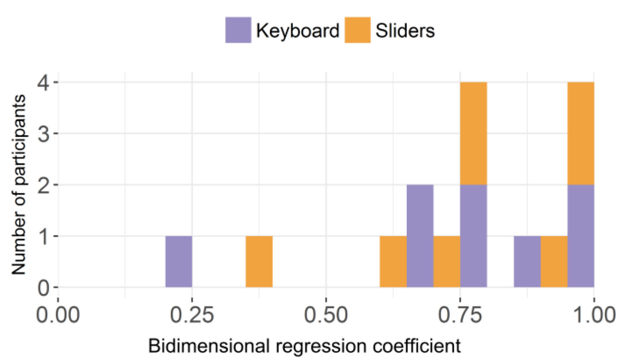


Fig. 14. Histogram of regression coefficients computed from a bidimensional regression analysis, for the 16 maps reconstructed by the participants. The higher the regression coefficient, the more similar the reconstructed map and the source map are.

[2.6, 11.2]; Sliders: *ratio* = 5.6, 95% CI [3.5, 9.7]). In addition, participants panned larger distances with the Sliders than with the Keyboard, as shown by the mean ratio between distances panned (Sliders/Keyboard), which was 4.0, 95% CI [−0.6, 8.5].

### 6.2.2 Map Comprehension.

*6.2.2.1 Bidimensional Analysis.* We report results about the bidimensional regression coefficient, which varies between 0 and 1, where 1 indicates the highest degree of similarity between the source map and the reconstructed map (see Figure 13 for examples). Regardless of the interaction techniques, 6 out of 16 maps were highly similar to the source map (*regression coefficient* > 0.8) and 8 were relatively similar to the source map (*regression coefficient* between 0.6 and 0.8). Only two maps strongly differed from the source map (P2 with the Keyboard; P8 with the Sliders), as shown in Figure 14. Five participants reconstructed at least one map whose regression coefficient was superior to 0.8 (P1, P3, P4, P5, and P6).

[8]Images were generated using Darcy, a software for bidimensional regression analysis: http://thema.univ-fcomte.fr/16-categories-en-francais/cat-productions-fr/cat-logiciels-fr/294-art-darcy.

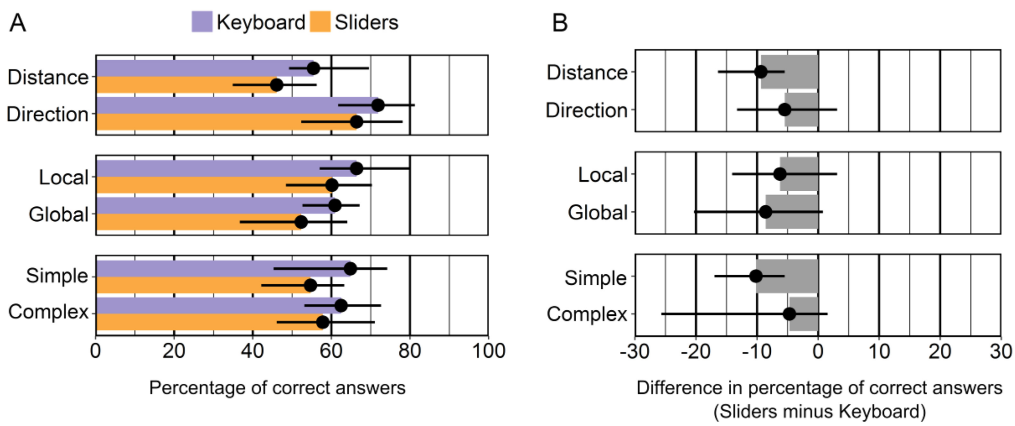


Fig. 15. (A) Percentage of correct answers for each type of questions, per interface ($N = 8$). (B) Mean differences of percentage of correct answers between interfaces (Sliders minus Keyboard, $N = 8$). Negative values indicate that the percentage of correct answers is higher for the Keyboard than for the Sliders. Error bars show 95% bootstrap CIs.

We computed the mean regression coefficient for each participant: two participants performed relatively poorly (P2 and P8, *mean regression coefficient* < 0.50); five performed well (P1, P3, P4, P6, and P7, *mean regression coefficient* between 0.70 and 0.85), and one performed very well (P5, *mean regression coefficient* > 0.95). Participants (P2 and P8) who did not perform well with one interface (*regression coefficient* < 0.4) did not obtain very high scores with the other interface either (*regression coefficient* < 0.65). On the contrary, participants who performed well with one interface tended to obtain similar scores with the other interface. This was particularly true for P5 who performed extremely well with both interfaces (*regression coefficients* > 0.95).

Results were similar for the two interfaces (Keyboard: *regression coefficient* = 0.74, 95% CI [0.54, 0.85]; Sliders: *regression coefficient* = 0.76, 95% CI [0.58, 0.87]). The mean difference between the interfaces (Sliders minus Keyboard), computed across all participants, was 0.02, 95% CI [−0.12, 0.17]. Differences were not consistent across the participants (half of them obtained better scores with the Sliders, the other half with the Keyboard).

*6.2.2.2 Multiple Choice Questions.* For each question, the chance level was 1/4 = 25%. Overall, the percentage of correct answers was reliably above the chance level (Keyboard: *score* = 63.7, 95% CI [56.6, 72.3]; Sliders: *score* = 56.2, 95% CI [44.9, 66.8], see Figure 15(A)). Figure 13(B) shows pairwise comparisons between the two interfaces (Sliders minus Keyboard) and illustrates that participants tended to perform slightly better with the Keyboard than with the Sliders (*mean difference* = −7.4 95% CI [−13.7, −3.9]), in particular for Distance and Simple questions.

*6.2.2.3 Subjective Comprehension and Memorization Strategies.* Participants were asked to report on a five-point Likert-scale whether they thought that they understood the map. Three participants answered 4 or 5 for both interfaces (P1, P5, and P6). Two participants (P4 and P7) answered 3 for both interfaces. P2, P3 and P8, respectively, answered 3, 3, and 2 for the Keyboard and 2, 4, and 3 for the Sliders. Overall, the median was 3.0 for the Keyboard and 3.5 for the Sliders.

Six participants tried to memorize the landmarks with respect to the cities, which they had memorized at the very beginning of the test. P2 said that she used the four cardinal points. P8 said that he remembered landmark locations relative to one another. P8 also said that he tried to remember the path he had navigated to find the targets as well as how he had explored them.

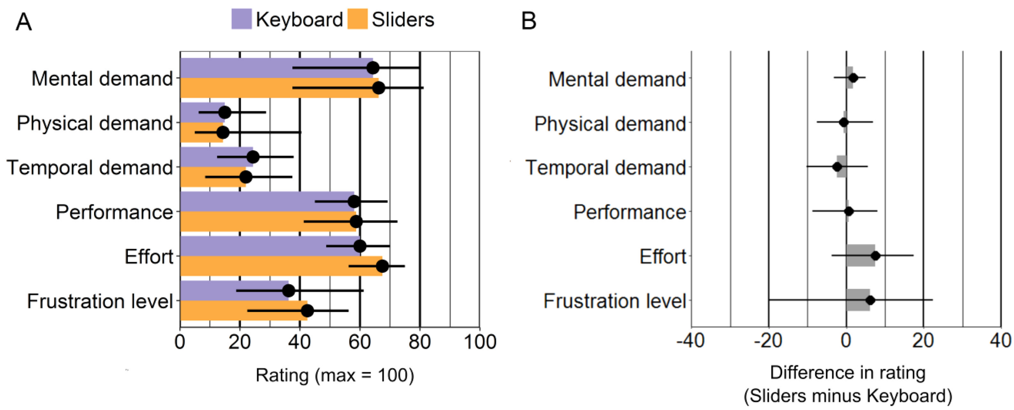


Fig. 16. (A) Mean workload (max: 100) for each dimension of the NASA-TLX questionnaire ($N = 8$). (B) For each dimension, mean difference between the interfaces (Sliders minus Keyboard, $N = 8$). A positive value indicates that participants rated the dimension higher for the Sliders than for the Keyboard. Error bars show 95% bootstrap CIs.

*6.2.3 Questionnaires and Preferences.* The average SUS score was 76.9 for the Keyboard (95% CI [68.1, 83.7]), and 72.8 for the Sliders (95% CI [60.6, 82.2]), showing very little differences between the two techniques in terms of perceived usability. Results from the NASA-TLX questionnaire (Figure 16(A)) show that the task was mentally demanding and that it required a relatively sustained effort. Pairwise differences computed for each dimension (Sliders minus Keyboard) did not reveal any reliable differences between the interfaces (see Figure 16(B)).

Concerning the preferences, six participants out of eight found the Keyboard more efficient; five found the Keyboard more pleasant to use, and five chose the Keyboard interface as their best choice overall.

*6.2.4 Qualitative Observations.*

*6.2.4.1 Interaction with the Keyboard and the Sliders.* In this section, we report qualitative observations concerning the use of the interfaces by the participants, based on videos and notes taken during the evaluations. Four participants (P2, P3, P4, and P6) expressed having difficulty moving the viewport as they intended, especially for Zoom in & Pan trials. This was particularly true with the Keyboard interface: when the viewport was too far away, they explicitly stated that they had difficulty in knowing which key they had to press in order to move the viewport toward the target.

Concerning the Sliders, only three participants (P1, P4, and P5) regularly took advantage of continuous zooming by deliberately selecting the most appropriate scale. Furthermore, when using the Sliders for panning, two participants (P2 and P6) experienced difficulties in understanding how to interpret their position, especially during the training session. In addition, four participants (P1, P2, P4, and P5) sometimes misused the sliders when they had to pan a large distance. For example, if they had to move the viewport toward the right, they first placed the slider far to the left, as one would do for “clutching”: by doing so, they thought that they will be able to pan larger distances but it only resulted in moving the viewport in the opposite direction.

Concerning the metaphors used, the two interfaces (Keyboard and Sliders) required the users to move a virtual viewport over a static map. The alternative would have been to drag the map under a static virtual viewport. As mentioned earlier, we opted for the first metaphor because we witnessed during preliminary observations that VI people are already used to it. Indeed, that same metaphor is used on screen readers where a virtual viewport is moved over a document.

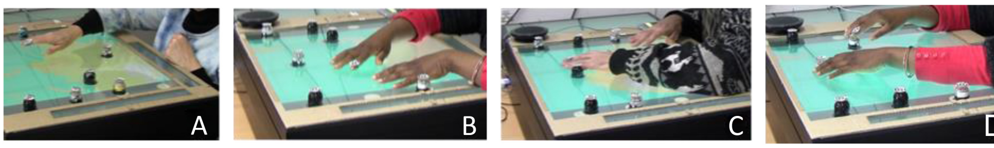


Fig. 17. Participants explored the map using one hand (A) or two hands (B–D). In that case, participants placed one hand at the corner of the table (C) or on a robot (D) and used the other hand to explore the map.

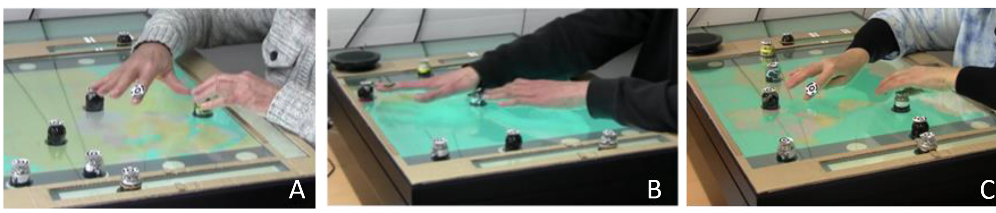


Fig. 18. Some ways participants estimated distance between robots and their relative positions.

However, all the participants tried to "move the map" instead of "moving the viewport" at least once during the experiment. This behavior was not frequent and disappeared as time went by. It was not possible to evaluate the exact number of occurrences of this behavior because the subject did not provide feedback on his ongoing actions and partial gestures would lead to ambiguous count.

The low number of incorrect actions ("clutching" or "moving the map") suggests that they were due to a lack of concentration and/or to the fact that the users were mimicking techniques used for other applications. We suggest that this is not an issue and that all these unexpected actions should disappear as users become more familiar with the interfaces and the exploration of "pan & zoom" maps. However, an experimental study regarding the mental manipulation of the model of the viewport and its consequence on the construction of a mental map based on the exploration of "pan & zoom" maps would be of interest.

Finally, it must be noted that on several occasions participants managed to recover from disorientation by developing suitable strategies. For example, P1, P3, and P8 zoomed out to the City level in order to relocate a particular City before finding a specific town and then the village they were looking for. P5, during the first session, spontaneously used the "center before zooming out" feature in order to successively display two Towns at the City level and compare distances. Some participants also used the "cancel" feature to relocate themselves.

*6.2.4.2 Interaction with the Robots.* In this section, we provide observations regarding how participants interacted with the landmarks (i.e., the robots). To explore the map, participants swept one hand (Figure 17(A)) or two hands over the surface to locate the robots and retrieve their names (Figure 17(B)–(D)). They also used one hand as an anchor, either placed at one corner of the table (Figure 17(C)) or on one robot (Figure 17(D)), while using the other hand to locate the remaining robots. We also observed that participants performed back-and-forth movements between robots previously selected, in order to check their names, but also to understand their relative position. During this process, several participants used their two hands to compare distances (see Figure 18), either by holding one hand on a robot as an anchor while the other hand was touching the other robots, or by counting how many hands/fingers could fit between two robots. This behavior was dynamic and it was not easy to categorize the gestures as uni- or bi-manual. As an illustration, when using the sliders for panning, many participants hold the two sliders at a time but move them alternatively.

### 6.3 Discussion of Study 2

*6.3.1 Visually Impaired Users can Understand "Pan & Zoom" Maps.* The accuracy of the reconstruction and the percentage of correct answers (reliably above chance level) indicate that, overall, participants understood and memorized the maps whose exploration required panning and zooming. We did not observe any differences between Local and Global questions. Most participants stated that they memorized the locations of the towns and villages with respect to their nearest city. Such a strategy probably helped them to answer Local questions, which concerned landmarks (towns or villages) that were located around a particular city. In addition, because they had memorized the configuration of the three cities, participants could draw inferences to answer Global questions. It should be noted that most of the questions required inferences as all of the landmarks mentioned in the questions could not be displayed within a single viewport. Hence participants had to combine pieces of information obtained from different viewport's positions in order to answer the questions.

Participants answered questions about directions better than distances. We notably observed that they frequently overestimated distances, and several users reported that they found it difficult to estimate distances. In fact, they did not use the "scale" voice command very often and most of them used the slider for zooming only as a switch to change the zoom level (e.g., from City Level to Town Level), and not to select a precise scale (e.g., from viewport representing 200km to viewport representing 150km). Altogether, it seems that participants were not comfortable with the concept of scale and rather relied on directions than distances in order to understand and memorize the map. Such results are coherent with empirical studies showing that blind users may need training to correctly estimate distances on tactile maps [72] and that errors in distance judgments are systematic [69].

Taken in a larger context, this result sheds light on VI users' spatial abilities. Difficulties for visually impaired users to correctly integrate various pieces of spatial information into one mental representation have been reported in a number of works. Different theories have been proposed to account for these difficulties (see [71] for a review). An early one stipulated that visual experience is essential to form mental spatial representations (the "deficiency" theory) but has now been rejected. The "inefficiency" theory states that the lack of visual experience necessarily leads to inefficient (or at least less efficient) spatial abilities. Our results tend to confirm the third theory, called the "difference" theory, which argues in favor of an amodal spatial representation system [39, 40]. In this theory, visual experience is not mandatory and other senses can be used to develop spatial abilities that may be of a different nature but still as functional as those developed by sighted users. For instance, in our study, one participant, who turned blind when he was four, performed extremely well with both interfaces. In addition, some VI participants spontaneously developed strategies that were similar to those developed by sighted users in Study 1. These results are in line with previous works on geographic maps for VI users [8].

*6.3.2 Participants Performed Better with the Keyboard than with the Sliders.* Participants tended to perform better with the Keyboard interface than with the Sliders interface in terms of navigation (shorter completion times and distances panned) and comprehension (higher percentage of correct answers). Besides, six participants out of eight found the Keyboard interface more efficient. Two explanations can be considered. First, in terms of interaction, the viewport could be moved diagonally by pressing one key only with the Keyboard interface, but with the Sliders interface the two sliders had to be moved simultaneously, resulting in a more complex set of actions. In Study 1, blindfolded participants also panned larger distances with the Sliders than with the Keyboard interface. In fact, even though participants were told to move the sliders for panning and then wait for feedback concerning the position of the viewport, some of them continuously moved the sliders, which resulted in incoherence between the feedback and the actual position of the viewport. This may have led participants to move the sliders too far and therefore to pan larger distances than necessary. Such an issue could be addressed by providing shorter feedback at a higher frequency. Also, the system could provide haptic or tactile feedback when the users move the sliders, which would help them estimate the distance they are moving the viewport.

In addition, the Sliders interface required users to rely on the absolute position of the viewport, a metaphor that users were not familiar with, as shown by incoherent uses of the sliders such as trying to perform "clutching" actions. We also observed that participants better answered *distance* questions with the Keyboard than with the Sliders, which may further indicate that they experienced difficulties in understanding how to correctly interpret the position of the sliders used for panning. However, five participants out of eight reported that the position of the sliders used for panning helped them to understand which part of the map was being displayed, which illustrates the potential usefulness of the sliders. Further training with the Sliders would certainly have been beneficial in helping participants be more comfortable with the concept of the absolute positioning of the viewport over a fixed map, as suggested by sighted participants in Study 1.

More generally, we compared two models for panning and zooming: *discrete* (with the Keyboard) vs. *continuous* (with the Sliders). Participants who preferred the Keyboard reported that they found it easier to pan with step-by-step movements. As for zooming, none of the participants mentioned a preference for one interface over the other. Unlike the sighted users in Study 1, very few participants used the slider for *fine* scaling (adjusting the scale within a given zoom level). However, it is unclear whether this was due to a lack of training or to the fact that selecting a more specific scale was not necessary to correctly perform the tasks. In addition, VI users are very familiar with keyboards, which provide *discrete* control, but rarely have access to input devices or interfaces that provide *continuous* control (such as sliders). This may also explain the advantage of the Keyboard over the Sliders.

*6.3.3 Participants Managed to Pan and Zoom but Sometimes Felt Disorientated.* Although more than 90% of the trials were successfully performed by the participants, data indicate that some participants experienced difficulties navigating the map, especially for *Zoom in & Pan* trials. We considered two potential explanations: users may have had difficulties in interacting with the Keyboard or the Sliders; users may have had difficulties in carrying out efficient strategies.

Data collected during the two studies provide evidences against the first explanation, stating that participants had difficulties in manipulating the Keyboard or the Sliders. The results from the first study, conducted with blindfolded participants, show that both interfaces enabled users to efficiently navigate the map in order to display and select specific targets. Similarly, during the training sessions of the second study, all VI users managed to easily display specific landmarks when they were given their direction, distance, and corresponding zoom level. Therefore, when participants were not engaged in a task that was cognitively demanding, they were able to interact with the Keyboard or the Sliders to correctly move the viewport or change the zoom level.

The second possible explanation is that participants may have had difficulty in carrying out efficient exploration strategies. In particular, for *Zoom in & Pan* trials, which required a complex set of actions, participants had to move the viewport around the point of reference to find surrounding landmarks. This proved difficult for several participants, especially with the Keyboard interface. For instance, they were going too far away from the point of reference, or they did not manage to systematically explore the map around the point of reference. This could be due to insufficient training or to the nature of feedback. The current position of the viewport was always provided with respect to its initial position (i.e., at the time of activation of the "pan" mode), both in terms of distance and direction; therefore, participants had to combine these two pieces of information in order to determine the current viewport's position and to decide which action to take. By focusing on one piece of information only (distance or direction), they may involuntarily move the viewport too far away (or too close) or in the wrong direction. As for zooming, it seems that participants were able to choose the correct level whenever necessary.

Overall, it seems that additional training would help users interpret the feedback. Alternative ways of giving feedback about the viewport's position could also be considered, such as providing a tactile viewport moving over a large overview of the map (similar to an inset map), or providing $x$- and $y$- coordinates (e.g., "100km left and 200km up") instead of polar coordinates (distance and direction).

*6.3.4 Relation between Navigation and Comprehension Performances.* Unsurprisingly, subjects who experienced navigation issues with one interaction technique were more likely to reconstruct the maps incorrectly: P2 and P4 panned excessive distances with one of the two interfaces (resp. Keyboard and Sliders), and their performance for map reconstruction was noticeably lower for the said technique. However, being able to correctly manipulate the interfaces did not systematically lead to good performances in terms of comprehension: P8 navigated very well but had the lowest regression coefficients. Nevertheless, in general, participants who reconstructed accurate maps obtained good performances in terms of navigation, which suggests that navigation skills are necessary but not sufficient to build accurate mental representations.

Results from the bidimensional regression analysis indicate that the majority of participants performed in a similar manner with both interfaces, which argue in favor of the importance of mental spatial skills, independent of the interface being used. If participants were able to build and manipulate map-like (allocentric) mental representations, they were likely to understand the maps with both interfaces. For example, all blindfolded participants of Study 1 obtained good results with both interfaces. Similarly, P5, who made the most accurate reconstructions and had the highest number of correct answers, was able to perform equally well with the two interfaces.

The question then arises as to why inter-individual differences can be observed. Participants' characteristics collected with the demographic questionnaire did not account for differences in spatial abilities: six participants reported having excellent mental representations (5/5) and two considered theirs to be neither good nor bad (3/5); all had been exposed to maps at school on several occasions but were not using maps anymore because of a lack of availability; participants' age was not correlated with performance. In addition, results were neither correlated with the age at onset of blindness nor with the degree of blindness (residual vision or not): in fact, P8, who turned blind at 17, and P2, who has residual vision, were the least successful. On the contrary, P5, who can be considered as early blind (onset of blindness at 4 years old), made the most accurate reconstructions.

Interestingly, P2 and P8 each reported strategies of memorization that were different from other participants: P2 used cardinal points and P8 tried to remember the path he had followed to find the landmarks. Therefore memorization strategies could better explain inter-individual differences than visual status (early or late blind) could, which argues in favor of the amodal theory of spatial cognition. Although these observations are qualitative and subjective, they may indicate that particular strategies of memorization must be encouraged when teaching VI users how to explore and understand "pan & zoom" maps. Such an explanation is in line with previous research work demonstrating the importance of teaching VI users how to develop efficient strategies for retrieving or encoding spatial information [4, 45, 74].

*6.3.5 Non-Visual Panning and Zooming is Cognitively Demanding.* The results show that VI people, regardless of their age at onset of blindness, are able to build and manipulate mental representations of "pan & zoom" maps, but results from the NASA-TLX questionnaire as well as observations also show that the task is cognitively demanding. This is consistent with literature on panning and zooming for sighed users. Bederson stated that "there is the potential that Zoomable User Interfaces tax human short-term memory because users must integrate in their heads the spatial layout of the information" [3]. This is especially true with haptic exploration that imposes sequential exploration within the viewport, and therefore makes further demands on short-term memory. Cognitive load was also certainly affected by the task itself (participants had to memorize both the names and configuration of many different landmarks) and by the fact that participants had to manipulate unfamiliar interfaces. Despite the training session, they were also not familiar with the concept of panning and zooming and surely had to be particularly focused to decide which action to take and how to take it. For example, some participants would sometimes use incorrect words for voice commands or make the opposite action of the one intended. Finally, the length of the experiment (2.5 hours in average) probably affected users' performances due to fatigue.

In the following section, we introduce the final study aiming at evaluating the whole system in a more realistic context (a genuine task, and independently moving robots). We specifically designed four additional features that could help users to better understand where the viewport is and navigate better, which would release cognitive resources.

## 7 STUDY 3: REALISTIC TASK AND INDEPENDENTLY MOVING ROBOTS

To investigate the use of BotMap in a more realistic context, we conducted sessions during which participants had to plan a journey through Africa using four navigation aids. These aids were designed to prevent users from feeling disorientated. In addition, because robots were moved by the evaluator in the first two studies, we wanted to collect feedback on the use of robots and to investigate whether using the system with independently moving robots led to specific usability issues.

### 7.1 Material and Methods

*7.1.1 Participants.* Three sighted (S) and three VI participants took part in these sessions. They were all familiar with the system, either because they took part in Study 1 (S3) or Study 2 (VI1 and VI3), or because they took part in pilot tests (S1, S2, and VI2). We recruited VI participants who performed relatively well during Study 2 (P5 and P7). Hence, we avoided a long training phase and kept the experiment relatively short (the robots' autonomy does not exceed one hour).

*7.1.2 Interface and Material.* We used the same setup as in the first two studies. We used the Keyboard interface only, which appeared as slightly more usable than the Sliders interface. The map was a map of Africa. The names and scales of the three zoom levels were changed accordingly. There were 6 metropolises (zoom level: Metropolis, scale: 9000km), 27 cities (zoom level: City, scale: 3000km) and 127 towns (zoom level: Town, scale: 1000km). Landmarks were chosen to be equally distributed over the map and so that there could never be more than six landmarks displayed at the same time. Finally, to make the scenario more realistic, the names of the countries were also given and for tourist locations the message "tourist" was played after the name of the country (e.g., "Johannesburg, metropolis of South Africa, tourist").

*7.1.3 Navigation Aids.* We implemented four additional features that users could access with voice commands: "*Home*", "*Where am I?*", "*Where is <name>?*", and "*Go to <name>*." When the "*Home*" feature is triggered, the map is reset to its original position and scale (zoom level: Metropolis), i.e., all metropolises are displayed within the viewport. The "*Where am I?*" feature provides

Table 8. List of Navigation Aids

| Voice Command | Feedback/output |
|---|---|
| *Home* | Reset the map to its initial position and zoom level |
| *Where am I?* | Provide feedback about the position of the viewport on the map + nearest landmarks |
| *Where is <name>?* | Provide feedback about the position of the said landmark (distance and direction) |
| *Go to <name>* | Center the said landmark and update the zoom level accordingly |

information about the viewport's position with respect to the entire map (e.g., "the viewport is on the upper-left of the map"), followed by information about the nearest landmark(s) of a particular zoom level (name, distance, and direction). At the Metropolis level, only the nearest metropolis is given; at the City level, information about the viewport's position is given, followed by the nearest metropolis and then by the nearest city; at the Town level, information about the viewport's position is given, followed by the nearest metropolis, then by the nearest city, then by the nearest town. "*Where is <name>?*" provides the distance and direction of the corresponding landmark with respect to the last selected landmark (e.g., "With respect to B, A is at 3 hours and 3000 kilometers"). If the last selected landmark is not displayed anymore, information is given with respect to the viewport's center (e.g., "With respect to the viewport's center, A is at 3 hours and 3000 kilometers"). Finally, "*Go to <name>*" enables users to center the map on a particular landmark with the corresponding zoom level (e.g., City zoom level for a city). Table 8 summarizes the four navigation aids.

*7.1.4 Task and Procedure.* Participants were reminded of the different features of the system and the four navigation aids were explained. They could explore a training map using all of the different commands, until they felt comfortable. They were then given the following scenario and were given 25 minutes to complete the task:

"As a reporter, you are making a documentary about social enterprises in Africa. You have already planned to meet three CEOs in three different places (Dakar, Djado, and Cairo), but you would also like to go sightseeing. You are therefore looking for one tourist city and one tourist town between Dakar and Djado, and one tourist city and one tourist town between Djado and Cairo. For environmental reasons, you aim to minimize the number of kilometers travelled. Using the interface, find an itinerary that fits all these criteria."

If participants completed the task in less than 15 minutes, another similar scenario was provided with different destinations. Participants could stop when they were satisfied with their itinerary. Then, participants were asked to answer a number of questions. For the navigation aids, they had to give their agreement on two statements using a five-points Likert scale: How useful the navigation aid was? Is the navigation aid easy to use? They were also asked to comment upon eight robot's attributes: discriminability, height, interactivity, shape, speed, stability, noise, number. In addition to these questions, we computed the number of times each navigation aid was used and the time needed for the robots to reach their new positions.

## 7.2 Results and Discussion

Apart from VI2, who planned two itineraries within the allotted time, participants planned a single itinerary. All participants managed to find two tourist cities and two tourist towns and most of these landmarks were either the best or the second best option to choose to minimize the number of kilometers travelled. On average, participants interacted with the system during 21 minutes

($SD = 4.2$), excluding training. Concerning the robots, they reached their new positions in 9 seconds on average ($SD = 2.8$).

*7.2.1 Strategies.* We used videos and logs to assess the strategies carried out by the participants. Four subjects (S2, S3, VI1, and VI2) used the same overall strategy: they first centered the map on the starting point (Dakar), and asked where the next destination (Djado) was located. Then, they performed panning and zooming actions in order to find surrounding tourist cities and towns. During these actions, they regularly checked that they were going in the appropriate direction by using the "Where Is Djado" command. Once they have identified a tourist city and a tourist town, they centered the map on the second destination (Djado) and repeated the procedure for the second half of the itinerary. Although S1 and VI3 used a similar strategy, they did it in a less systematic way. For example, S1 found a tourist city (Niamey) and zoomed in to find a tourist town in its neighborhood. Because he was unsuccessful, he went back to the starting point (Dakar) and panned in direction to Niamey, while looking for a city. Three participants (S1, VI1, and VI3) started exploring the map without any clear strategy. Afterward, they zoomed out or reinitialized the map to center onto the starting point, and they start exploring the map in a more systematic way.

*7.2.2 Navigation Aids.* All navigation aids were found very easy to use ($Mdn = 5$) and very useful ($Mdn = 5$). However, there was a small preference for the "*Where is ⟨name⟩?*" and "*Go to ⟨name⟩*" features in terms of usefulness ($Mdn = 5$), compared to "*Home*" ($Mdn = 3$) and "*Where Am I?*" ($Mdn = 3.5$). "*Home*" and "*Where Am I?*" were barely used (respectively, 2 times and 5 times across all sessions): participants did not feel disorientated and therefore did not find it necessary to use them. However, they acknowledged that they could be potentially useful. The "*Where is ⟨name⟩?*" feature was found to be very useful to go from one landmark to another without deviating, to remain orientated and also to estimate distances between two landmarks. Two participants (S3 and VI3) also used it to quickly locate one landmark within the viewport. It was used 76 times across all sessions. "*Go to ⟨name⟩*" was not used as often (25 times across all sessions): participants mainly used it to go back to one of the three landmarks indicated by the evaluator, without having to pan or zoom, thereby saving time. However, two participants (S3 and VI2) were skeptical about using "*Go to ⟨name⟩*," as both the viewport's position and zoom level could be updated at the same time without any explicit feedback.

In addition, four participants (S1, S2, VI1, and VI2) mentioned that they would enjoy having additional information on the map (e.g. museums and capitals) and being able to filter which landmarks to display. Two participants (S3 and VI1) also indicated that they would like to be able to retrieve a list of landmarks (e.g., tourist towns or museums) near a specific landmark. By doing so, they would only zoom in if the list is not empty or if they wanted more details about a landmark given in the list, which will allow them to save time.

To sum up, participants valued the four navigation aids, and the "*Where is ⟨name⟩?*" feature appeared to be essential. "*Go to ⟨name⟩*" was also found to be useful, but additional feedback should be provided to help users understand how the map is updated (viewport's position and zoom level). Participants found the system very comprehensive and stated that the proposed functionalities were sufficient to use it extensively and independently. In particular, two VI users stated that they would like to use it to explore the main points of interest of the city in which they live in (VI1) and the capitals of the countries of the former USSR (VI2).

*7.2.3 Independently Moving Robots.* Participants' comments about the robots' attributes were very positive and provide interesting insights about the design of actuated tabletop TUIs for VI users. Concerning the physical properties of the robots, five participants found their *height* ideal (3cm). Only one indicated that even though their height was not a problem, he would have preferred them to be smaller (S2). All participants found their *shape* ideal, especially because they do not take up a lot of place. Participants found that the robots were sufficiently *stable*, even though some stated that they needed to explore the map carefully (S2, S3, VI2, and VI3): greater stability would have been appreciated, but this was not critical. Interestingly, three participants (S1, S3, and VI1) indicated that the *number* of robots used was sufficient and that using more robots would result in the exploration being more tedious and should be avoided. The three others (S2, VI2, and VI3) indicated that more robots could be used to display more information. Five participants (S1, S2, S3, VI1, and VI3) found the *noise* made by the robots very useful as it helped them to know whether the robots were still moving or whether the system was working. Two participants (VI2 and VI3) even mentioned that the noise could be louder, so that participants could be positive that the robots are indeed moving. When discussing the *interactivity* of the robots, three participants (S1, S3, and VI2) said that they were very responsive. Two suggested that the robots could also be used as input devices, for example, to filter which information to display or to retrieve the names of landmarks within a certain distance (S1 and S2). When commenting upon the *discriminability* of the robots, three participants (S2, VI1, and VI3) stated that all robots should be similar and that using different sounds or shapes would not at all be useful. Two participants (S1 and VI2) suggested that different shapes could be used to help differentiate the robots (e.g., the larger the city, the higher the robot). Finally, none of the participants found the *speed* too slow. However, the three VI participants stated that if the robots could move faster, it would be better, but that the current speed was not an issue (VI1, VI2, and VI3).

## 8 GENERAL DISCUSSION AND PERSPECTIVES

### 8.1 Tangible "Pan & Zoom" Maps are Usable by People with Visual Impairment

In this section, we recap and discuss the main results framed around four research questions that were addressed in this study.

*8.1.1 Design of Non-Visual Interaction Techniques for Panning and Zooming on an Actuated Tabletop TUI.* Both results from Study 1 and Study 2 indicate that the Keyboard and Sliders interfaces enabled users to pan and zoom without vision, with a slight advantage of the Keyboard over the Sliders in terms of navigation performances. In Study 2, 90% of the trials were successfully performed, and results suggest that unsuccessful trials were due to a lack of training or inappropriate navigation or memorization strategies rather than to the manipulation of the input techniques per se. More generally, the "center" functionality as well as the voice command to retrieve the list of the landmarks currently displayed appeared to be very relevant. In addition, providing distance and direction feedback proved essential to enable users to locate the viewport when panning. In particular, five participants out of eight found the position of the sliders helped them to visualize the actual position of the viewport. However, training was required to understand how to interpret these pieces of information. Although the feedback that we designed for panning enabled users to perform most of the tasks (clock-face directions and relative position of the viewport with respect to its position when activating the "panning" mode), a number of "zoom-in and pan" trials were unsuccessful and participants found that the task was cognitively demanding. Therefore, alternative ways of providing distance and directions might be considered to try reducing the cognitive load and facilitating navigation.

Another way to facilitate navigation is to provide users with additional features, such as the ones developed for the Study 3. The four navigation aids were found to be very useful and easy to use by the participants, especially the "Where Is ⟨name⟩" feature, which enabled them to remain orientated and also to compute distances between two landmarks. Together with the other voice commands (list, scale, and so on) and functionalities (pan, zoom, and center), participants were able to successfully and independently find an itinerary between different landmarks displayed at different zoom levels, which further demonstrated the overall usability of the system. We therefore recommend designers of "pan & zoom" maps for VI users to integrate similar navigation aids in their systems.

Finally, the last study provided insights into the use of robots for actuated tabletop TUIs. In particular, different properties of the robots were found as ideal by most participants, notably their height (2.5cm), shape (2.5cm wide and circular) and noise (feeble but not silent). Usability could be improved if the robots were heavier or more stable, or could reposition themselves when they are involuntarily moved by the user. Interestingly, the low number of robots used was not seen as critical, and several participants indicated that for the specific task they had to perform, using a larger number of robots would not be beneficial.

*8.1.2 Which Panning and Zooming Modes (Discrete vs. Continuous) Should be Used for Visually Impaired Users?* Six VI participants found the Keyboard more efficient than the Sliders, and results suggest that participants performed better with the Keyboard than with the Sliders. In particular, participants panned larger distances with the Sliders than with the Keyboard. In addition, when zooming with the Sliders, only a few took advantage of the possibility to select a fine scale within a particular zoom level. However, the participants that performed best were able to perform equally well with both interfaces. We discussed in Section 6.3.2 how these observations may be related to a more frequent use of discrete than continuous pointing devices by VI people as well as to a lack of familiarity with absolute positioning. We suggest that with additional training users would be able to perform equally well with the two interfaces. Therefore, we recommend to use discrete control for novice users, but to consider implementing continuous control for expert users; this is similar to Google Maps and OpenStreetMap, which provide both types of control for sighted users.

*8.1.3 Visually Impaired Users Can Understand Large Tangible Maps Whose Exploration Requires Panning and Zooming.* Very importantly, we showed that VI users can understand maps whose exploration requires panning and zooming, regardless of their age and age at onset of blindness. These results are in line with previous studies conducted with blindfolded participants [35] as well as research on zoomable diagrams [41]. One limitation of our second study is that the maps were composed of only 10 landmarks and that there were only three zoom levels. However, 160 landmarks were used in the third study and participants were able to compute itineraries. Additional work is required to investigate to what extent using a larger number of landmarks and zoom levels would impact comprehension.

In addition, as discussed in Study 2, participants' performances appeared to depend on their spatial abilities, and more precisely their ability to generate and/or manipulate survey-like mental representations. In particular, it is interesting to note that the participant with the highest average comprehension score managed to perform very well with both interfaces (comprehension and navigation). We suggest that training could benefit users in terms of comprehension and navigation, and could also help reduce cognitive load.

*8.1.4 Is Map Comprehension Influenced by the Interaction Technique?* The interfaces relied on two different metaphors: the Keyboard did not provide users with any information about the current position of the viewport (relative displacement of the viewport) while the Sliders did (absolute positioning of the viewport). Although we would have expected the Sliders interface to lead to better comprehension scores, as it makes it possible to infer the position of the viewport on the map, participants tended to perform better with the Keyboard interface. As mentioned before, this may be due to a more frequent use of keyboards by VI users (to navigate a document for example), but also to the implementation of the Sliders, for which there could be inconsistencies between the actual positions of the sliders and the actual position of the viewport. Therefore, on that topic again, it would be interesting to evaluate the effect of more training on the use of the Sliders interface (which might require a higher level of abstraction), and to investigate whether navigation aids can compensate for the lack of information inherent to the Keyboard interface.

### 8.2 Limitations and Improvements

*8.2.1 Overall Usability.* Based on participants' feedback and observations, we identified three aspects of BotMap's design that could be improved. First, some participants had difficulty remembering all voice commands and sometimes used incorrect words. Future versions of the prototype should limit the number of voice commands and some of them could be replaced by physical buttons or tangible interactions. A more appropriate keyboard was also suggested by blindfolded participants, with more space between the keys used for panning and zooming (however, it should be noted that none of the VI users reported issues with the keyboard). Finally, similarly to what is done with visual sliders, the tangible sliders for panning could be improved either by physically representing the whole height/width of the viewport instead of its center only (e.g., with a length-adjustable bar), or by using more precise and easy to use sliders. To help users estimate how many kilometers they are moving the viewport, haptic and/or audio cues could also be provided during the sliders displacements.

*8.2.2 Implementation.* Currently, BotMap relies on robots, an underlying screen, and a camera placed above the table, leading to a quite complex setup. With recent robots such as the latest Ozobots,[9] which can be controlled remotely using a mobile application, or Cellulos [48], which are hand-held robots that can be used on non-interactive surfaces, the underlying screen would be useless.

Another limitation of the current version of BotMap concerns the number and proximity of robots, which impose restrictions on the number of map elements that can be simultaneously displayed. However, the recent development of Zooids [16], and more generally of Swarm User Interfaces, demonstrate that it is possible to develop actuated tabletop TUIs based on a much larger number of robots that can avoid collision, even when the surface is "crowded."

Finally, increasing the robots' speed could improve the usability of the system. Currently, there is a 9 seconds delay in average between the confirmation of a panning/zooming action and the end of the repositioning phase. Using optimized algorithms to assign robots to landmarks (e.g., the Hungarian algorithm [36]) as well as to compute trajectories [66] could greatly reduce this time. In addition, instead of quite slow Ozobots (44mm/s), faster robots can be used, such as Zooids [16] (44cm/s).

*8.2.3 Map Content.* Another point that needs to be addressed concerns the authoring of content for non-visual "pan & zoom" maps. For the second and third study, we assigned each landmark to a particular zoom level "by hand," which was a time-consuming process. In order to improve the accessibility of maps for VI users, there is a need to develop suitable algorithms that will automatically define the appropriate number of zoom levels and assign each landmark to one zoom level, while taking into account the nature of the map elements (e.g., a city vs. a point of interest such as a restaurant or a shopping mall) as well as their density. It would then be interesting to compare the evolution of the performance for the slider and the keyboard interaction as a function of the number of zoom levels. In addition, it will be necessary to investigate cases where one landmark (e.g., a city) is replaced by several landmarks when zooming in (e.g., neighborhoods). In that sense, previous works on non-visual zooming algorithms could be worth investigating

[9] http://ozobot.com/products/ozobot-evo.

[49, 56, 64], such as the one proposed and evaluated by Rastogi et al. [56] for the exploration of tactile diagrams, and the one proposed by Palani et al. [49] for the exploration of floorplans on a smartphone.

The current implementation of BotMap enables VI users to access maps of various complexities with landmarks only (in the third study, the map was composed of 160 landmarks). However, many types of maps include lines and areas. There are several ways this limitation can be addressed. For example, Ducasse et al. [11] proposed to use retractable reels as a way of drawing physical interactive lines between two landmarks. A combination of robots and retractable reels could be designed to provide VI users with access to various types of tangible graphics that are dynamic and include lines.

Sonification [41] can also be used to enable users to explore non-physical areas. For instance, a specific sound and verbal description can be used when the user's hand is entering an area. Bardot et al. [1] designed a solution where a smartwatch provided collocated vibrations and verbal descriptions when the hand was entering specific areas. In their study, VI users were able to explore and understand a digital map of a fictitious country made of adjacent states.

### 8.3 Perspectives for Future Research

In this section, we provide more general reflections concerning the improvement of non-visual "pan & zoom" maps for VI users:

*Developing affordable and reliable systems.* Better tangible devices are required to increase the accessibility of dynamic maps. One particular challenge is to design robots that are stable in addition to being small, affordable, and fast. Speed and ease of use (simple calibration, independent of light conditions, and so on) should also been considered to allow VI users to independently access "pan & zoom" maps. Finally, one could consider making tangible and dynamic maps that will not only include points (landmarks) but also physical lines (e.g., boundaries) and/or textured areas (e.g., national parks).

*Automating the adaptation of content.* Due to the limited amount of information that can be rendered tactilely, geospatial data need to be adapted in order to be accessible. This time-consuming process becomes even more complex when there is a large content, and when each element must be assigned a particular zoom level "by hand." Further research is needed to facilitate the adaptation of "pan & zoom" maps as well as to propose innovative ways to compensate for the restricted number of landmarks that can be displayed simultaneously. For example, a single robot could represent several landmarks at the same time, and users could locate these landmarks more precisely by zooming in. Users should also be able to quickly retrieve the names of landmarks located near a particular landmark by interacting with the corresponding robot.

*Reducing cognitive workload.* Research on visual panning and zooming interfaces has investigated various ways of reducing users' cognitive workload, especially by helping them to better navigate and avoid *desert fog*: use of animations, overviews, artificial landmarks, and cues, and so on. Although we already proposed a few navigation aids, further research is needed to design additional features and better evaluate their efficiency in terms of navigation performance, comprehension, and satisfaction.

*Tangible vs. non-tangible UIs.* As described in the related work section, a number of accessible maps for VI users exist and are based on different technologies, some of which presenting a number of advantages in terms of cost and/or portability. Therefore, it will be interesting to compare the usability of hybrid (pan & zoom) maps with the usability of digital maps, and identify in which contexts and for what content each type of map is preferable.

*Investigating other applications.* In this study, we focused on geographical maps, but many applications could benefit from the ability to pan and zoom. It is obviously useful for all other "spatial" graphics, such as graphs with large datasets (e.g., scatterplots, metro maps, and sky maps), technical drawings with many levels of details, and so on, but it could also be used in GUIs for browsing and searching different items (e.g., a virtual music store).

## 9 CONCLUSION

We described the design, implementation, and evaluation of an actuated tabletop TUI, named BotMap, that enables VI users to independently explore "pan & zoom" maps. Each landmark is represented by a robot, and whenever the map needs to be refreshed, the robots move to their new position. To interact with the map, we proposed two interfaces, the Keyboard and the Sliders, as well as a number of voice commands and navigation aids. We conducted three user studies. The first, conducted with blindfolded participants, demonstrated that both interfaces can be used to perform panning and zooming operations of various complexities without vision. The second study, conducted with VI users, demonstrated that users can understand maps whose exploration requires panning and zooming, and that they were able to pan and zoom, even though some felt disorientated on occasion and found that the task was cognitively demanding. We discussed a number of factors that may have explained differences in terms of navigation and comprehension (strategies of memorization, training, use of discrete vs. continuous controls, abilities to build map-like mental representations of space). In the final study, participants had to plan a journey through Africa using four navigation aids. This study showed the potential of these aids to facilitate navigation and gave interesting insights into the design of actuated tabletop TUIs for VI users. We concluded by discussing to what extent the prototype could be improved, notably in terms of implementation, and proposed a number of perspectives for further research on non-visual panning and zooming. We suggest that the pieces of information related to the design, development, and evaluation of BotMap as well as the perspectives that we identified will facilitate and encourage the design and deployment of actuated tangible «pan & zoom» maps for VI, and, ultimately, empower VI people by giving them the opportunity to independently explore and interact with complex data in the same way that sighted users do.

## REFERENCES

[1] Sandra Bardot, Marcos Serrano, and Christophe Jouffrais. 2016. From tactile to virtual: Using a smartwatch to improve spatial map exploration for visually impaired users. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI'16)*. ACM, 100–111. http://doi.org/10.1145/2935334.2935342

[2] SE Battersby. 2008. User-centered design for digital map navigation tools. In *Proceedings of The 17th International Research Symposium on Computer-based Cartography*. 1–15.

[3] Benjamin B. Bederson. 2011. The promise of zoomable user interfaces. *Behav. Inform. Technol.* 30, 6 (2011), 853–866. http://doi.org/10.1080/0144929X.2011.586724

[4] Edward P. Berla and Lawrence H. Butterfield. 1977. Tactual distinctive features analysis: Training blind students in shape recognition and in locating shapes on a map. *J. Special Edu.* 11, 3 (1977), 335–346.

[5] Anke M. Brock, Philippe Truillet, Bernard Oriola, Delphine Picard, and Christophe Jouffrais. 2015. Interactivity improves usability of geographic maps for visually impaired people. *Hum.–Comput. Interact.* 30 (2015), 156–194.

[6] John Brooke. 1996. SUS: A "quick and dirty" usability scale. In *Usability Evaluation in Industry*, P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland (Eds.). Taylor & Francis, London, UK, 189–194.

[7] Geoff Cumming. 2013. The new statistics why and how. *Psychol. Sci.* 25 (2013), 7–29.

[8] Franco Delogu, Massimiliano Palmiero, Stefano Federici, Catherine Plaisant, Haixia Zhao, and Olivetti Belardinelli. 2010. Non-visual exploration of geographic maps: Does sonification help? *Disab. Rehab.: Assistive Technol.* 5, 3 (2010), 164–174. http://doi.org/10.3109/17483100903100277

[9] Pierre Dragicevic. 2015. *HCI statistics without p-values.* [Research Report] RR-8738, Inria, 32 pages.

[10] Julie Ducasse, Anke Brock, and Christophe Jouffrais. 2018. Accessible interactive maps for visually impaired users. In *Mobility in Visually Impaired People – Fundamentals and ICT Assistive Technologies*, E. Pissaloux and R. Velasquez (Eds.). Springer, 537–584.

[11] Julie Ducasse, Marc J.-M. Macé, Marcos Serrano, and Christophe Jouffrais. 2016. Tangible reels: Construction and exploration of tangible maps by visually impaired users. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2186–2197. http://doi.org/10.1145/2858036.2858058

[12] F. De Felice, F. Renna, G. Attolico, and A. Distante. 2007. A haptic/acoustic application to allow blind the access to spatial information. In *Proceedings of the Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)*. IEEE, 310–315. http://doi.org/10.1109/WHC.2007.6

[13] Martin J. Gardner and Douglas G. Altman. 1986. Confidence intervals rather than P values: Estimation rather than hypothesis testing. *Br. Med. J.* 292, 6522 (1986), 746–750.

[14] Gary R. VandenBos (Ed.) APA. 2010. *Manual of the American Psychological Association*. (A. P. Association, Ed.). https://doi.org/10.1006/mgme.2001.3260

[15] F. Gaunet, J. L. Martinez, and C. Thinus-Blanc. 1997. Early-blind subjects' spatial representation of manipulatory space: Exploratory strategies and reaction to change. *Perception* 26, 3 (1997), 345–366.

[16] Mathieu Le Goc, Lawrence H. Kim, Ali Parsaei, Jean-daniel Fekete, Pierre Dragicevic, and Sean Follmer. 2016. Zooids: Building blocks for swarm user interfaces. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST'16)*. ACM, 97–109. http://doi.org/10.1145/2984511.2984547

[17] Yves Guiard and Michel Beaudouin-Lafon. 2004. Target acquisition in multiscale electronic worlds. *Int. J. Hum.-Comput. Stud.* 61, 6 (2004), 875–905.

[18] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Human Mental Workload*, P. A. Hancock and N. Meshkati (Eds.). Elsevier, 139–183. http://doi.org/10.1016/S0166-4115(08)62386-9

[19] Sabine Hennig, Fritz Zobl, and Wolfgang W. Wasserburger. 2017. Accessible web maps for visually impaired users: Recommendations and example solutions. *Cartogr. Perspect.* 88 (November 2017), 6–27. http://doi.org/10.14714/CP88.1391

[20] Lars Erik Holmquist, Johan Redström, and Peter Ljungstrand. 1999. Token-based acces to digital information. In *Proceedings of the Handheld and Ubiquitous Computing: First International Symposium (HUC'99), Karlsruhe, Germany, September 27–29*. Springer, Berlin, 234–245. http://doi.org/10.1007/3-540-48157-5_22

[21] Kasper Hornbaek, Benjamin B. Bederson, and Catherine Plaisant. 2003. Navigation patterns & usability of zoomable user interfaces. *Interactions* 10, 1 (2003), 11. http://doi.org/10.1145/604575.604582

[22] C. J. Huang, Ellen Yi-luen Do, and D. Gross. 2003. Mousehaus table: A physical interface for urban design. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology*. 41–42.

[23] Hiroshi Ishii and Brygg Ullmer. 1997. Tangible bits: Towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. 234–241. http://doi.org/10.1145/258549.258715

[24] Mihail Ivanchev, Francis Zinke, and Ulrike Lucke. 2014. Pre-journey visualization of travel routes for the blind on refreshable interactive tactile displays. In *Proceedings of the International Conference on Computers Helping People with Special Needs (ICCHP'14)*, LNCS, vol. 8548. Springer International Publishing. http://doi.org/10.1007/978-3-319-08599-9

[25] R. D. Jacobson. 1998. Navigating maps with little or no sight: An audio-tactile approach. In *Proceedings of Content Visualization and Intermedia Representations*. 95–102.

[26] Gunnar Jansson, Imre Juhasz, and Arina Cammilton. 2006. Reading virtual maps with a haptic mouse: Effects of some modifications of the tactile and audio-tactile information. *Br. J. Vis. Impair.* 24, 2 (2006), 60–66. http://doi.org/10.1177/0264619606064206

[27] Hans-Christian Jetter, Svenja Leifert, Jens Gerken, Sören Schubert, and Harald Reiterer. 2012. Does (multi-)touch aid users' spatial memory and navigation in "panning" and in "zooming & panning" UIs? In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI'12)*. ACM, 83–90. http://doi.org/10.1145/2254556.2254575

[28] Catherine Emma Jones and Valérie Maquil. 2016. Towards geospatial tangible user interfaces: An observational user study exploring geospatial interactions of the novice. *Adv. Intel. Syst. Comput.* 582 (2016), 104–123. http://doi.org/10.1007/978-3-319-29589-3_7

[29] Susanne Jul and George W. Furnas. 1998. Critical zones in desert fog: Aids to multiscale navigation. In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology (UIST'98)*. 97–106. http://doi.org/10.1145/288392.288578

[30] Hesham M. Kamel and James A. Landay. 2002. Sketching images eyes-free. In *Proceedings of the Fifth International ACM Conference on Assistive Technologies (Assets'02)*. New York, USA: ACM Press, 33–42. https://doi.org/10.1145/638249.638258

[31] Shaun K. Kane, Meredith Ringel Morris, Annuska Z. Perkins, Daniel Wigdor, Richard E. Ladner, and Jacob O. Wobbrock. 2011. Access overlays: Improving non-visual access to large touch screens for blind users. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST'11).* ACM, 273–282. http://doi.org/10.1145/2047196.2047232

[32] Shaun K. S. K. Kane, Meredith Ringel M. R. Morris, and Jacob O. J. O. Wobbrock. 2013. Touchplates: Low-cost tactile overlays for visually impaired touch screen users. In *Proceedings of the ASSETS'13 – SIGACCESS International Conference on Computers and Accessibility.* ACM.

[33] R. M. Kitchin. 2000. Collecting and analysing cognitive mapping data. In *Cognitive Mapping: Past, Present and Future*, R. Kitchin and S. Freundschuh (Eds.). Routledge, 9–23.

[34] Roberta L. Klatzky, Nicholas A. Giudice, Christopher R. Bennett, and Jack M. Loomis. 2014. Touch-screen technology for the dynamic display of 2D spatial information without vision: Promise and progress. *Multisens. Res.* 27, 5–6 (2014), 359–378. http://doi.org/10.1163/22134808-00002447

[35] Roberta L. Klatzky, Jack M. Loomis, Susan J. Lederman, Hiromi Wake, and Naofumi Fujita. 1993. Haptic identification of objects and their depictions. *Attention Percept. Psychophys.* 54, 2 (1993), 170–178.

[36] Harold W. Kuhn. 1955. The Hungarian method for the assignment problem. *Naval Res. Logis. Q.* 2, 1–2 (1955), 83–97.

[37] Uwe Laufs, Christopher Ruff, and Jan Zibuschka. 2010. MT4j – A Cross-platform multi-touch development framework. In *Proceedings of the ACM EICS 2010, Workshop: Engineering Patterns for Multi-touch Interfaces.* ACM, 52–57.

[38] J. M. Loomis, R. L. Klatzky, and Susan J. Lederman. 1991. Similarity of tactual and visual picture recognition with limited field of view. *Perception* 20, 2 (1991), 167–177.

[39] Jack M. Loomis, Roberta L. Klatzky, and Nicolas A. Giudice. 2013. Representing 3D space in working memory: Spatial images from vision, hearing, touch, and language. In *Multisensory Imagery*, S. Lacey and R. Lawson (Eds.). Springer, New York, NY, 131–155. http://doi.org/10.1007/978-1-4614-5879-1

[40] J. M. Loomis and R. L. Klatzky. 2008. Functional equivalence of spatial representations from vision, touch, and hearing: Relevance for sensory substitution. In *Blindness and Brain Plasticity in Navigation and Object Perception*, J. Rieser, D. Ashmead, F. Ebner, and A. Corn (Eds.). New York: Taylor & Francis, 155–184. https://doi.org/10.4324/9780203809976

[41] Douglass Lee Mansur. 1984. *Graphs in Sound: A Numerical Data Analysis Method for the Blind.* Technical report. https://www.osti.gov/biblio/7044393.

[42] Valérie Maquil, Thomas Psik, and Ina Wagner. 2008. The colortable: A design story. In *Proceedings of the 2nd International Conference on Tangible and Embedded Interaction (TEI'08).* 97–104.

[43] David McGookin, Euan Robertson, and Stephen S. A. Brewster. 2010. Clutching at straws: Using tangible interaction to provide non-visual access to graphs. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI'10).* ACM, 1715–1724. http://doi.org/10.1145/1753326.1753583

[44] Susanna Millar and Zainab Al-Attar. 2001. Illusions in reading maps by touch: Reducing distance errors. *Br. J. Psychol.* 92, 4 (2001), 643–657. http://doi.org/10.1348/000712601162392

[45] Valerie S. Morash, Allison E. Connell Pensky, Steven T. W. Tseng, and Joshua A. Miele. 2014. Effects of using multiple hands and fingers on haptic performance in individuals who are blind. *Perception* 43, 6 (2014), 569–588.

[46] Mathieu Nancel, Julie Wagner, Emmanuel Pietriga, Olivier Chapuis, and Wendy Mackay. 2011. Mid-air pan-and-zoom on wall-sized displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 177–186.

[47] Sile O'Modhrain, Nicholas A. Giudice, John A. Gardner, and Gordon E. Legge. 2015. Designing media for visually-impaired users of refreshable touch displays: Possibilities and pitfalls. *IEEE Trans. Haptics* 8, 3 (2015), 248–257. http://doi.org/10.1109/TOH.2015.2466231

[48] A. Ozgür, S. Lemaignan, W. Johal, M. Beltran, M. Briod, L. Pereyre, and P. Dillenbourg. 2017. Cellulo: Versatile handheld robots for education. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI'17).* 119–127. https://doi.org/10.1145/2909824.3020247

[49] Hariprasath Palani, Uro Giudice, and Nicholas A. Giudice. 2016. Evaluation of non-visual zooming operations on Touchscreen devices. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* Springer International Publishing, 162–174. http://doi.org/10.1007/978-3-319-40244-4_16

[50] Gian Antonio Pangaro, Dan Maynes-Aminzade, and Hiroshi Ishii. 2003. The actuated workbench: 2D actuation in tabletop tangible interfaces. *Interfaces* 4, 2 (2003), 181–190. http://doi.org/http://doi.acm.org/10.1145/571985.572011

[51] James Patten and Hiroshi Ishii. 2007. Mechanical constraints as computational constraints in tabletop tangible interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07).* 809. http://doi.org/10.1145/1240624.1240746

[52] Esben Warming Pedersen and Kasper Hornbæk. 2011. Tangible Bots : Interaction with active tangibles in tabletop interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 2975–2984.

[53] Benjamin Poppinga, Charlotte Magnusson, Martin Pielot, and Kirsten Rassmus-Gröhn. 2011. TouchOver map: Audio-tactile exploration of interactive maps. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI'11)*. ACM, 545–550. http://doi.org/10.1145/2037373.2037458

[54] Ivan Poupyrev, T. Nashida, and M. Okabe. 2007. Actuation and tangible user interfaces: The Vaucanson duck, robots, and shape displays. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction*. 205–212.

[55] Roman Rädle, Hans-Christian Jetter, Simon Butscher, and Harald Reiterer. 2013. The effect of egocentric body movements on users' navigation performance and spatial memory in zoomable user interfaces. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces (ITS'13)*. 23–32. http://doi.org/10.1145/2512349.2512811

[56] Ravi Rastogi, T. V. Dianne Pawluk, and Jessica Ketchum. 2013. Intuitive tactile zooming for graphics accessed by individuals who are blind and visually impaired. *IEEE Trans. Neural Syst. Rehab. Eng.: A Publ. IEEE Eng. Med. Biol. Soc.* 21, 4 (2013), 655–63. http://doi.org/10.1109/TNSRE.2013.2250520

[57] Carlo Ratti, Yao Wang, Hiroshi Ishii, Ben Piper, and Dennis Frenchman. 2004. Tangible user interfaces (TUIs): A novel paradigm for GIS. *Trans. GIS* 8, 4 (2004), 407–421.

[58] Eckard Riedenklau. 2016. Development of actuated tangible user interfaces: New interaction concepts and evaluation methods. Dissertation for degree of Dr-Ing. Faculty of Technology at the Bielefeld University.

[59] Eckard Riedenklau, Thomas Hermann, and Helge Ritter. 2010. Tangible active objects and interactive sonification as a scatter plot alternative for the visually impaired. In *Proceedings of the 16th International Conference on Auditory Display (ICAD'10)*. 1–7.

[60] Eckard Riedenklau, Thomas Hermann, and Helge Ritter. 2012. An integrated multi-modal actuated tangible user interface for distributed collaborative planning. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*. 169–174.

[61] André Rodrigues, André Santos, Kyle Montague, and Tiago Guerreiro. 2017. Improving smartphone accessibility with personalizable static overlays. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 37–41. http://doi.org/10.1145/3132525.3132558

[62] Jaime Sánchez and Natalia de la Torre. 2010. Autonomous navigation through the city for the blind. In *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility*. 195–202.

[63] Jeff Sauro and James R. Lewis. 2010. Average task times in usability tests: What to report? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2347–2350.

[64] Bernhard Schmitz and Thomas Ertl. 2010. Making digital maps accessible using vibrations. In *Proceedings of the International Conference on Computers Helping People with Special (ICCHP'10), Part I*. LNCS, vol. 6179, K. Miesenberger, J. Klaus, W. Zagler, and A. Karshmer (Eds.). Springer, 100–107.

[65] Bernhard Schmitz and Thomas Ertl. 2012. Interactively displaying maps on a tactile graphics display. In *Proceedings of the Spatial Knowledge Acquisition with Limited Information Displays (SKALID'12)*. 13–18.

[66] Jamie Snape, Jur Van Den Berg, Stephen J. Guy, and Dinesh Manocha. 2011. The hybrid reciprocal velocity obstacle. *IEEE Trans. Robot.* 27, 4 (2011), 696–706.

[67] Saiganesh Swaminathan, Thijs Roumen, Robert Kovacs, David Stangl, Stefanie Mueller, and Patrick Baudisch. 2016. Linespace: A sensemaking platform for the blind. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI'16*. ACM, 2175–2185. http://doi.org/10.1145/2858036.2858245

[68] Waldo R. Tobler. 1994. Bidimensional regression. *Geogr. Anal.* 26, 3 (1994), 187–212.

[69] Barbara Tversky. 1981. Distortions in memory for maps. *Cogn. Psychol.* 13, 3 (1981), 407–433.

[70] J. Underkoffler and H. Ishii. 1999. Urp: A luminous-tangible workbench for urban planning and design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 386–383. http://doi.org/10.1145/302979.303114

[71] Simon Ungar. 2000. Cognitive mapping without visual experience. In *Cognitive Mapping: Past, Present and Future*, R. Kitchin and S. Freundschuh (Eds.). Routledge, Oxon, UK, 221–248.

[72] Simon Ungar, Mark Blades, and Christopher Spencer. 1997. Teaching visually impaired children to make distance judgement from a tactile map. *J. Vis. Impair. Blindness* 91, 2 (1997), 163–174.

[73] Simon Ungar, Mark Blades, and Christopher Spencer. 1995. Mental rotation of a tactile layout by young visually impaired children. *Perception* 24, 8 (1995), 891–900.

[74] Simon Ungar, Mark Blades, and Christopher Spencer. 1997. Teaching visually impaired children to make distance judgments from a tactile map. *J. Vis. Impair. Blindness* 91, 163–174.

[75] Steven Wall and Stephen Brewster. 2006. Feeling what you hear: Tactile feedback for navigation of audio graphs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1123–1132. http://doi.org/10.1145/1124772.1124941

[76] Maarten W. A. Wijntjes, Thijs van Lienen, Ilse M. Verstijnen, and Astrid M. L. Kappers. 2008. The influence of picture size on recognition and exploratory behaviour in raised-line drawings. *Perception* 37, 4 (2008), 602–614. http://doi.org/10.1068/p5714

[77] Limin Zeng and Gerhard Weber. 2012. ATMap: Annotated tactile maps for the visually impaired. In *Cognitive Behavioural Systems: COST 2102 International Training School*, LNCS, vol. 7403. Springer, Berlin, 290–298. http://doi.org/10.1007/978-3-642-34584-5

[78] Haixia Zhao, Catherine Plaisant, Ben Shneiderman, and Jonathan Lazar. 2008. Data sonification for users with visual impairment. *ACM Trans. Comput.-Hum. Interact.* 15, 1 (2008), 1–28. http://doi.org/10.1145/1352782.1352786

Received July 2017; revised February 2018; accepted April 2018