AGILE: GIScience Series, 5, 7, 2024. https://doi.org/10.5194/agile-giss-5-7-2024
Proceedings of the 27th AGILE Conference on Geographic Information Science, 4–7 Sept. 2024.
Editors: Alison Heppenstall, Mingshu Wang, Urska Demsar, Rob Lemmens, and Jing Yao.
This contribution underwent peer review based on a full paper submission.

FreeMapRetrieve: Freehand Gestures for Retrieve Operations in Large-Screen Map Environments

Jonas Hurst [1], Auriol Degbelo [2], and Christian Kray [1]

[1]Institute for Geoinformatics, University of Münster
[2]Chair of Geoinformatics, TU Dresden

Correspondence: Auriol Degbelo (auriol.degbelo@tu-dresden.de)

Abstract. Current research has investigated freehand gestures for Pan and Zoom operations on maps on large screens. Freehand gestures for Retrieve operations, however, have remained largely unexplored. To address this gap, this work introduced two mechanisms for Retrieve on large displays in a research-through-design study: Pointer-to-Feature (i.e. moving a pointer to a geographic feature and then performing a hand gesture to achieve Retrieve) and Feature-to-Pointer (i.e. moving a geographic feature to a stationary pointer at the screen's centre through panning/zooming of the map, before performing a hand gesture to achieve Retrieve). The evaluation of a prototype (FreeMapRetrieve) regarding usability and data exploration utility showed that both mechanisms work well. The two techniques are comparable for large polygons but Pointer-to-Feature is slightly more efficient for smaller polygons. Reflections on the design process yield lessons learned that are relevant to designers of gesture-based interaction for maps on large displays.

Keywords. map interaction, mid-air gestures, large displays, gesture recognition, retrieve, details-on-demand

1 Introduction

Maps are useful for various tasks such as exploratory data analysis (Edsall et al., 2009), data journalism (Griffin, 2020), citizen participation (Rinner and Bird, 2009) and wayfinding (Schwering et al., 2023). When used on large displays, they can help communicate science results to the public, e.g. in an exhibition or a science centre, where large surfaces are important for greater visibility (Bartoschek et al., 2014). While maps have become ubiquitous, consolidated design guidelines specific to map interaction are still needed (Roth, 2013b; Kray et al., 2017; Degbelo, 2022). These guidelines, derived empirically wherever possible, will form a key ingredient of a science of map interaction (Roth, 2013b) and a science of interaction (Pike et al., 2009) more broadly. The long-term goal of the current work is to formulate such guidelines for gestural interaction with maps. The contribution of this article is to be placed in that context, with a focus on the *Retrieve* operation using gestures on large displays.

Why 'Retrieve'. Freehand gestures for Pan and Zoom operations within large high-resolution display environments have been evaluated sufficiently in the past. Yet, research on gestures to enable Retrieve operations for maps on large screens is currently lacking. In line with Roth (2013a), Retrieve operations (a.k.a. details-on-demand) are interactions that request specific details about a map feature of interest. In this work, Retrieve means querying a spatial feature for its associated non-spatial attributes. For spatial datasets with lots of non-spatial attributes, it is often not possible to convey all information about a spatial feature through cartographic visualization alone. In such cases, a Retrieve mechanism is necessary to reveal attributes which are not visualized. Therefore, Retrieve operations are essential in conveying all information available to the person interacting with the spatial dataset and allow them to fully comprehend it. It is one of the basic tasks in exploratory spatial data analysis. The research question addressed in this work is: How to best design gestures for Retrieve operations in large-screen map environments?

Contributions. There is a good deal of gesture elicitation studies on large screens, which provide insights into users' mental models of surface gestures for specific tasks (e.g. Austin et al., 2020; Wittorf and Jakobsen, 2016; Du et al., 2019). By contrast, gesture performance evaluation studies, which provide insight into the cost and benefits of using a specific gesture for a given task (e.g. Sluÿters et al., 2023; Hatscher et al., 2017) seem less frequent, despite their necessity for a more thorough understanding of the adequacy of gestures for specific tasks. This work contributes a gesture performance evaluation study, which helps to learn about viable gesture combinations for Retrieve for map interaction on large displays.

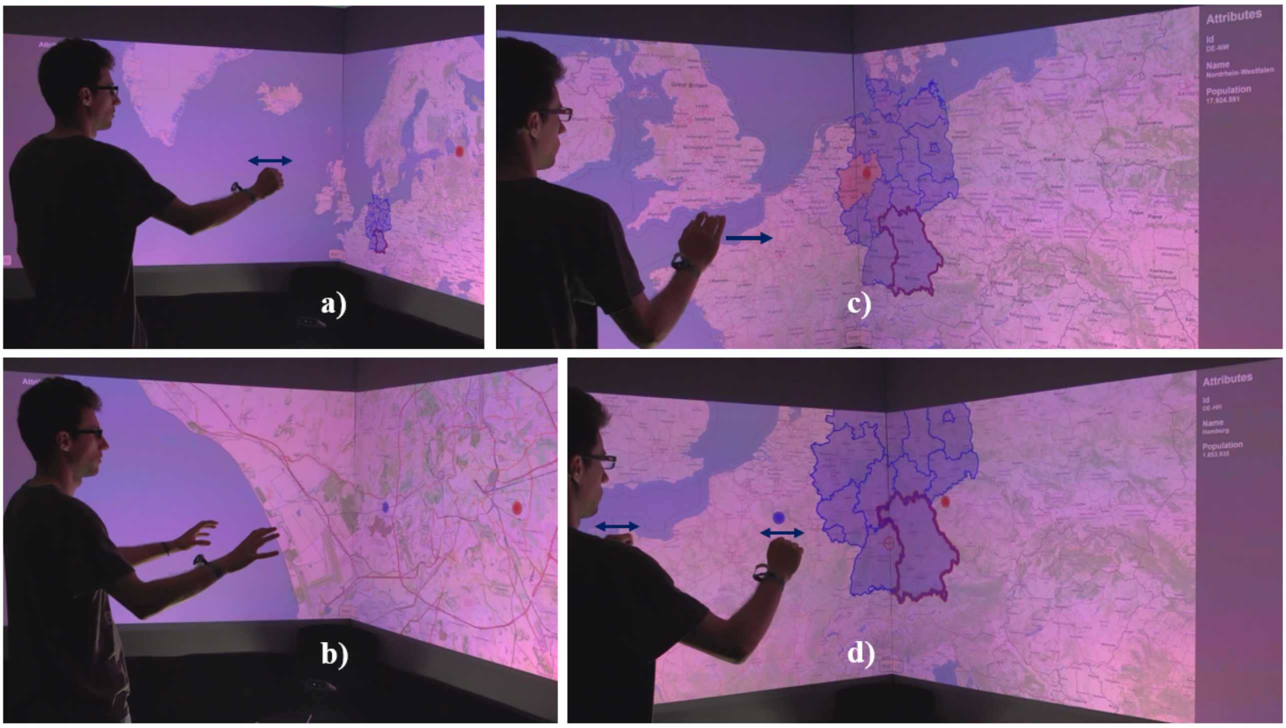


Figure 1. Freehand gestures implemented and evaluated during the work: a) a user grabbing the map during a Pan operation; b) a user performing a zooming-out operation; c) a user moving their hand to a feature of interest before performing a push gesture during a Retrieve operation (Pointer-to-Feature); d) a user performing panning & zooming during a Retrieve operation (Feature-to-Pointer).

Retrieving a spatial feature's attributes is a composite task, consisting of two distinct actions: selection (i.e. mark something as interesting, see Yi et al. (2007)) and elaboration (i.e. show more details about the selected feature, see Yi et al. (2007)). Two mechanisms for Retrieve operations were implemented and evaluated in this work, following the four stages of the research-through-design approach from Zimmerman and Forlizzi (2014), see Figure 2. Both use the same set of core gestures to achieve the Retrieve operation: grabbing (Zoom and Pan) and push (Elaborate), but the two mechanisms rely on two distinct interaction metaphors: selection as moving a dynamic pointer to the desired feature (hence the name Pointer-to-Feature, see Figure 1c) vs. selection as moving the desired feature to a static pointer at the centre of the interaction screen (hence the name Feature-to-Pointer, see Figure 1d). From the point of view of metaphor theory, both strategies rely on the IN-OUT image schema for CONTAINEMENT discussed e.g. in (Hurtienne et al., 2015; Mandler, 1992). In the case of Pointer-to-Feature, the trajector is the pointer, while the trajector is the spatial entity itself for Feature-to-Pointer. From the point of view of gesture execution, Pointer-to-Feature (P2F) relies on the action *pointing* to select spatial features while Feature-to-Pointer uses the action *grab* for feature selection. Hence, from the practical point of view, Feature-to-Pointer (F2P) implies one gesture less to learn in the overall process of realizing a retrieve operation. The contributions of this work are (i) two methods to realize Retrieve during map interaction on large displays, (ii) lessons learned from their evaluation in a lab-based study (N=25), and (iii) an open-source prototype (FreeMapRetrieve) that realizes gesture-based Pan, Zoom and Retrieve operations on large displays.

2 Background

A Retrieve operation builds upon simpler operations such as Select, Zoom and Pan. Overall, there are various proposals of gestures to realize these different operations in the literature, but little consensus beyond single studies about what the most appropriate gestures are, nor guidance about how to reuse the proposed gestures beyond the experimental context.

2.1 Gestures for Select

There is much work investigating gesture-based selection in Human-Computer Interaction (e.g. Walter et al. (2014); Yoo et al. (2015)) but the selection of objects on a map is peculiar for at least three reasons:

- Geographic space is a multi-scale information space: Hence, selection almost always necessitates Pan and Zoom operations in the process of accessing the target of interest (e.g. a lake) on a map.
- Irregularly-shaped targets: The shape of the targets (e.g. countries) differ widely.

STAGE 1: PROBLEM	STAGE 2: DESIGN	STAGE 3: EVALUATION	STAGE 4: REFLECTION
- map interaction goal: procure - map interaction objective: identify - map interaction operator: retrieve	- interaction primitives: explore, abstract, select, elaborate - interaction metaphor: F2P vs P2F - interaction targets: small & large - interactive system: based on Azure Kinect	- success criteria: efficiency, effectiveness, data exploration utility, perceived workload, exploration time - theoretical apparatus: theories of working memory, human processor model, cognitive learning theory	- guidelines for interaction design - lessons learned about theoretical apparatus

Figure 2. Aspects of map interaction and gestural interaction considered at each of the four stages of the research-through-design framework used in the work. Through the study, we learn about viable design alternatives for Retrieve and the effectiveness of the theories considered during the design process.

- Dynamic targets: The position, the size, and sometimes even the symbol (Roth et al., 2011) used to represent the target feature change.

Guiard and Beaudouin-Lafon (2004) proposed that Fitt's law applies to the task of pointing in multi-scale spaces. It is unclear, however, whether this would also hold for irregularly shaped objects. Moreover, pointing is only one subtask of gesture execution, as discussed by Erazo and Pino (2015), who proposed gesture units to predict performance time while doing tasks with hand gestures.

The work by Lin et al. (2019) is perhaps the most related to the current work. They compared four freehand gesture combinations that can be relevant for Retrieve: index/click, index/thrust, palm/click and palm/thrust. The index vs. palm gestures can be used for selection, while the click vs. thrust gestures can be used for elaboration. They reported that the performances of the gesture combinations were dependent on the size of the target. For the large target, the fastest times were for the index/thrust and palm/thrust gestures; for the smallest target, the fastest times were for index/click. Across target sizes, the participants expressed a preference for the index/click and index/thrust gestures. Hence, not one gesture combination stands out as being the most performant in all conditions. Since Retrieve requires a combination of multiple gestures, minimizing the number of motor operations to switch between these is desirable to minimize the participants' fatigue overall. In this work, palm/thrust was chosen as switching between palm and grab (i.e. the gestures used for Zoom/Pan, see Section 2.2) requires fewer gesture strokes than switching from index to grab.

2.2 Gestures for Pan and Zoom

From the interaction viewpoint, panning necessitates two subtasks: selection (of any point of the map) + moving (that point to the east/west/north/south). That is, panning involves a selection activity, albeit a selection with a different purpose than the basic intent of 'marking something as interesting' (Yi et al., 2007) in information visualization. Pan and Zoom have been investigated in non-spatial contexts in several elicitation studies (Wittorf and Jakobsen, 2016; Yoo et al., 2015; Gentile et al., 2019) as well. For the specific task of the interaction with maps, previous work has come up with suggestions almost as varied as the number of studies (see Table 1). In essence, there is currently no consensus as to what gestures are best for panning and zooming on the map in general, let alone for the specific task of Retrieve. In the absence of guidelines to select one of these gestures for a task, we were guided by Jacob and Sibert (1992), who suggested that users typically do not think of zooming and panning as two separate operations, but rather think of them as integral operations in the broader context of locating specific entities on the map. A consequence of this fact for design is that the interaction operations involved in the execution of panning and zooming gestures should make it easy for users to switch between the two and possibly execute them simultaneously. Moreover, the notion of integrality should also be extended to other operations occurring in the Retrieve process, in order to minimise the strokes required to switch to other gestures. To comply with this requirement, we use a grabbing gesture for Pan and Zoom. During a Pan operation, the user grabs the map with one hand and realizes panning through movements in the direction of panning desired. During a Zoom operation, the user grabs the map with two hands and uses movements of their hands in opposite directions to see more/less spatial details. A Pan gesture is also executed if the users move their two grabbing hands in the same direction.

2.3 Gesture selection for Retrieve

With the plethora of options available, we were guided by the principle of gesture stroke optimization, i.e. pick the constituent gestures of Retrieve in such a way that the number of strokes required to switch to other gestures is minimized overall. This leads to the following gestures for the two interaction strategies.

P2F: grab-to-zoom, grab-to-pan, point-to-select (coarse mode), push-to-elaborate, lift-hand-to-change-mode (coarse-to-fine), lower-hand-to-change-mode (fine-to-coarse).

F2P: grab-to-zoom, grab-to-pan, pan-and-zoom-to-select (coarse mode), push-to-elaborate, lift-hand-to-change-mode (coarse-to-fine), lower-hand-to-change-mode (fine-to-coarse).

The formal breakdown of the gestures is shown in Table 2.

3 Prototype design and implementation

The gestures introduced in the previous section were implemented in the FreeMapRetrieve prototype using the Python programming language. Initial tests before implementing the prototype have shown that pyKinectAzure

Table 1. Examples of gestures proposed in previous work, which are relevant to the four atomic operations needed to realize Retrieve (large displays only). This suggests at least 7x7x6x8 = 2352 options to realize the Retrieve gesture, but there is currently little guidance about *viable* combinations. The atomic gestures selected during the work for the performance evaluation are underlined. The selection process is described in Section 2.

Pan	Zoom	Select	Elaborate
1. drawing a line (Siddhpuria et al., 2017) 2. dragging (Nancel et al., 2011) 3. movement from a resting hand to a pointing hand (Fikkert et al., 2009) 4. one-hand grabbing gesture (Boulos et al., 2011) 5. move hand while pointing index (Sluÿters et al., 2023) 6. the joystick gesture (Stellmach et al., 2012) 7. one hand wipe (Bartoschek et al., 2014)	1. full-circle clock/anti-clockwise (Siddhpuria et al., 2017) 2. turning the hand clock/anti-clockwise (Nancel et al., 2011), 3. movement from cupped hands to pointing hands (Fikkert et al., 2009) 4. two-hand grabbing gesture (Boulos et al., 2011) (also in virtual reality, see Newbury et al. (2021)) 5. pinch in/out (Sluÿters et al., 2023) 6. the hand-zoom gesture (Stellmach et al., 2012) 7. two hand spread (Bartoschek et al., 2014)	1. point (Walter et al., 2014; Yoo et al., 2015) 2. swipe (Walter et al., 2014) 3. push (Yoo et al., 2015; Hespanhol et al., 2012) 4. dwell (Hespanhol et al., 2012) 5. drawing a lasso (Hespanhol et al., 2012) 6. grab (Hespanhol et al., 2012)	1. dwell (Walter et al., 2014; Yoo et al., 2015) 2. swipe (Walter et al., 2014) 3. push (Walter et al., 2014) 4. point (Walter et al., 2014) 5. grip (Walter et al., 2014) 6. wave (Walter et al., 2014) 7. AirTap (Vogel and Balakrishnan, 2005) 8. thumb trigger (Vogel and Balakrishnan, 2005)

Table 2. Formal breakdown of the gestures implemented using Roth (2013a), Yi et al. (2007) and Wigdor and Wixon (2011)'s taxonomies. PG = Phase of gesture: R = Registration, C = Continuation, T = Termination. The prototype supports the selection of both large and small polygons through two modes: a coarse mode and a fine mode.

Map operation	Interaction intent	PG	Gesture metaphor	Body Action
Overall use of system		R	Stand in front of screen and point at it	Move body into area of operation and lift at least one hand
		C	⟹ *iterative use of Pan, Zoom and Retrieve operations in no particular order*	
		T	Step away from screen or stop pointing at it	Move out of area of operation or lower both hands
Pan	Explore	R	Grab map with one hand	Make a fist with one pointing hand
		C	Drag map in desired direction	Move hand
		T	Release map	Open hand again
Zoom in	Elaborate	R	Grab map with both hands	Make a fist with both hands pointing at screen
		C	Pull map apart	Increase distance between hands
		T	Release map with both hands	Open both hands again
Zoom out	Abstract	R	Grab map with both hands	Make a fist with both hands pointing at screen
		C	Compress the map	Decrease distance between hands
		T	Release map with both hands	Open both hands again
Retrieve *(Pointer-to-feature)*	Select *coarse mode*	R	— *none* —	
		C	Point at desired feature	Move one arm in the direction of desired feature
		T	— *none* —	
	Select *fine mode, optional*	R	Turn on a switch	Lift up non-pointing hand over shoulder
		C	Point hand up/down/left/right	Move hand relatively up/down/left/right
		T	Turn off a switch	Lower non-pointing hand below shoulder again
	Elaborate	R	Push (as if feature was mid-air in front of user)	Swiftly move pointing hand towards desired feature
		C	— *none* —	
		T	— *none* —	
Retrieve *(Feature-to-pointer)*	Select *coarse mode*	R	— *none* —	
		C	⟹*Iterative combination of Pan and Zoom operations in no particular order*	
		T	— *none* —	
	Select *fine mode, optional*	R	Turn on a switch	Lift up non-panning hand over shoulder
		C	⟹*arbitrary number of Pan operations*	
		T	Turn off a switch	Lower non-pointing hand below shoulder again
	Elaborate	R	Push (as a virtual button mid-air)	Swiftly move one hand in the general direction of its pointer
		C	— *none* —	
		T	— *none* —	

(Fernandez, 2022) can be used reliably. Hence, we used it instead of the native Azure Kinect Sensor and Body Tracking SDKs. Figure 3 shows the architecture of the FreeMapRetrieve and its main components. An Azure Kinect camera is used to observe the space in front of the large display because previous work (e.g. Sosa-León and Schwering (2022); Tölgyessy et al. (2021)) has reported good skeleton tracking performance.

We adapted the open-source model from Takahashi (2020) to infer the hand states. Initial tests have shown that while this model can detect open and closed hand states and distinguish them from each other reliably, the pointing hand state (only index finger extended) is often confused with a closed hand state. As none of the gestures implemented rely on a pointing hand state, we retrained the model for open and closed hand states only. The resulting model was able to differentiate between open and closed hand states with an accuracy of over 98% on a validation dataset (i.e. unseen samples). Figure 3 (bottom) shows the two hand poses distinguished by the prototype. To achieve a good trade-off between robustness and latency, the majority of detected hand states in the latest five camera frames was used to determine whether a Pan/Zoom operation should be registered, continued or terminated. The Azure Kinect camera records 30 frames per second (fps), which implies that max $N = \frac{1000*0.1}{30} \approx 4$ frames are needed to keep the 0.1-second response time (Miller, 1968; Nielsen, 1993; Card et al., 1991) for a feeling of instantaneous system reaction. (5 is slightly above but still maintains the feeling of instantaneousness/direct manipulation).

Once a gesture for a specific map operation is recognized (InteractionController), this gesture is translated into emulated touch screen interactions through the TouchController (Figure 3). We mapped the freehand gestures onto touch screen gestures understood by major map application frameworks (e.g. OpenStreetMap, Leaflet, OpenLayers, Google Maps). Therefore, the prototype can be used to interact with most maps on websites or stand-alone desktop applications. Details on the touchscreen gestures mapped to during the emulation process are available in the supplementary material.

Finally, map applications on the Web (which run in the browser) do not know anything about the gesture elements detected within the Python program. To solve this problem, we created a Chrome web browser extension that communicates with the Python program through a WebSocket connection to indicate the screen location the user is currently pointing at. The browser extension currently supports OpenStreetMap, Google Maps and locally hosted web pages. More pages can be supported simply by adding them to the extension's manifest.json file. Further implementation details are available in the supplementary material (see Section 8). The prototype is agnostic to the main hand used by the user during the interaction, i.e. it is appropriate for right-handed and left-handed users alike.

4 Evaluation

The two techniques to perform Retrieve operations were evaluated in a between-group experiment, to prevent bias due to learning effects. The participants were assigned pseudo-randomly to a condition (P2F or F2P) by strictly alternating the method for every new participant. The users could choose to do the experiment in English or German.

4.1 Procedure

All participants signed an informed consent form before starting the experiment. Then, they filled out a questionnaire asking about personal background information, their frequency of usage of hand gestures to interact with computer systems, their self-assessment of their skills to read and interpret thematic maps as well as their familiarity with the topic of renewable energy in different countries. Afterwards, they performed three tasks:

Task 1 – Training: All participants underwent a training protocol in which they were systematically taught all gestures needed to control a map on the screen in front of them. The training protocol was identical for all participants, only differing in the explanations about how to perform Retrieve operations in the relevant condition. The protocol is available as supplementary material.

Task 2 – Gesture Performance Evaluation: The participants used their assigned Retrieve gesture to find out the population of six German states (three large polygons and three small polygons). They marked the end of the retrieve task by saying the exact population number out loud. This task was audio and video-recorded. The map's zoom level was set to 5 at the beginning of each retrieve task (i.e. the map was zoomed out) to avoid the introduction of a systematic advantage for the P2F condition.

Task 3 – Exploration: To evaluate the gestures' suitability in practical data exploration scenarios, participants performed a video-recorded open-ended think-aloud exercise after a short break. Participants were shown a choropleth world map visualizing indicator 7.2.1 of the UN's sustainable development goals. This is the share of renewable energy used in covering a country's total energy demand. When performing a Retrieve operation on a specific country, participants are presented with a line chart of how the share of renewable energy has developed for this country throughout the last years (see supplementary material).

At the end of the tasks, they filled out questionnaires from the System Usability Scale (SUS) (Brooke, 1995) and a subset of the NASA's Task Load Index (TLX) questionnaire (Hart, 2006) to rate their experience.

4.2 Variables

The following variables were considered during the study:

Independent variables: retrieve mechanism (P2F vs. F2P) and size of polygon (large vs. small).

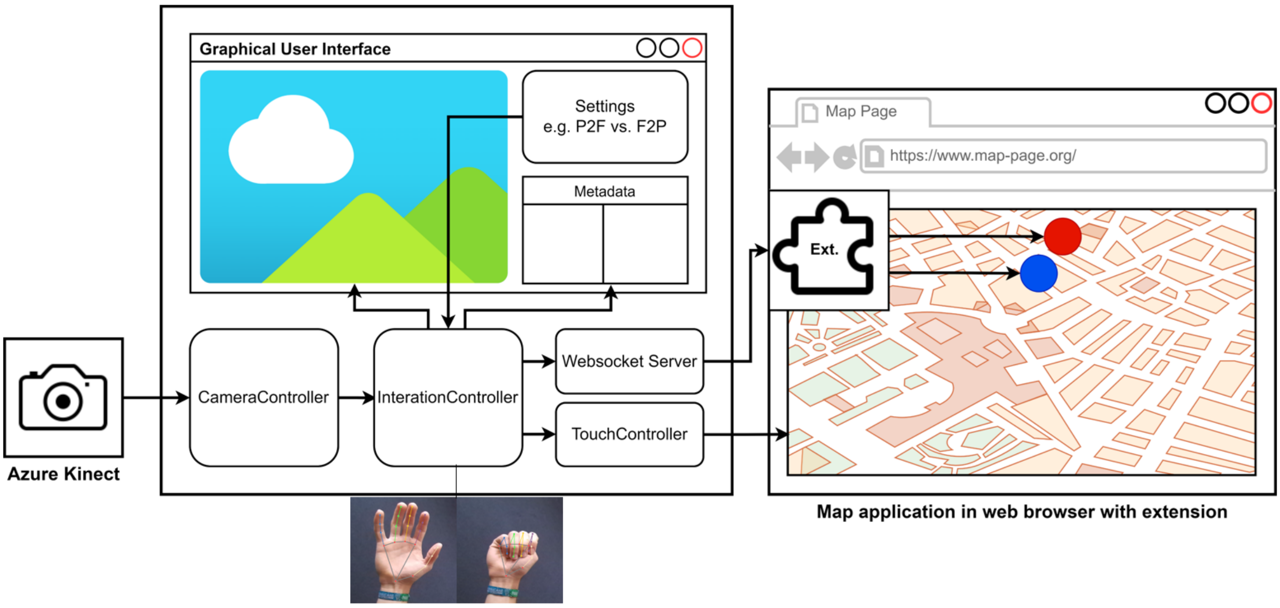


Figure 3. Simplified architecture of the FreeMapRetrieve prototype for gesture-controlled interactions in web-based map applications.

Dependent variables: efficiency (time a participant took to retrieve the feature's attribute and start to say it out loud), effectiveness (number of slips and lapses, usability, data exploration utility (number of insights got, depth of the insights, temporal evolution of insight depth), perceived workload (physical demand, mental demand, frustration) and exploration time (time until a participant claims to receive no further insight).

Control variables (i.e. held constant): screen environment, training on how to use the gestures, data and visualization shown to the participants, initial zoom level (=5) for all retrieve tasks.

Subject variables (i.e. user characteristics): age, gender, experience with gestures, experience with reading and interpreting thematic maps, experience with worldwide distribution of renewable energy.

4.3 Hypotheses

The following hypotheses were formulated before the experiment, based on theories from the literature and the first author's own experience with the gestures during the implementation of the prototype (the reasoning behind each hypothesis is also briefly explained).

Effectiveness. We looked into both slips and lapses. In line with Stanton (2009), lapses are a failure of a human's memory when the user forgets how a particular gesture is performed. Slips are a failure in a user's execution of a gesture when they perform the gesture wrongly, resulting in either not being recognized by the system at all or recognized by the system but performed imprecisely resulting in a Retrieve operation from the wrong polygon.

- **H1a**: For larger polygons, there will not be a significant difference in slips between P2F and F2P, as movements of the pointer during the Elaborate gesture will happen within the polygon itself due to its large size.

- **H1b**: For smaller polygons, the number of slips will be higher for P2F, as the pointer in P2F can easily move into the exterior of the polygon due to its small size.

The interaction tasks take place shortly after the users are taught the gestures for the first time, and hence put some demands on the short-term memory (also called immediate memory) of the users. Here, we build on findings from previous work (Miller, 1956; Baddeley, 1994) that information is stored in short-term memory using chunks (i.e. coherent, meaningful units in mind). Lapse errors will become more frequent after the maximum chunk capacity has been reached. Since we are not aware of HCI or GIScience work that informs about the limits in the context of gestural interaction, we turned to findings from Psychology as a starting point. Miller (1956) originally proposed 7 ∓ 2 and follow-up studies (Cowan, 2001) proposed 4 ∓ 1 (and even 2, see Gobet and Clarkson (2004)). We used 4 ∓ 1 at this point because Cowan (2001) documents a broad range of observations consistent with the limit of four entities. The 4 ∓ 1 guideline applies to *independent* chunks, hence, we can expect lapses to occur more frequently beyond 4-5 distinct, independent gestures. The number of distinct gestures involved in the current experiment is six, which is within the limit range. Since independence does not hold in this case (i.e. there are similarities in the way of executing the gestures), the number of lapses may be minimal overall in both conditions. Nonetheless, in case lapses do occur:

- **H1c**: Lapses will be more frequent during pointing or panning in fine mode than in coarse mode (the fine pointing mode demands more gestures to be executed and this increases the probability of forgetting some of those that were just learned).

Efficiency. We used the human processor model (Card et al., 1986) to estimate the cognitive, perceptual, and motor operations involved in the retrieve task. This revealed that only half the motor operations are necessary for P2F compared to F2P. Thus:

- **H2a**: Efficiency will be higher for P2F than F2P for large polygons.

- **H2b**: For small polygons, however, participants using F2P will be more efficient in retrieving the correct polygon's attributes, as both the pointer and the map remain stationary once the polygon to retrieve from is positioned correctly.

Usability. The SUS score will be higher for P2F because users are very likely familiar with this interaction concept. In Desktop environments, Retrieve operations have a similar modus operandi: place the mouse cursor on top of a feature to select it, then click on the feature to get additional information (**H3**).

Perceived Workload. The users' perceived mental and physical demands and frustration level while completing the task were all measured on a seven-point Likert scale.

- **H4a**: Mental demand will be lower among participants using P2F, as users are likely already familiar with it from Desktop environments, while the concept of F2P is uncommon in Desktop environments.

- **H4b**: Physical demand will be higher among participants using F2P, as about twice the motor operations are involved in selecting a polygon before performing the Elaborate gesture on it.

- **H4c**: Frustration will be higher among participants using P2F, stemming from involuntary pointer movements while performing the Elaborate gesture and therefore retrieving from the wrong polygon.

Data Exploration Utility. Data exploration can be conceptualized as an activity where users *learn* about the data. This enables borrowing insight from cognitive load theory (CLT) to formulate the hypotheses. CLT foresees three types of loads during learning (Debue and Van De Leemput, 2014): intrinsic (complexity of the material to learn; it also depends on the user's expertise), extraneous (complexity of the instruction format), and germane (how much the user actually invests in learning; it is related to their motivation to some extent). Assuming that intrinsic and germane loads are similar for the two conditions, the difference must come from the extraneous load.

- **H5a**: The overall time spent during the exploration will be smaller for F2P than for P2F, as a corollary of H4a and H4b. If F2P yields higher mental and physical demands, this will shrink the cognitive resources available for the data exploration itself and may lead to participants spending less time on the activity.

- **H5b**: The depth of insight will be comparable in both conditions, as this relates to other aspects, most notably, the participants' background knowledge.

Like in Saraiya et al. (2005), insights are assigned a domain value to measure the deepness of an insight a participant has received. The domain value of an insight is a score between one and five, which depends on the complexity of the insight. The sum of all domain values is then a measure of the deepness of insight into the dataset a participant has achieved. Further details are provided in the supplementary material.

4.4 Participants

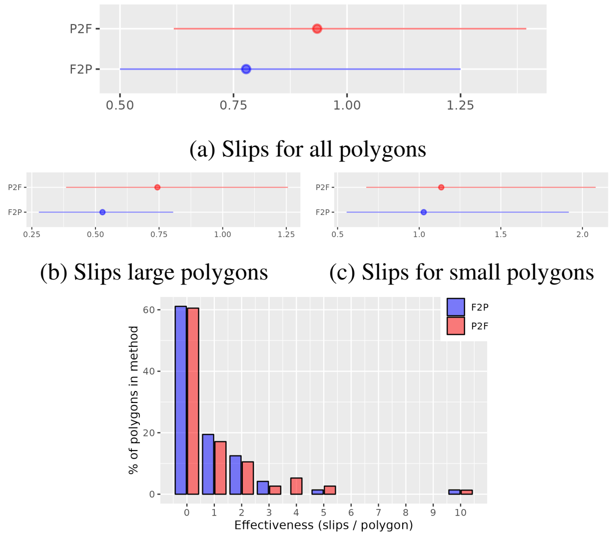
Twenty-six participants, recruited through personal contacts and word of mouth, participated in the experiment. One participant's observations were excluded from the analysis because they disclosed at the end of the study that they were unable to properly form a fist with their hand due to anatomical deformations in their thumb, resulting in imprecise Pan and Zoom operations. Another participant initially had problems keeping their body within the field of view of the Azure Kinect camera. Thus, two measurements from task two when retrieving from large polygons were excluded as this can be seen as an error in measurement. The mean age of the participants included in the analysis (13 Female, 12 Male) was 26.9 (sd: 9.9) and the median was 24. Snacks and refreshments were provided to participants during the study between the tasks. Only one participant reported using hand gestures less than monthly to interact with computer systems. All others reported to never use them. The study was pilot-tested and approved by the institutional ethics board.

5 Results

We used an estimation approach (Dragicevic, 2016), with mean point estimates and confidence intervals (CIs) to compare the two conditions. All CIs were calculated using the bootES package (Kirby and Gerlanc, 2013) with $R = 5000$ bootstrapped resamples. Confidence intervals, which do not overlap indicate statistical significance.

5.1 Effectiveness

On average, there were 0.934 (95%-CI: [0.62, 1.40]) slips per polygon for P2F (Figure 4a) and 0.778 (95%-CI: [0.5, 1.25]) slips per polygon for F2P. For large polygons, the average slips per polygon was 0.744 (95%-CI: [0.39, 1.26]) for P2F and 0.528 (95%-CI: [0.28, 0.81]) for F2P (Figure 4b). Hence, both conditions were comparable as anticipated (H1a). For small polygons, there were on average 1.135 slips [0.68, 2.05] per polygon (P2F) and 1.028 slips [0.56, 1.89] per polygon (F2P) respectively (Figure 4c). Hence, there is not enough evidence to support H1b. Figure 4d shows the distribution of slips in the two conditions.



(a) Slips for all polygons

(b) Slips large polygons

(c) Slips for small polygons

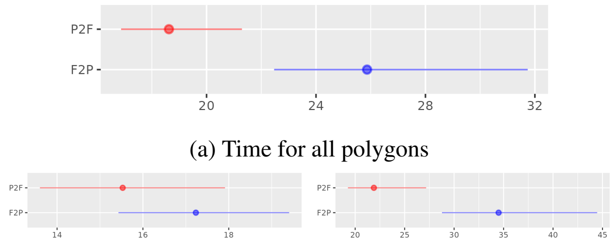
(d) Histogram for effectiveness (slips per polygon), normalized to 100%

Figure 4. Mean effectiveness measured in slips per polygon.

Overall, there were almost no lapses during the experiment as anticipated. There was only one instance per condition where this happened: one participant using P2F forgot how to properly execute the Retrieve gesture. A participant using F2P for retrieving a feature's attributes forgot that zooming into the map was an option to facilitate the Selection part of the Retrieve gesture.

5.2 Efficiency

When considering all polygon sizes, the time-on-task for a Retrieve operation was on average 18.62 (95%-CI: [16.87, 21.29]) seconds for P2F and 25.86 (95%-CI: [22.47, 31.74]) seconds for F2P (Figure 5a). Looking at large polygons only, the average was 15.53 (95%-CI: [13.6, 17.91]) seconds for P2F and 17.23 (95%-CI: [15.43, 19.41]) seconds for F2P ((Figure 5b). This indicates that H2a is not supported. If only small polygons are considered, participants took on average 21.89 (95%-CI: [19.27, 27.17]) seconds in the P2F condition, and 34.5 (95%-CI: [28.77, 44.46]) seconds in F2P (Figure 5c). For small polygons, the average bootstrapped difference of means is 12.609 seconds with a 95%-CI of [5.64, 22.93] seconds. Hence, there is evidence that P2F was more efficient than F2P, in the opposite direction of what was anticipated (H2b).



(a) Time for all polygons

(b) Time for large polygons

(c) Time for small polygons

Figure 5. Mean efficiency (seconds) to retrieve from a polygon.

5.3 Usability

The mean SUS score for participants using the P2F mechanism was 77.5 (95%-CI: [70.58, 83.16]). For participants using the F2P mechanism, the mean score was 78.54 (95%-CI: [70.34, 82.5]). A SUS score in the range of 71-85 indicates that the users perceived the overall interaction experience as "good" according to Bangor et al. (2008).

5.4 Perceived workload

Figure 6 shows the results for perceived workload. The mean score for the physical load was 47.69 (95%-CI: [35.38, 58.46]) for P2F and 43.33 (95%-CI: [31.67, 54.17]) for F2P. For mental load, the mean score was 49.23 (95%-CI: [37.69, 59.23]) for P2F and 41.67 (95%-CI: [31.67, 50.0]) for F2P. Lastly, the mean frustration was 39.23 (95%-CI: [28.46, 50.0]) for P2F and 37.5 (95%-CI: [26.67, 50.83]) for F2P. There is still no definite answer to the question 'what is an acceptable workload value?' in the literature. Hertzum (2021) reported means of 49 (mental demand), 30 (physical demand) and 40 (frustration) across 127 studies for general activities (i.e. activities not tied to a specific domain). Taking these as benchmark values suggests that the workload of working with the FreeMapRetrieve prototype is 'within the norm'. The deviation of the physical load from that benchmark value is consistent with the nature of the interaction.

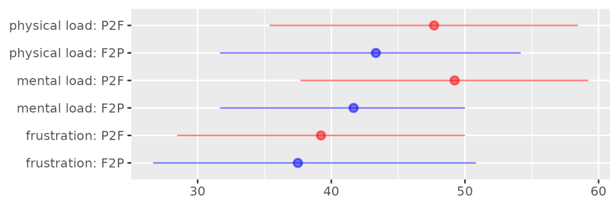


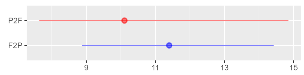
Figure 6. Mean physical and mental load and frustration.

5.5 Data exploration utility

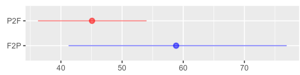
We could extract 250 insight statements (i.e. 19 statements per user on average, P2F) and 270 insight statements (i.e. 23 statements per user on average, F2P) from the video analysis. A good deal (70%) was classified as trivial, about

20% were intermediate and about 10% were of high domain value (Figure 7a). The two conditions were comparable as H5b anticipated. The mean exploration time was slightly higher in the F2P condition (11 mins vs 10 mins), but because the confidence intervals overlap Figure 7b), the two conditions can be considered comparable. Thus, there is no support for H5a, but this is consistent with the idea that H5a is a consequence of H4a and H4b. The domain value scores were slightly higher in F2P (Figures 7c and 7e) and this can be traced back to the contributions of four participants from the F2P condition (Figure 7d).

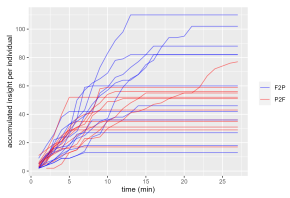
(a) Frequencies



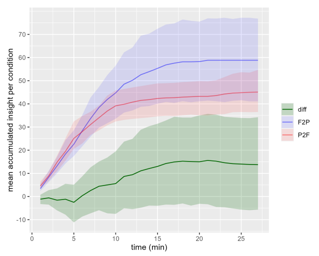
(b) Mean total exploration time



(c) Mean total insight score



(d) Patterns of exploration evolution per individual



(e) Patterns of exploration evolution per condition

Figure 7. Data exploration utility.

5.6 Impact of participants' background

We performed an effects analysis of the participant's gender, previous experience with thematic maps, and familiarity with the renewable energy topic on the efficiency/effectiveness results. There was no evidence of an effect, which indicates that the two mechanisms are agnostic to the participants' background, i.e. they are equally good for participants irrespective of the factors analyzed.

6 Discussion

Overall, the evaluation of the prototype has shown good results w.r.t. usability and workload. In addition, the number of slips was relatively low with roughly 80% of the participants completing the tasks with 1 slip or less. Only a neglectable number of lapses was observed in both conditions as mentioned above. Finally, while there is no benchmark to compare the data exploration utility observations, there is evidence that both mechanisms support data exploration, with about 20 insight-based statements in each condition. All these observations validate the two designs (P2F and F2P) and suggest that the combination of gestures are *viable* for interfaces attempting to enable gesture-based interaction for Retrieve w.r.t. maps on large displays. As Table 1 suggests, means to systematically unveil these viable design paths are still needed in map-based gesture interaction research. We contend that gesture performance evaluation studies, along with benchmarking against thresholds (e.g. SUS values, perceived workload values), as done in this work can be used to that end. For the entire gesture performance evaluation framework, its stages and the relevant dimensions to consider at each stage, see Figure 2. Next, we reflect on the design process, articulate guidelines and point to opportunities for future work. For the design guidelines, we provide heuristic-based and counterfactual-based formulations. A counterfactual proposition is of the form: 'If design was D then interaction would be I' (Oulasvirta and Hornbæk, 2021). It makes immediately clear to the reader what to expect when the heuristics are not followed.

6.1 Reflections about system and interaction design

One lesson learned is that a simple task such as Retrieve necessitates rethinking 'successfulness' or reported user preferences from elicitation studies for single gestures examined as atomic units. We have discussed in Section 2 that Retrieve = Select + Elaborate, while Pan/Zoom = Select + Move. It would not have made sense to use the same gesture for selection in both cases. We used palm and grab as two different gestures for the two different selection operations, even though dwelling (Hespanhol et al., 2012) or index (Lin et al., 2019) were rated highest in previous work. The guiding principle here, called the rule of gesture strokes optimization (Section 2.3), seems absent from even the most recent guidelines for gesture vocabulary design (Xia et al., 2022). It may be listed under complexity ('The state or quality of a gesture being intricate or complicated', see Xia et al. (2022)), but this would be a different notion of complexity (i.e. global complexity for all gestures considered instead of 'local' complexity for a single gesture). Gesture stroke optimization could be introduced as a new dimension belonging to the physical factors to consider during gesture design (next to complexity, efficiency, ergonomics and occlusion, see Xia et al. (2022)). Besides, similar projects could find the two heuristics below for user activity detection on maps, which has been identified as an important requirement for intelligent maps (Degbelo and Kray, 2018; Degbelo et al., 2023):

- G1: Classify the hand state based on the N latest video frames to increase classification accuracy,

where N should be selected to maintain the feeling of instantaneous response times (upper limit of 0.1 seconds according to Nielsen (1993)). For instance, use $2 <= N <= 5$ when recording at 30 fps.
If N=1, the accuracy would drop to the accuracy of the hand state classifier (in the case of the current prototype, this accuracy was 98%, which implies about 1 false classification every 2 seconds); if N >= 8, then the feeling of direct manipulation would disappear (N = 8 corresponds to ≈ 0.2 *seconds).*

- O1: consider gesture stroke optimization across all gestures involved in the target task, in the process of designing Retrieve gestures. This is a way of acknowledging that sub-gestures are *integral* operations in the context of the broader common goal.
If gesture stroke optimization is not taken into account, then the efficiency of the whole gesture set would be compromised. (The extent to which that efficiency is compromised remains to be characterized empirically, hence it is mentioned here as an opportunity for further work).

Third, both Pointer-to-Feature and Feature-to-Pointer have achieved good performance during the study. Hence, the gestures chosen are good candidates for Retrieve.

- G2: Use Pointer-to-Feature as Retrieve strategy when there are many small geographic entities to explore.
If F2P is used as a strategy with small polygons, then users would become 50% slower (Figure 5c).

- O2: Use P2F or F2F as Retrieve strategy when there are many large polygons to explore.
If P2F or F2P would not be used as strategies, then the performance of the alternative strategies needs to be assessed empirically.

6.2 Reflections about the theoretical apparatus

As for effectiveness, both conditions were comparable as anticipated w.r.t. slips for large polygons (H1a) and lapses (H1c). Since lapses were predicted based on theories of short-term memory, these may be said to have stood the test for now:

- G3: Use theories of short-term memory (in particular the 4 ∓ 1 rule) for prediction of lapses during the design of gesture-based interaction for maps.

The guiding principle for the predictions of efficiency values was the relative number of motor operations (Card et al., 1986), as estimated from the human processor model. Both H2a and H2b were not supported (with H2b going in the opposite direction than anticipated), hence:

- O3: More work on models to estimate the time for gesture interaction with maps on large displays is needed to advance the science of map interaction. The Human-Processor-Model (Card et al., 1986; Jastrzembski and Charness, 2007) and the Gesture Unit Model (Erazo and Pino, 2015, 2018) did not lead to realistic estimates in this study.

The usability values did not differ as anticipated (H3). A difference would have been an indication of a similar effect to legacy biases observed during elicitation studies (Morris et al., 2014), where previous experience with a type of technology affects user outcomes (i.e. the type of gestures produced). There does not seem to be such 'legacy effect' from the observations in the current study. Finally, the results for the perceived workload (H4a, H4b and H4c) and data exploration utility (H5a and H5b) are consistent with the observations related to H2b and H1c and reinforce what has been said above regarding G3 and O3. A clear statement about CLT cannot be made at this point based on the observations.

6.3 Limitations

One limitation of the study is that the scope of the results applies only to able-bodies participants. The need to exclude one participant's data from the statistical analysis highlights how even small anatomical deformations in the hand can impact a person's ability to perform the gestures properly and the need for different gestures to help that user group perform Retrieve operations successfully. Besides, the study was conducted as a between-group experiment. Despite the care taken to systematically assign the participants randomly to the conditions, the impact of the differences in profiles between the participants of the two groups cannot be entirely ruled out. This being said, we used the values from the background questionnaires (i.e. similarity with maps, familiarity with the renewable energy topic), and treated them as categorical variables to perform a head-to-head comparison between 'similar' users from the two conditions. The results indicated a similar tendency to the ones reported in Section 5.

7 Conclusion and future work

We have presented Pointer-and-Feature and Feature-to-Pointer, two basic mechanisms to realize gesture-based Retrieve operations for maps on large displays, and the lessons learned from their design, implementation and evaluation. We found that both mechanisms work well for Retrieve operations on large polygons and hence can be used as a starting point for more complex tasks (e.g. simultaneous selection of multiple spatial entities for comparison purposes). We also learned that the Pointer-to-Feature mechanism should be preferred if the geographic dataset involves the exploration of a large number of small polygons. The two mechanisms are robust against participants' background, i.e. they perform equally well irrespective of participants' previous experience with thematic maps and the topic of exploration at hand.

The spatial entities explored in this work were all represented as polygons. Hence, one immediate direction for future work is the inclusion of more types of symbols representing geographic entities: points (as generalized polygons) and lines (rivers, roads), to assess the performance of the two techniques in these scenarios. Another direction for future work includes the extension of the open-source prototype with a configuration panel where designers can select different options for sub-gestures relevant to a map interaction goal, and chain them as they like during prototyping and testing. This would facilitate the testing of alternative possibilities (e.g. hover to select, popup menus after Retrieve operations) and the discovery of additional viable design paths for gesture-based interaction with maps on large displays. At last, the design of effective gesture vocabularies for collaborative interaction on maps with large displays would be a follow-up task to this work, which poses interesting challenges for future work.

8 Data and software availability

The code of FreeMapRetrieve and the scripts for the analysis of the user study data are available at https://github.com/jonas-hurst/MapGestureController and https://github.com/jonas-hurst/FreeMapRetrieve-Statistic respectively. A demo, the stimuli used during the experiment, the training protocol and a detailed description of the components of FreeMapRetrieve are available at https://doi.org/10.6084/m9.figshare.24077664.

Acknowledgements

We thank Samuel Navas Medrano and Violeta Sosa-León for their feedback during the work. Auriol Degbelo was funded by the German Research Foundation through the project NFDI4Earth (DFG project no. 460036893, https://www.nfdi4earth.de/) within the German National Research Data Infrastructure (NFDI, https://www.nfdi.de/).

References

Austin, C. R., Ens, B., Satriadi, K. A., and Jenny, B.: Elicitation study investigating hand and foot gesture interaction for immersive maps in augmented reality, Cartography and Geographic Information Science, 47, 214–228, https://doi.org/10.1080/15230406.2019.1696232, 2020.

Baddeley, A.: The magical number seven: Still magic after all these years?, Psychological Review, 101, 353–356, https://doi.org/10.1037/0033-295X.101.2.353, 1994.

Bangor, A., Kortum, P. T., and Miller, J. T.: An empirical evaluation of the system usability scale, International Journal of Human-Computer Interaction, 24, 574–594, https://doi.org/10.1080/10447310802205776, 2008.

Bartoschek, T., Pape, G., Kray, C., Jones, J., and Kauppinen, T.: Gestural interaction with spatiotemporal linked open data, OSGeo Journal, 13, 60–67, https://doi.org/10.7275/R57D2SBM, 2014.

Boulos, M. N. K., Blanchard, B. J., Walker, C., Montero, J., Tripathy, A., and Gutierrez-Osuna, R.: Web GIS in practice X: a Microsoft Kinect natural user interface for Google Earth navigation, International journal of health geographics, 10, 45, https://doi.org/10.1186/1476-072X-10-45, 2011.

Brooke, J.: SUS - A quick and dirty usability scale, Usability Evaluation in Industry, 189, 4–7, https://doi.org/10.1002/hbm.20701, 1995.

Card, S., Moran, T., and Newell, A.: The model human processor - An engineering model of human performance, in: Handbook of Perception and Human Performance. Vol. 2: Cognitive Processes and Performance, edited by Boff, K. R., Kaufman, L., and James, P. T., pp. 1–35, book: Handbook of perception and human performance. Issue: 45–1, 1986.

Card, S. K., Robertson, G. G., and Mackinlay, J. D.: The information visualizer, an information workspace, in: Proceedings of the SIGCHI conference on Human factors in computing systems Reaching through technology - CHI '91, edited by Robertson, S. P., Olson, G. M., and Olson, J. S., pp. 181–186, ACM Press, New Orleans, Louisiana, USA, https://doi.org/10.1145/108844.108874, 1991.

Cowan, N.: The magical number 4 in short-term memory: A reconsideration of mental storage capacity, Behavioral and Brain Sciences, 24, 87–114, https://doi.org/10.1017/S0140525X01003922, 2001.

Debue, N. and Van De Leemput, C.: What does germane load mean? An empirical contribution to the cognitive load theory, Frontiers in Psychology, 5, https://doi.org/10.3389/fpsyg.2014.01099, 2014.

Degbelo, A.: FAIR geovisualizations: definitions, challenges, and the road ahead, International Journal of Geographical Information Science, 36, 1059–1099, https://doi.org/10.1080/13658816.2021.1983579, 2022.

Degbelo, A. and Kray, C.: Intelligent geovisualizations for open government data (vision paper), in: 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, edited by Banaei-Kashani, F., Hoel, E. G., Güting, R. H., Tamassia, R., and Xiong, L., pp. 77–80, ACM Press, Seattle, Washington, USA, https://doi.org/10.1145/3274895.3274940, 2018.

Degbelo, A., Schmidt, B., Henzen, C., Lechler, S., Lubahn, B., and Zander, F.: Desiderata for intelligent maps: A multiperspective compilation, KN - Journal of Cartography and Geographic Information, 73, 183–198, https://doi.org/10.1007/s42489-023-00142-w, 2023.

Dragicevic, P.: Fair statistical communication in HCI, in: Modern Statistical Methods for HCI, edited by Robertson, J. and Kaptein, M., pp. 291–330, Springer, Cham, https://doi.org/10.1007/978-3-319-26633-6_13, 2016.

Du, G., Degbelo, A., and Kray, C.: User-generated gestures for voting and commenting on immersive displays in urban planning, Multimodal Technologies and Interaction, 3, https://doi.org/10.3390/mti3020031, 2019.

Edsall, R., Andrienko, G., Andrienko, N., and Buttenfield, B.: Interactive maps for exploring spatial data, Manual of Geographic Information Systems, pp. 837–858, publisher: ASPRS, Bethesda, MD, 2009.

Erazo, O. and Pino, J. A.: Predicting task execution time on natural user interfaces based on touchless hand gestures, in: Proceedings of the 20th International Conference on Intelligent User Interfaces (IUI 2015), edited by Brdiczka, O., Chau, P., Carenini, G., Pan, S., and Kristensson, P. O., pp. 97–109, ACM, Atlanta, Georgia, USA, https://doi.org/10.1145/2678025.2701394, 2015.

Erazo, O. and Pino, J. A.: Predicting user performance time for hand gesture interfaces, International Journal of Industrial Ergonomics, 65, 122–138, https://doi.org/10.1016/j.ergon.2017.07.010, 2018.

Fernandez, I. G.: pyKinectAzure, https://github.com/ibaiGorordo/pyKinectAzure, accessed: September 10, 2023., 2022.

Fikkert, W., van der Vet, P. E., van der Veer, G. C., and Nijholt, A.: Gestures for large display control, in: Gesture in Embodied Communication and Human-Computer Interaction, 8th International Gesture Workshop (GW 2009), edited by Kopp, S. and Wachsmuth, I., vol. 5934 of *Lecture notes in computer science*, pp. 245–256, Springer, Bielefeld, Germany, https://doi.org/10.1007/978-3-642-12553-9_22, 2009.

Gentile, V., Fundarò, D., and Sorce, S.: Elicitation and evaluation of zoom gestures for touchless interaction with desktop displays, in: Proceedings of the 8th ACM International Symposium on Pervasive Displays, edited by Khamis, M., Sorce, S., Cauchard, J. R., and Gentile, V., ACM Digital Library, pp. 1–7, Association for Computing Machinery, New York, NY, USA, https://doi.org/10.1145/3321335.3324934, 2019.

Gobet, F. and Clarkson, G.: Chunks in expert memory: Evidence for the magical number four … or is it two?, Memory, 12, 732–747, https://doi.org/10.1080/09658210344000530, 2004.

Griffin, A. L.: Trustworthy maps, Journal of Spatial Information Science, https://doi.org/10.5311/JOSIS.2020.20.654, 2020.

Guiard, Y. and Beaudouin-Lafon, M.: Target acquisition in multiscale electronic worlds, International Journal of Human-Computer Studies, 61, 875–905, https://doi.org/10.1016/j.ijhcs.2004.09.005, 2004.

Hart, Sandra, G.: NASA-task load index (NASA-TLX); 20 years later, in: Human Factors and Ergonomics Society Anual Meting, pp. 904–908, San Francisco, California, USA, https://doi.org/10.1037/e577632012-009, 2006.

Hatscher, B., Luz, M., Nacke, L. E., Elkmann, N., Müller, V., and Hansen, C.: GazeTap: towards hands-free interaction in the operating room, in: Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI 2017), edited by Lank, E., Vinciarelli, A., Hoggan, E. E., Subramanian, S., and Brewster, S. A., pp. 243–251, ACM, Glasgow, United Kingdom, https://doi.org/10.1145/3136755.3136759, 2017.

Hertzum, M.: Reference values and subscale patterns for the task load index (TLX): a meta-analytic review, Ergonomics, 64, 869–878, https://doi.org/10.1080/00140139.2021.1876927, 2021.

Hespanhol, L., Tomitsch, M., Grace, K., Collins, A., and Kay, J.: Investigating intuitiveness and effectiveness of gestures for free spatial interaction with large displays, in: The International Symposium on Pervasive Displays (PerDis 2012), edited by José, R. and Huang, E. M., p. 6, ACM, Porto, Portugal, https://doi.org/10.1145/2307798.2307804, 2012.

Hurtienne, J., Klockner, K., Diefenbach, S., Nass, C., and Maier, A.: Designing with image schemas: resolving the tension between innovation, inclusion and intuitive use, Interacting with Computers, 27, 235–255, https://doi.org/10.1093/iwc/iwu049, 2015.

Jacob, R. J. K. and Sibert, L. E.: The perceptual structure of multidimensional input device selection, in: Conference on Human Factors in Computing Systems (CHI 1992), edited by Bauersfeld, P., Bennett, J., and Lynch, G., pp. 211–218, ACM, Monterey, California, USA, https://doi.org/10.1145/142750.142792, 1992.

Jastrzembski, T. S. and Charness, N.: The model human processor and the older adult: parameter estimation and validation within a mobile phone task, Journal of Experimental Psychology: Applied, 13, 224–248, https://doi.org/10.1037/1076-898X.13.4.224, 2007.

Kirby, K. N. and Gerlanc, D.: BootES: An R package for bootstrap confidence intervals on effect sizes, Behavior Research Methods, 45, 905–927, https://doi.org/10.3758/s13428-013-0330-5, 2013.

Kray, C., Schmid, F., and Fritze, H.: Guest editorial: map interaction, GeoInformatica, 21, 573–576, https://doi.org/10.1007/s10707-016-0290-x, 2017.

Lin, J., Harris-Adamson, C., and Rempel, D.: The design of hand gestures for selecting virtual objects, International Journal of Human–Computer Interaction, 35, 1729–1735, https://doi.org/10.1080/10447318.2019.1571783, 2019.

Mandler, J. M.: How to build a baby: II. Conceptual primitives., Psychological Review, 99, 587–604, https://doi.org/10.1037/0033-295X.99.4.587, 1992.

Miller, G. A.: The magical number seven, plus or minus two: some limits on our capacity for processing information, Psychological Review, 63, 81–97, https://doi.org/10.1037/h0043158, 1956.

Miller, R. B.: Response time in man-computer conversational transactions, in: AFIPS'68 (Fall, part I) - Proceedings of the December 9-11, 1968, fall joint computer conference, part I, pp. 267–277, ACM Press, San Francisco, California, USA, https://doi.org/10.1145/1476589.1476628, 1968.

Morris, M. R., Danielescu, A., Drucker, S., Fisher, D., Lee, B., Schraefel, M. C., and Wobbrock, J. O.: Reducing legacy bias in gesture elicitation studies, Interactions, 21, 40–45, https://doi.org/10.1145/2591689, 2014.

Nancel, M., Wagner, J., Pietriga, E., Chapuis, O., and Mackay, W.: Mid-air pan-and-zoom on wall-sized displays, in: Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11, edited by Tan, D. S., Amershi, S., Begole, B., Kellogg, W. A., and Tungare, M., pp. 177–186, ACM Press, Vancouver, Canada, https://doi.org/10.1145/1978942.1978969, 2011.

Newbury, R., Satriadi, K. A., Bolton, J., Liu, J., Cordeil, M., Prouzeau, A., and Jenny, B.: Embodied gesture interaction for immersive maps, Cartography and Geographic Information Science, 48, 417–431, https://doi.org/10.1080/15230406.2021.1929492, 2021.

Nielsen, J.: Response times: the 3 important limits, https://www.nngroup.com/articles/response-times-3-important-limits/, accessed: September 03, 2023, 1993.

Oulasvirta, A. and Hornbæk, K.: Counterfactual thinking: what theories do in design, International Journal of Human–Computer Interaction, pp. 1–15, https://doi.org/10.1080/10447318.2021.1925436, 2021.

Pike, W. A., Stasko, J., Chang, R., and O'Connell, T. A.: The science of interaction, Information Visualization, 8, 263–274, https://doi.org/10.1057/ivs.2009.22, 2009.

Rinner, C. and Bird, M.: Evaluating community engagement through argumentation maps - A public participation GIS case study, Environment and Planning B: Planning and Design, 36, 588–601, https://doi.org/10.1068/b34084, 2009.

Roth, R. E.: An empirically-derived taxonomy of interaction primitives for interactive cartography and geovisualization, IEEE Transactions on Visualization and Computer Graphics, 19, 2356–2365, https://doi.org/10.1109/TVCG.2013.130, 2013a.

Roth, R. E.: Interactive maps: What we know and what we need to know, Journal of Spatial Information Science, 6, 59–115, https://doi.org/10.5311/JOSIS.2013.6.105, 2013b.

Roth, R. E., Brewer, C. A., and Stryker, M. S.: A typology of operators for maintaining legible map designs at multiple scales, Cartographic Perspectives, pp. 29–64, https://doi.org/10.14714/CP68.7, 2011.

Saraiya, P., North, C., and Duca, K.: An insight-based methodology for evaluating bioinformatics visualizations, IEEE Transactions on Visualization and Computer Graphics, 11, 443–456, https://doi.org/10.1109/TVCG.2005.53, 2005.

Schwering, A., Krukar, J., Seep, J., and Qamaz, Y.: Individualization in spatial behaviour and map reading, AGILE: GIScience Series, 4, 1–5, https://doi.org/10.5194/agile-giss-4-41-2023, 2023.

Siddhpuria, S., Katsuragawa, K., Wallace, J. R., and Lank, E.: Exploring at-your-side gestural interaction for ubiquitous environments, in: Proceedings of the 2017 Conference on Designing Interactive Systems (DIS'17), edited by Mival, O. H., Smyth, M., and Dalsgaard, P., pp. 1111–1122, ACM, Edinburgh, Scotland, UK, https://doi.org/10.1145/3064663.3064695, 2017.

Sluÿters, A., Sellier, Q., Vanderdonckt, J., Parthiban, V., and Maes, P.: Consistent, continuous, and customizable mid-air gesture interaction for browsing multimedia objects on large displays, International Journal of Human–Computer Interaction, 39, 2492–2523, https://doi.org/10.1080/10447318.2022.2078464, 2023.

Sosa-León, V. A. L. and Schwering, A.: Evaluating automatic body orientation detection for indoor location from skeleton tracking data to detect socially occupied spaces using the Kinect v2, Azure Kinect and Zed 2i, Sensors, 22, 3798, https://doi.org/10.3390/s22103798, 2022.

Stanton, N.: Human-error identification in human–computer interaction, in: Human- Computer Interaction Fundamentals, edited by Sears, A. and Jacko, J. A., pp. 123–134, Taylor & Francis, 2009.

Stellmach, S., Jüttner, M., Nywelt, C., Schneider, J., and Dachselt, R.: Investigating Freehand Pan and Zoom, in: Mensch & Computer 2012, edited by Reiterer, H. and Deussen, O., pp. 303–312, Oldenbourg Wissenschaftsverlag and Oldenbourg, München, https://doi.org/10.1524/9783486718782.303, 2012.

Takahashi, S.: hand-gesture-recognition-using-mediapipe, https://github.com/Kazuhito00/hand-gesture-recognition-using-mediapipe, accessed: September 02, 2023, 2020.

Tölgyessy, M., Dekan, M., and Chovanec, L.: Skeleton tracking accuracy and precision evaluation of Kinect V1, Kinect V2, and the Azure Kinect, Applied Sciences, 11, 5756, https://doi.org/10.3390/app11125756, 2021.

Vogel, D. and Balakrishnan, R.: Distant freehand pointing and clicking on very large, high resolution displays, in: Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology (UIST'05), edited by Baudisch, P., Czerwinski, M., and Olsen, D. R., pp. 33–42, ACM, Seattle, Washington, USA, https://doi.org/10.1145/1095034.1095041, 2005.

Walter, R., Bailly, G., Valkanova, N., and Müller, J.: Cuenesics: using mid-air gestures to select items on interactive public displays, in: Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services - MobileHCI '14, pp. 299–308, ACM Press, Toronto, Canada, https://doi.org/10.1145/2628363.2628368, 2014.

Wigdor, D. and Wixon, D.: Brave NUI world: Designing natural user interfaces for touch and gesture, Morgan Kaufmann/Elsevier, Burlington, MA, 2011.

Wittorf, M. L. and Jakobsen, M. R.: Eliciting mid-air gestures for wall-display interaction, in: Proceedings of the 9th nordic conference on human-computer interaction, gothenburg, sweden, october 23 - 27, 2016, p. 3, ACM, https://doi.org/10.1145/2971485.2971503, 2016.

Xia, H., Glueck, M., Annett, M., Wang, M., and Wigdor, D.: Iteratively designing gesture vocabularies: A survey and analysis of best practices in the HCI literature, ACM Transactions on Computer-Human Interaction, 29, 1–54, https://doi.org/10.1145/3503537, 2022.

Yi, J. S., Kang, Y. a., Stasko, J., and Jacko, J.: Toward a deeper understanding of the role of interaction in information visualization, IEEE Transactions on Visualization and Computer Graphics, 13, 1224–1231, https://doi.org/10.1109/TVCG.2007.70515, 2007.

Yoo, S., Parker, C., Kay, J., and Tomitsch, M.: To dwell or not to dwell: An evaluation of mid-air gestures for large information displays, in: Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction (OZCHI 2015), edited by Ploderer, B., Carter, M., Gibbs, M. R., Smith, W., and Vetere, F., pp. 187–191, ACM, Parkville, Victoria, Australia, https://doi.org/10.1145/2838739.2838819, 2015.

Zimmerman, J. and Forlizzi, J.: Research through design in HCI, in: Ways of Knowing in HCI, edited by Olson, J. S. and Kellogg, W. A., pp. 167–189, Springer New York, New York, New York, USA, https://doi.org/10.1007/978-1-4939-0378-8_8, 2014.