Exploring Multiple and Coordinated Views for Multilayered Geospatial Data in Virtual Reality

Maxim Spur [1],* , **Vincent Tourre [1]** , **Erwan David [2]** , **Guillaume Moreau [1]**
and Patrick Le Callet [3]

[1] Architectural and Urban Ambiances Laboratory, Centrale Nantes, 44321 Nantes, France;
vincent.tourre@ec-nantes.fr (V.T.); guillaume.moreau@ec-nantes.fr (G.M.)

[2] Department of Psychology, Goethe University Frankfurt, 60323 Frankfurt am Main, Germany;
david@psych.uni-frankfurt.de

[3] LS2N, Polytech'Nantes, Université de Nantes, 44306 Nantes, France; patrick.lecallet@univ-nantes.fr

* Correspondence: maxim.spur@crenau.archi.fr

Abstract: Virtual reality (VR) headsets offer a large and immersive workspace for displaying visualizations with stereoscopic vision, as compared to traditional environments with monitors or printouts. The controllers for these devices further allow direct three-dimensional interaction with the virtual environment. In this paper, we make use of these advantages to implement a novel multiple and coordinated view (MCV) system in the form of a vertical stack, showing tilted layers of geospatial data. In a formal study based on a use-case from urbanism that requires cross-referencing four layers of geospatial urban data, we compared it against more conventional systems similarly implemented in VR: a simpler grid of layers, and one map that allows for switching between layers. Performance and oculometric analyses showed a slight advantage of the two spatial-multiplexing methods (the grid or the stack) over the temporal multiplexing in blitting. Subgrouping the participants based on their preferences, characteristics, and behavior allowed a more nuanced analysis, allowing us to establish links between e.g., saccadic information, experience with video games, and preferred system. In conclusion, we found that none of the three systems are optimal and a choice of different MCV systems should be provided in order to optimally engage users.

Keywords: virtual environments and data visualization; coordinated and multiple views; geospatial data visualization; immersive analytics

1. Introduction

Analysis and decision-making in geospatial domains often rely on visualizing and understanding multiple layers of spatial data. Cartographers have, for centuries, created methods of combining multi-layered information into single maps in order to provide multidimensional information about locations. More recently, the Semiology of graphics [1], and research on visual perception [2] led to advances in understanding how e.g., visual channels can be best employed to clearly represent as much data as possible in an effective and (often space-) efficient way [3].

While the above led to established practices for displaying many types of geospatial information, creating effective maps showing multilayered information remains a nontrivial task, even for domain experts, and research is ongoing [4]. With an ever-increasing amount of spatial data being collected and generated at faster rates and rising demand to get ahead of this data, there may not always be the time and resources to craft bespoke map visualizations for each new analysis task that requires understanding a multitude of layers. It may also not be practical to display too much information on one map, no matter how well designed, when the maps are too dense or feature-rich [5].

An alternative solution is displaying multiple maps of the same area at the same time. With computerization, this approach of multiple coordinated views (MCVs) [6] was adapted to this use-case [7], which can juxtapose different maps, or layers of a map, on one or multiple screens, and synchronize interactions between them, such as panning, zooming, the placement of markers, etc.

A downside to spatial juxtaposition is a reduction in the visible size of each map, as limited by screen space, and the head/eye movements required to look at different maps. On the other hand, MCVs should be employed when the different views "bring out correlations and or disparities", and can help to "divide and conquer", or "create manageable chunks and provide insight into the interaction among different dimensions", as recommended in the guidelines set forth by [8].

Commodity-grade virtual reality headsets (VR-HMDs) are steadily increasing their capabilities in terms of resolution and field of view, offering an omnidirectional and much more flexible virtual workspace than what is practical or economical with positioning monitors, prints, or projections in a real environment. Another benefit VR-HMDs provide is stereoscopic vision, which allows for a more natural perception of three-dimensional objects. Furthermore, VR devices, such as the HTC Vive or the Oculus Quest, usually come with controllers that are tracked in all axes of translation and rotation, presenting users with direct means of three-dimensional interaction with the virtual environment.

A crucial advantage of VR over AR for our application is the complete control over the environment, even in small offices, whereas the translucent nature of AR-HMDs requires a controlled real environment—a large enough and clutter-free real background to place the virtual objects. Another advantage with currently available headsets is the typically much larger field-of-view of VR-HMDs, providing less need for head movements and, crucially, showing more data at the same time, which is essential for preattentive processing [9].

These potential advantages appear to be applicable, even to the display of flat topographic maps without three-dimensional (3D) features, and allow for different kinds of spatial arrangements than otherwise feasible (no restrictions on monitor numbers or placement). A case has been made for separating and vertically stacking different data layers of a map [10], a MCV application that seems to be most suited for such an immersive system.

For this work, we continued the evaluation started in [11], in which we developed an implementation of this stacking system (further referenced as the stack), specifically with the VR case in mind, as we believe this is where its advantages are most pronounced and could be best utilized. In our hypothesis, its main benefit here is its ability to balance a trade-off of MCVs: stacking layers in this way allows them to be visually larger, and still closer together than through other means of juxtaposition—both being factors that we believe to be beneficial for visual analytics tasks involving the understanding and comparison of multiple layers of geospatial data, or maps. The nature of a vertical stacking system should also provide a secondary benefit of showing the layers in a more coherent way than other, more spread-out layouts, such as grids.

We set up a controlled user study to evaluate this stack's performance in visualizing multi-layered maps in a realistic decision-making task relevant to current work in urbanism. In it, we compared the stack to two more traditional methods of MCVs (Figure 1): temporal multiplexing or blitting, where all layers occupy the same space and a user toggles their visibility, and spatial multiplexing in a grid, showing all layers side-by-side.

Even though these other, more traditional methods have been tested and would work well in the traditional desktop computer environment, we implemented them using the same VR environment and means of interaction as our proposed stacking method. This allows for a fairer comparison on an even footing and controls for the "wow-effect" of using VR, particularly for test participants with little experience in it.

The design space for comparative or composite (geospatial) visualization encompasses more than these options [12], but other methods, such as overloading and nesting (e.g., by using a lens or by swiping), appear more practical for just two layers, and have indeed been investigated for

that purpose [5]. To our knowledge, no studies exist to date on comparatively evaluating these map comparison techniques with more than two layers, or executed in VR.
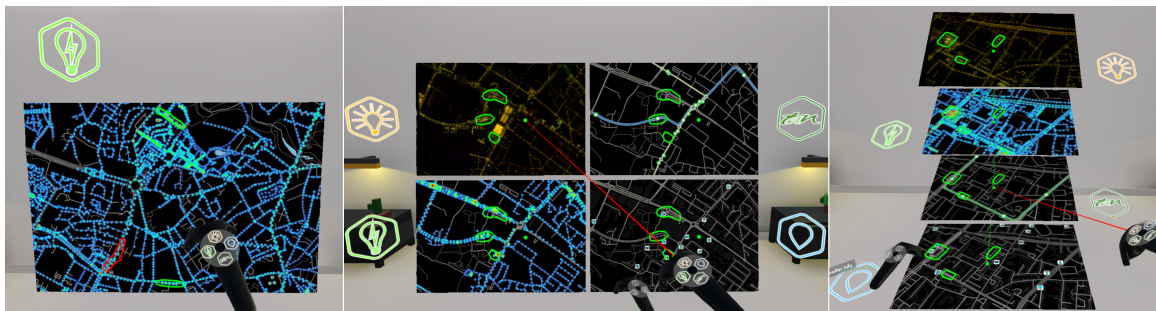
Figure 1. (**Left**): Temporal multiplexing (blitting); center and right: spatial multiplexing—in a grid (**center**), and in our proposed stack (**right**). Shown as implemented for the user study, with controller interaction.

Our paper contributes to research on multilayered map visualization:

- with a novel spatial multiplexing approach based on a stack of maps, derived from a study of the available design space for comparison tasks and its application in VR; and,
- an evaluation of this stack wholly done in VR, in comparison to two more traditional systems within a controlled user study.

1.1. Related Work

1.1.1. Urban Data Visualization

We chose the rapidly expanding field of urban data visualization as a particular domain in geospatial visualization to focus on. In [13], many current examples of urban data visualization are given from the point of view of visual [14] and immersive analytics [15]. Typical systems present one type of information, or closely related data like transportation on an interactive map [16], or superimpose few and sparse layers [17]. While most systems work with a flat map view, some started utilizing and showing in perspective projection the 3D shape of cities and buildings [18]. While this provides a better sense of the urban shape, occlusion of data can occur—this is addressed in [19] by "exploding" the building models. Vertical separation of data layers has also been done for legibility purposes when there was no occlusion to mitigate in [20].

1.1.2. Immersive Analytics

The emerging field of immersive analytics [21] aims to combine the advances of immersive technologies with visual analytics [22] and has already resulted in applications for large-scale geospatial visualizations—world maps of different shapes [23], global trajectories [24], and even complex flight paths [25]. However, urban environments have so far mostly been immersively explored only in 3D city models [26], or by adding data to one spatial layer [27]. Most recently, different spatial arrangements of small multiples for spatial and abstract data have been investigated in VR [28].

1.1.3. Multiple and Coordinated Views

In [29], arguments for immersive analytics are reiterated, with a call for more research into its application to coordinated and multiple views (interchangeably abbreviated to CMVs or MVCs) [6]—a powerful form of composite visualization by juxtaposition [12]. Many of the systems mentioned above contain MCVs in the shape of a map view augmented by connected tables or charts, others [30,31] also link related maps in innovative ways. While studies have been conducted to explore and compare the efficacy of different compositions of such coordinated geospatial map views [5],

to our knowledge they have so far only evaluated the case of two map layers, and also not within immersive environments.

1.2. System Design

1.2.1. Visual Composition Design Patterns

As explained in [5], combining multiple layers of geospatial data into one view can be a straightforward superposition, as long as the added information is sparse and the occlusion of the base map or blending of color or texture coding is not an issue. This is not the case when the map layers are dense and feature-rich, and this is where other design patterns of composite visualization views, as defined in [12], should be explored:

- Juxtaposition: placing visualizations side-by-side in one view;
- Superimposition:overlaying visualizations;
- Overloading: utilizing the space of one visualizations for another;
- Nesting: nesting the contents of one visualization inside another; and,
- Integration: placing visualizations in the same view with visual links.

Superimposition methods, as opposed to the plain superposition described above could be made useful if applied *locally*, e.g., like a lens [5]. With more than two layers though, a lens-based comparison interface quickly becomes less trivial to design, e.g., a "base" layer becomes necessary, as well as either controls or presets for the size, shape and placement of a potentially unwieldy number of lenses [32]. When mitigating this by using less lenses than layers, it becomes necessary to fall back to temporal multiplexing, as discussed below.

Global superimposition through fading the transparency of layers is again only applicable to sparse map layers and not to the general case of dense and heterogeneous ones, where e.g., colors that are typically employed for visualization could blend and become unreadable.

Overloading and nesting can be dismissed for map visualizations. Though they are related to superimposition, they are defined to lack a "one-to-one spatial link" between two visualizations, which is central to most map comparison tasks.

This leaves juxtaposition and its augmented form, integration, which adds explicit linking of data items across views. Those are familiar and relatively easy to implement design patterns that have been shown to increase user performance in abstract data visualization [33]. The challenges in designing effective juxtaposed views, as [12] describe, lie in creating "efficient relational linking and spatial layout". The first challenge could be addressed by relying on the integration design pattern, and the second one is where we propose the vertical stack method as an alternative to be evaluated against a more classical, flat grid of maps.

In Gleicher's earlier work [34], juxtaposition is also talked about in the temporal sense: "alternat[ing] the display of two aligned objects, such that the differences 'blink'". Lobo et al. [5] refer to this as *temporal multiplexing* or blitting and see it also as a version of superimposition. As one of the most common composition techniques (e.g., flipping pages in an atlas, or switching map views on mobile device), and one we observed being used by the urbanists in our laboratory working with geographic information systems, we also included it in our comparative study.

1.2.2. The Stack

Because the strategy a user will employ for comparing layers will be a sequential scan [35], we believe a design where the distances between these layers are minimal would fare better. Figure 2 shows how the stack helps this sequential scan by presenting each layer in the same visual way—all layers share the same inclination and distance relative to the viewer and are thus equally distorted by perspective. Increasing this inclination allows for the layers to be stacked closer together

without overlap—reducing their distance, while still preserving legibility up to reasonable angles [36]. This would also allow more layers of the same size in the same vertical space that a grid would occupy.
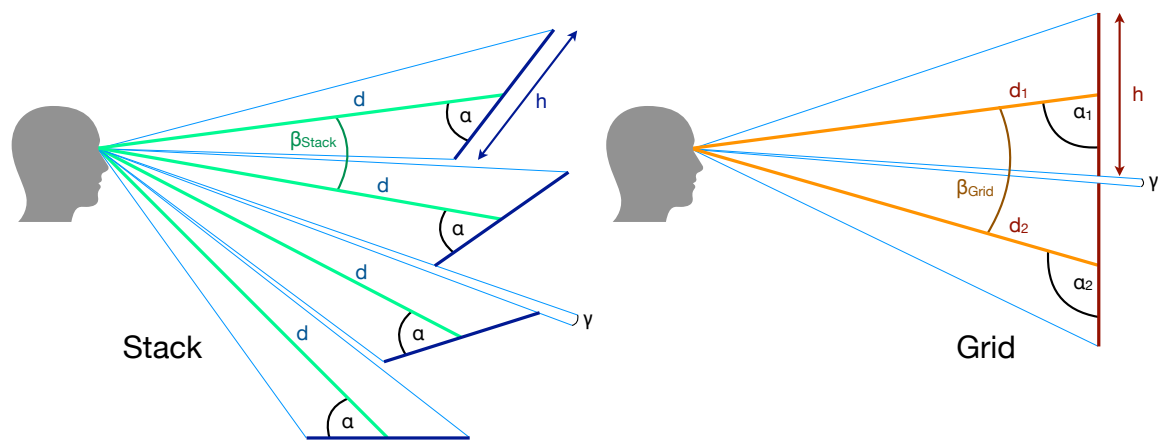
Figure 2. Spatial arrangement of the layers in the stack and grid systems, as seen from the side, highlighting the larger visual distance (β) between layers in the grid, assuming same height (h) of the maps, same minimum distance to viewer (d_1), and same gap between layers (γ). The viewing angles (α) and distances (d) are constant in the stack and different in the grid, depending on viewer position.

Scanning through a stack also requires eye movement in one direction only—all representations of an area are aligned vertically, as opposed to being spread out in the two dimensions of a grid. Additionally, the way that the individual maps are arranged in the stack mirrors the way maps are traditionally, and still today, often viewed in professional settings: as a flat print or display on a table, inclined to the viewer—even in VR applications [27]. All of the maps share the same relative inclination and distance to the viewer, as well as the same visual distance to each other, avoiding occlusion (all of which is constantly being adjusted by tracking the HMD's position), and thus the same perspective distortion, making them easier to compare visually [37]. This way, points on one map have the same visual distances to each other across each map. Conversely, in a grid of coplanar maps, the distances and angles for each map all differ from each other, resulting in differently distorted maps. The tracking and adjusting of tilt and distance is also a major differentiation from a static "shelf" of maps, which can be seen as a grid where each map lies flat instead of being upright, resulting in the same inconsistent distortion characteristics of each item [28]. A number of possible arrangements of the stack system and comparisons to a grid are shown and explained in Appendix A.

The binocular display of VR gives an immediate stereoscopic cue that the maps are inclined and not just distorted. This hels the visual system to decode the effect of perspective and removes the need for kinetic depth cues [38]. In addition, this inclination is also made clearly visible to a user by framing the map layers in rectangles, which helps to indicate the perspective surface slant. Studies exist that show how picture viewing is nearly invariant with respect to this kind of inclination or "viewing obliqueness", bordering on imperceptibility in many cases [36].

2. Materials and Methods

2.1. Task Design and Considerations

In Schulz' *A Design Space of Visualization Tasks* [39], the concepts of data, tasks, and visualization are combined in two ways to ask different questions:

- *Data + Task* = **Visualization?** and
- *Data + Visualization* = **Task?**

The first combination asks which visualization needs to be created for a given task and data, whereas the second can follow as an evaluation process, once a visualization has been defined: how well can tasks be performed on these data using this visualization?

To perform this evaluation, the task and data had to be well defined. Usually, the effectiveness of geovisualization systems is evaluated with simplified tasks, such as detecting differences between maps or finding certain features on a map [5]. While these methods can often be generalized to map legibility, we aimed at defining a task that could more directly test how well a system can facilitate an understanding of multiple layers of spatial data. Such task design was key to this project, and conducting the experiment with it, the applicability of that methodology to evaluate geovisualization systems could also be investigated.

Gleicher argues that "much, if not most of analysis can be viewed as comparisons" [35]. He describes comparison as consisting in the broadest sense of items or targets, and an action, or what a user does with the relationship between items.

Those actions have in earlier literature on comparison been limited to identifying relationships, but Gleicher extends the categorization of comparative tasks to include: measuring/quantifying/summarizing, dissecting, and connecting relationships between items, as well as contextualizing and communicating/illuminating them.

In our case, one can see how a user would do most of the above actions when comparing the data layers—our items/targets—to come to and communicate a conclusion.

Following Gleicher's considerations on what makes a "difficult comparison", our task needs to be refined:

- the number of items being compared;
- the size or complexity of the individual items; and,
- the size or complexity of the relationships.

The first two issues we directly addressed by simplifying the choice a user had to make. We divided the area of the city for which we had coverage of all four data layers into twelve similarly-sized regions—one for each scenario. In each, we outlined three items—the candidate areas, out of which a user would then only select the one most "problematic" candidate. This considerably reduces the number of items, their size and complexity. Having a fixed number of candidates in all scenarios allows us to compare completion time and other metrics, such as the number of times participants switched their attention from one candidate to another, in a more consistent way. Different numbers of candidates could help to generalize findings, but would also require accordingly longer or a larger number of experiment sessions, an endeavor that we relegate to future work.

To address the relationships' complexity, we provided candidates that were as similar to each other as possible, while differing in ways that are interesting from the urbanist point of view. For example: one scenario's candidates are all segments with a roundabout, have similar energy consumption, but one of them is close to a tram stop, and another has stronger light pollution. We aimed for the users to balance fewer aspects, while providing insights to urbanists regarding the remaining differences that mattered the most in a decision.

2.2. Use Case: Urban Illumination

We developed a use case around ongoing research into public city illumination, and citizens' view of it, with the help of a group of experts consisting of urbanists, architects, and sociologists. The topic requires understanding and appreciation of multilayered spatial information by naïve study participants previously not exposed to this data to make informed choices. It relies on four data layers (Figure 3):

- Light pollution: an orthoimage taken at night over the city,
- Energy consumption: a heatmap visualization of the electrical energy each street lamp consumes,

- Night transportation: a map of public transit lines that operate at night and their stops, including bike-sharing stations; and,
- Night POIs: a map of points of interest that are relevant to nighttime activities.
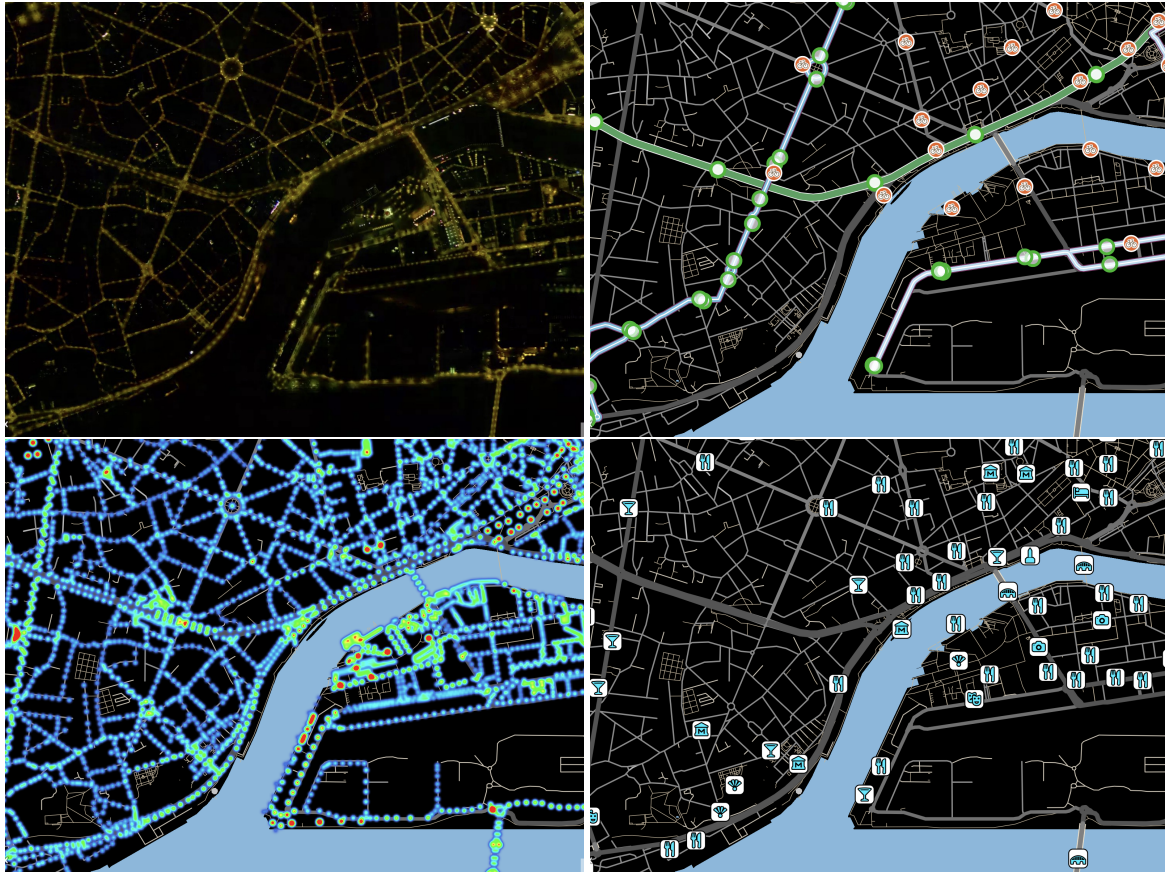
Figure 3. Crops of the four data layers, as used as textures in the MCV systems: light pollution, nighttime transportation, street light energy consumption, and points of interest at night.

Thus, the participants in our study have the task of gaining an understanding of these four layers for certain regions, in order to then decide which one of the pre-defined zones within each region are most "problematic" in terms of public lighting. Each layer's significance for that purpose is explained multiple times to the participants throughout the introductory part of the sessions, and reminders are available on demand. For the purpose of the study in urbanism, the participants play the role of a local citizen without special domain knowledge, so this explanation and the legibility of these map layers is key for this task. The participants are told that their input will help shape the considerations in how to update urban illumination, as new regulations are being discussed to limit light pollution and the energy consumption associated with inefficient lighting. The explanations include a direction to reduce excessive lighting where not necessary, while keeping in mind that critical areas, such as transportation hubs or highly frequented public places, should stay well-lit. A "problematic area" could also be a place that does not receive enough light, based on these factors. Thus, this urban study aims to gauge the citizens' perspective, which factors drive their decisions, and how they balance them.

2.3. Implementation and Interaction

2.3.1. Technology

On the *Mapbox* platform, we created the four data layers as "map styles", using data from *OpenStreetMaps* (nighttime transportation network and points of interest), as well as from data that were

provided by the local metropolitan administration (light pollution and public street lamp information). The Mapbox API allowed us to use these map styles as textures directly within the *Unity 3D* game engine [40] to build our three layering systems for navigating the maps and selecting candidate areas.

2.3.2. Configuration

The layers were presented as floating rectangular surfaces with a 3:2 aspect ratio in front of the seated user. The blit system's single view measured 1.5 m × 1 m (width by height), approximately 2 m away from the user, and it was oriented perfectly vertical, providing a 90 degree viewing angle if facing it centrally. The grid system's four views measured 1.125 m × 0.75 m, approx. 2.2 m away from the user, and tilted by 10 degrees in order to provide a 80 degree viewing angle if viewed centrally. Finally, the stack's layers measured 1.2 m × 0.8 m, and kept at a distance of approximately 1.8 m. Tracking the position of the user's HMD in virtual space, each layer of the stack was constantly being oriented to have a viewing angle of 45 degrees, and moved to have the same distance as the lowest layer in the stack as well as to preserve the visual distance between the layers, as illustrated in Figure 2, ensuring that all layers looked equally distorted by perspective, regardless of the viewer's position. All of the systems were positioned to hover slightly above the virtual room's floor.

These arrangements and configurations were not mathematically designed to provide the exact same fields of view for each system. The different distortion characteristics of the systems, and the difference between seeing only one or all four layers at the same time would prevent a total visual equivalency—it is not clear how to balance these factors. Instead, we iterated on different arrangements with the previously mentioned group of experts during the development and pre-testing phases, asking for their feedback and adjusting the systems in real time, so as to subjectively appear as equivalent to each other as possible to them. Each system was tuned to maximize the experts' satisfaction with them, which resulted in the above dimensions and positions. As a further counterargument to a more mathematical approach, we anticipated the users to sit at slightly different positions relative to the views, and even move around slightly, which would render more precise optimizations less effective. The goal was to make viewing and interacting with all three systems as comfortable and as equivalent to each other as possible, given their differences, in order to allow a fair comparison of these systems.

2.3.3. Interaction

For displaying the VR scene, we employed the first-generation HTC Vive HMD and controllers. The two controllers are physically identical, providing six-axis tracking in space, a large trackpad, and a number of buttons. We used both, dividing them into two main functions: one controller for interacting with the map and the other for selecting and confirming candidates, as well as for progressing through the tutorial system.

The map interactions consist of panning and zooming—both are accomplished by moving the controller in space while holding its side button: motions that are parallel to the map plane resulted in translation, or panning of the map image, while motion perpendicular to it results in zooming it in or out: "pushing" or "pulling" the map. If on the blit system, the touchpad of that controller also served for switching between map layers—an annular menu hovered over it with symbols representing each layer; swiping the thumb to a symbol would switch the view to that layer. The same symbol would also appear hovering above the map view, and next to each data layer in the other systems for consistency (Figures 1 and 4).

Figure 1 also shows an additional visual aid that can be enabled by pressing the map controller's trigger: a "laser" pointer emanating from its tip. It places a spherical marker where it intersects with a map view, and in the two spatial multiplexing systems also duplicates them across the other layers (Figure 1). In the case of the stack, those markers are also connected with vertical lines. The result is additional visual linking across the map layers—implicit in the grid, and explicit in the stack, which thus gets elevated to an integrated view design pattern, according to [12].

Figure 4. The interface for selecting and confirming candidate zones ((**left**) and (**middle**), respectively), and optional information panels for the map layers (**right**).

The other controller used the same touchpad method for switching between candidates. The annular menu contained the numbers 1–3 and, once selected, the corresponding candidate's outline turns from green to red, and the same number appears in a corner above the map view. Touching the middle of the pad resets the choice. If an outline is obscuring underlying data, the controller's trigger input can be used to fade its opacity gradually down to zero. Once the participant wishes to finalizes their choice, the touchpad needs to be pressed down for a moment until the confirmation interface appears: a red cube, offset in position by 20 cm away from the controller. To confirm, the tip of the controller has to be moved inside this cube and held there for three seconds (Figure 4).

At the beginning of each session, a text field was floated above the map navigation controller, instructing users to take it in their dominant hand, as the fine movements that are related to navigation and pointing require more dexterity than the actions of the candidate selection controller.

2.4. Participants

26 participants took part in the experiment (9 female, 17 male), with ages ranging from 18 to 45 years (M = 21, SD = 6.33).

Twenty-one reported as currently being students, with 17 holding at least a bachelor's degree. As per study design, none had a background in urbanism. Most of the participants had either never tried VR before (10), or for only less than five hours total (13). Two have had between five and twenty hours of VR experience, and one more than twenty hours. We also asked about experience with 3D video games: nine participants reportedly never play those, nine others only a few times per year. Four play a few times per month, one a few times per week, and three play every day.

The participants' responses to questions about their familiarity with the city we visualized, its map and their comfort of reading city maps were normally distributed on the visual analog scales we employed in our questionnaires. For screening, we tested all of the participants on site for unaided visual acuity and colorblindness, with all passing these tests.

2.5. Stimuli

We divided the parts of the city map that were covered by the additional data layers into twelve smaller and constrained parts (limited by pan and zoom) for our twelve "scenarios", with each containing three outlines demarcating the candidate zones, as explained in Section 2.1. Pairing one such scenario with one layering system resulted in one stimulus. One of these scenarios was chosen to always be used for the first tutorial to introduce the data layers and explain interaction with the blit system, based on its characteristics: it showed the most well-known part of the city, and seeing all candidate zones required panning the map with the controllers. The other eleven scenarios appeared in a random order for the remainder of each session: two for the remaining part of the tutorial in order to introduce the other two systems, and nine for the actual tasks.

2.6. Design and Procedure

The experiment followed a within-subject design. We exposed each participant equally to all three layering systems and, thus, could inquire directly about relative preferences. Using the scenarios, we split the sessions into a a tutorial and an evaluation phase: the first consisting of three scenarios (one for each system to be introduced) and the remaining nine allocated for the phase in which we collected data for the analysis after the participants were trained to use the systems. Each scenario itself is further divided into an instruction phase, the task phase, and the questionnaire part, where the participants reflect on the just completed task. The instructions were full, interactive step-by-step manuals inside the tutorial phase, and shortened to quick reminders in the following evaluation phase. Once a task was complete (a candidate zone selected and confirmed), data recording was stopped and the questionnaire part commenced.

2.6.1. Tutorial Phase

After a brief introduction on how to handle the VR headset and controllers by the experimenter, the tutorial phase—consisting of three scenarios—began. The first scenario showed the same map region and used the blit layering system for each user, to ensure maximum consistency in their training. Blitting was chosen here, as this is the closest to what participants were likely to already be familiar with from using digital maps, and because it appears as just a single map at a time, which allowed explaining the significance of each map layer in sequence and without interference.

At first, all of the controller interaction is disabled. The tutorial system gradually introduced and enabled panning, zooming, and blitting the maps, and selecting and confirming candidates by asking the participant to perform simple tasks and waiting for their successful completion. While introducing the blitting mechanism, each layer was explained in detail—participants could not switch to the next one before confirming their understanding of the summary. The first such training scenario concluded with a reiteration of the common task of all scenarios: selecting and confirming one of the three candidate zones, followed by the instruction to take off the HMD and proceed with the questionnaire, which introduced the types of questions that will be answered throughout the session.

The second and third scenarios were used to introduce the grid and stack systems in a similar fashion, forgoing explaining each layer again. They similarly ended with the actual task and the questionnaire. Here, the order was balanced between the participants: half were first exposed to the grid, the other half to the stack. The map regions were randomized from this point on.

2.6.2. Evaluation Phase

The remaining nine scenarios were divided into three blocks of three tasks: a repeating permutation of the three layering systems (G(rid), S(stack), B(blit)), balanced across the participant pool, e.g., GSB-GSB-GSB, or BGS-BGS-BGS. The tutorial system was pared down to only instruct participants in the blitting system to switch between all layers at least once at the beginning, and to show reminders of the controller functions as well as the layer descriptions if requested. Instructions appeared to evaluate the scenario, pick a candidate, and proceed to the questionnaire.

2.6.3. Balancing

This choice of procedure—always introducing the blit system first—could appear to impact the balancing. However, we believe that presenting this system first allowed us to create the easiest in-system tutorial for introducing each layer separately, as opposed to overwhelming the participants with all the data layers at once. In this aspect, it is also the simplest system to start with for explaining the interaction methods necessary to navigate the maps and selecting candidate zones. It is also the closest to what we believe the existing experiences of our participants are with multilayered maps: spatial multiplexing is rather uncommon outside of professional applications, whereas typical digital maps usually offer an interface for switching between data layers. The participants are assumed to

already be "unbalanced" in this regard: the blit system does not introduce any new concepts to them, and rather only teaches about the general interaction common to all three systems. Therefore, we think our approach of introducing the blit system first, and then randomly the grid or the stack second and last, was the most balanced possible, considering real-life experiences. The two "novel" systems were introduced in a balanced way, and the trial phase after this introduction was completely balanced.

2.7. Apparatus and Measurements

We employed two separate devices for our experiment: One PC handling the execution of the layering systems within the HTC Vive VR setup, and a second one with the sole task of hosting the digital questionnaire, created with the tool that is described in [41]. Thus, instructions for the participants and all interactions for the complete duration of the experiment could be fully automated between the prototype system and the questionnaire. The interactive questionnaire instructed participants to put on the HMD, the prototype system told them when to take off again and go back to the questionnaire to continue with their evaluations, and so on until the all tasks have been completed. Switching between these devices provided the benefit of giving the participants regular pauses from wearing the HMD, the excessive use of which might have lead to unnecessary fatigue.

As alluded to above, we employed the first-generation HTC Vive setup (HMD plus the two Vive controllers), albeit with the SMI binocular eye-tracking retrofit, capable of sampling gaze positions at 250 Hz with a reported accuracy of 0.2 degrees. The system ran on a Windows workstation powerful enough to render the prototype system at the HMD's maximum frame rate of 90 Hz and its maximum resolution of 1080×1200 pixels per eye at constant performance.

The interactive digital questionnaire ran on a separate PC in a Chrome browser window, close enough so the participants could stay in their swivel chair to access it, but far enough not to interfere with the VR tasks. It recorded subjective data by asking the participants about their reasons for choosing the candidates they did after each trial, as well as by providing text fields for free-form feedback about each individual system after all trials were done. In this post-hoc phase, it also presented pairwise preference questions about the systems under three aspects: legibility of the map layers within these systems, their ease of use, and their visual design.

To complement this declarative feedback, the VR prototype system also recorded performance aspects (completion times, interactions with the system) and visuo-motor data via oculometry and tracking of the HMD. Eye movements bring a wealth of information—they are overt clues about an observer's attention deployment and cognitive processes in general, and are increasingly being tracked for evaluating visual analytics [42]. In the context of map reading, measuring gaze allows for us to know precisely which map layers participants chose to observe in particular, and at which times. Furthermore, gaze features and their properties, such as saccades and fixations can be derived, in this case by processing with the toolkit developed for the *Salient360!* dataset [43] and further refined in [44], using a velocity-based parsing algorithm [45].

3. Results

3.1. Preferences

We asked the participants about their opinions regarding different aspects of the three layering systems. To quantify their preferences, we used pairwise comparison questions (randomized and balanced) to create pairwise comparison matrices (PCMs) for each of these three aspects:

1. Map legibility: which system showed the map layers in the clearest way and made them easier to understand for you?
2. Ease of use: which system made interacting with the maps and candidate areas easier for you?
3. Visual design: which system looked more appealing to you?

The resulting PCMs could then be used in different ways. First, we determined rankings for the systems by counting the number of times one "won" against the others in the comparisons under each

of the three aspect. Thus, systems "winning" twice we called the "preferred" system by the participant under that aspect. Winning once placed them second, and zero wins resulted in it being last place; what we called the "disliked" system. Figure 5a shows these results.
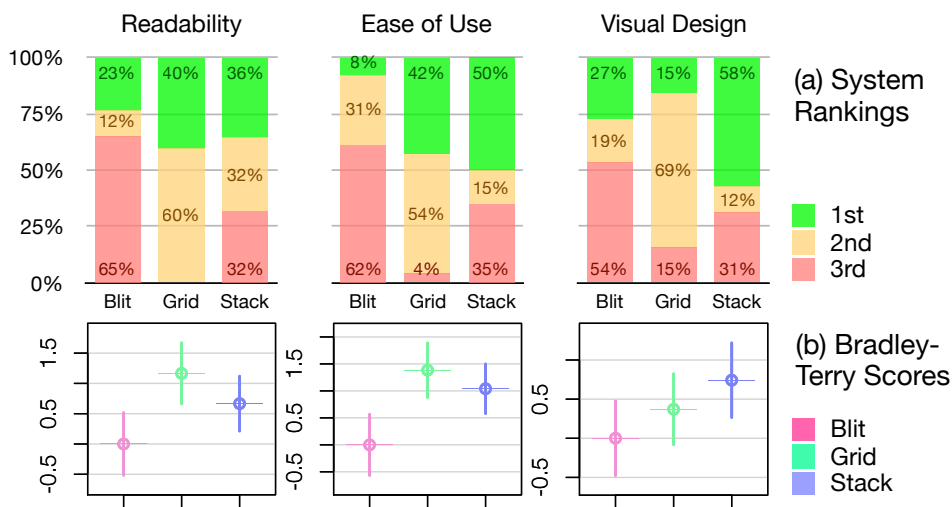
Figure 5. (**a**) How often (in percent) each system was ranked *first*, *second*, or *last* in terms of readability (legibility of the map layers), ease of use, and visual design by the users; (**b**) Pairwise comparisons scored with the Bradley–Terry model.

By this measurement, more than half of the participants placed the blit system last under each aspect. The grid system was almost evenly split between being rated first or second under the ease of use and readability aspects, while being a clear second in terms of visual design—here, the stack system proved to be the favorite. The stack also shows a slight advantage with its "firsts" in terms of ease of use over the grid.

These PCMs could also be fed into a Bradley–Terry model [46] to generate relative scores for each system (c.f. Figure 5b). These results mostly mirror the results from our approach above (c.f. Figure 5a) in a less granular way, with a key difference: it gives the grid system a clearer advantage over the stack in terms of readability and, to a smaller degree, ease of use. This can be explained by the evidently polarizing nature of the stack system, in stark contrast to the grid. The grid was rarely if ever rated last in any aspect, and usually second place, whereas the stack garnered the highest number of first place ratings, as well as a considerable number last places. This "love it or hate it" kind of response can also be seen in the blit system to a lesser degree, in that it similarly had a very low number of second place ratings, but it was also rated as being worst far more often than being first.

Condensing those pairwise comparisons further, we counted how many times the participants ranked each system first and last across the three aspects (Figure 6a). This method shows how the blit system received zero "first place" rankings from the majority of our participants—under no aspect was it "the best" among the three systems for them. It also shows how only the stack system ever received a "first place" under all three aspects; in fact, from a quarter of our participants.

The chart showing the number of "lasts" received allows similar observations: the blit system received the maximum number of "lasts" most often, and the stack system had an overall lower number of "lasts" than it. However, the grid received by far the least number of lasts: only a fifth of the participants deemed it to be worse than the other two systems under a single aspect. Most participants thus found either the blit, or less frequently, the stack system to be worse than the grid system under at least one aspect; once again showing it to be a "middle ground" system.
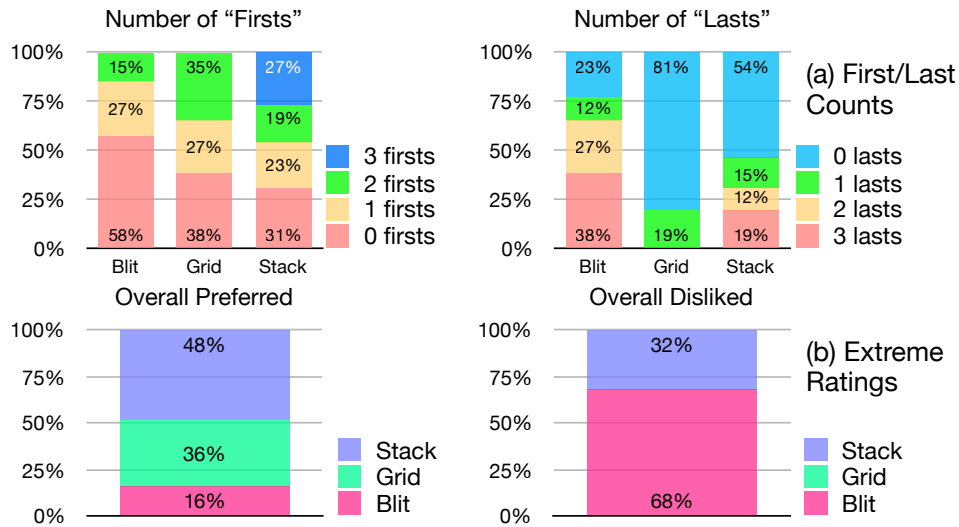
Figure 6. (**a**) Which proportion of participants gave the systems a number of zero to four *first* or *last* rankings across all aspects (legibility, ease of use, and visual design); (**b**) which proportion gave the most "firsts" (and therefore "preferred") or "lasts" (and therefore "disliked") to each system in the pairwise comparisons.

In our last distillation of the PCMs, we measure how many of the total extreme ratings, or "firsts" and "lasts" (as per the previous measurement above) each system received, allowing us to present the overall "most preferred" or "most disliked" system in Figure 6b. By this measure, the stack was preferred overall by almost half of our participants, followed by the grid with more than a third. Two thirds of the participants showed strong dislike for the blit system, and none of them did so for the grid system—once again showing its noncontroversial nature.

3.2. Participant Characteristics

The two most interesting juxtapositions between the participants' characteristics and their system preferences came from their reported habits regarding playing 3D video games and consulting maps of cities, as illustrated in Figure 7.
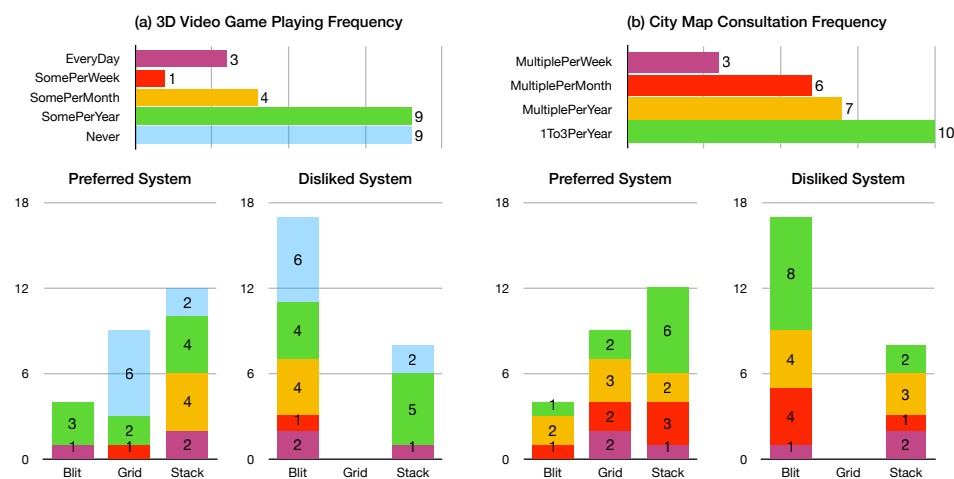
Figure 7. Left (**a**): the numbers of participants who play 3D-based video games at different frequencies, and their distribution among those who "preferred" and "disliked" each system, according to the measurement in Figure 6; right (**b**): same for the frequency of consulting city maps.

Out of those identifying as never playing such games, most prefer the grid system, and some the stack—none picked the blit system as their favorite as measured by the method in Figure 6. The stack

system shows a tendency to be favored by those who play at least a few times per month. Those who use maps most frequently (at least multiple times per month) disproportionally prefer one of the spatial multiplexing systems, especially those who most rarely use maps preferred the stack system most often.

Concerning the "disliked" system charts, the blit system overwhelmingly earned the least favorite status from from all kinds of participants, however mostly from those with the least gaming experience, and those who very infrequently consult city maps. None placed the grid system last, as seen before in Figure 6b.

3.3. Subgroupings

Because our main goal was to evaluate the legibility of the systems, we used the rankings from Figure 5 to split the participants' data into three subgroups (plus the total) for further analysis: *pB*, *pG*, and *pS*. Those refer to data from users who *preferred* the B(lit), G(rid), or S(tack) system, i.e., ranked it first in the pairwise comparisons under that aspect. This could be done, since the number of participants who did so were roughly comparable: out of the 26 total participants, six fell into *pB*, 10 into *pG*, and nine into *pS*.

Similarly, looking at how the user characteristics from Figure 7 seemed to interact with their system preferences, we created comparable subgroups based on the frequency with which the participants play 3D video games, and consult city maps. For the video game frequency, we created the groups *zG*, *sG*, and *fG* for:

- (z)ero (never, n = 9),
- (s)ome (some per year, n = 9), or
- (f)requent (at least some per month, n = 8) 3D video (G)aming,

    and for the map consultation frequency *rM*, *sM*, and *fM* for:

- (r)are (1–3 per year, n = 10),
- (s)ome (>3 per year, n = 7), or
- (f)requent (at least multiple per month, n = 9) city (M)ap consultation.

Figures 8–11 display these subgroups, as well as a (T)otal measurement for comparing each MCV system across all users.
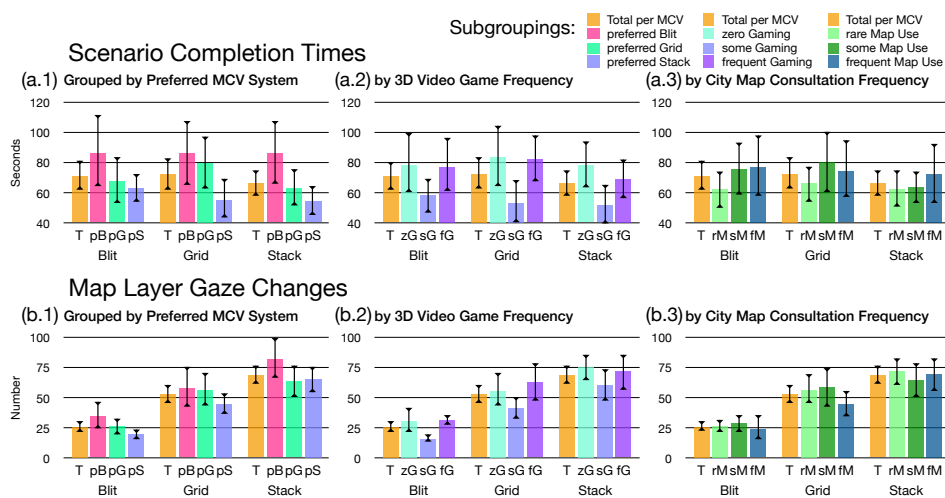
Figure 8. Performance measurements for each MCV system (blit/grid/stack): scenario completion times (**a**), and the number of times gaze has shifted from one map to another per scenario (**b**). Further subgrouped by the participants' preference for an MCV system regarding its legibility (1), the frequency with which they play video games (2), and the frequency with which they consult city maps (3).
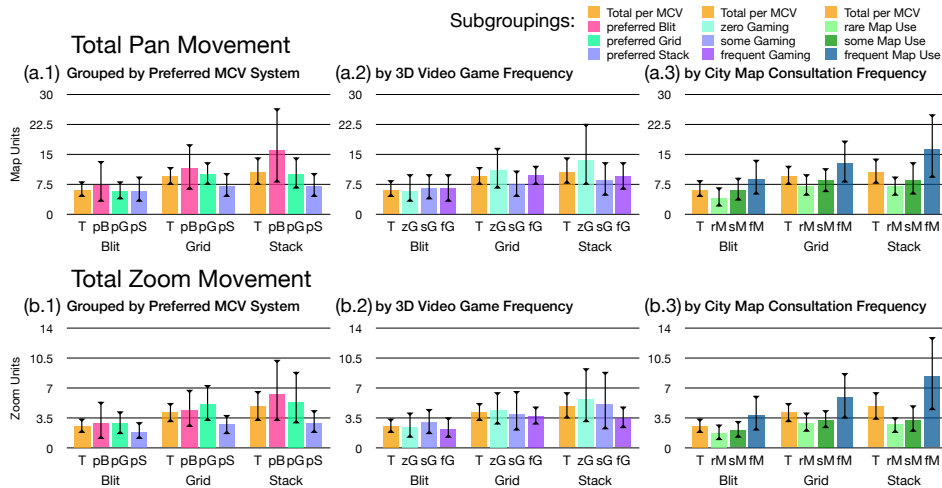
Figure 9. Map navigation measurements for each MCV system (blit/grid/stack): how much they panned the map per scenario (**a**), and how much they zoomed in/out of the map (**b**). Further subgrouped by the participants' preference for an MCV system regarding its legibility (1), the frequency with which they play video games (2), and the frequency with which they consult city maps (3).
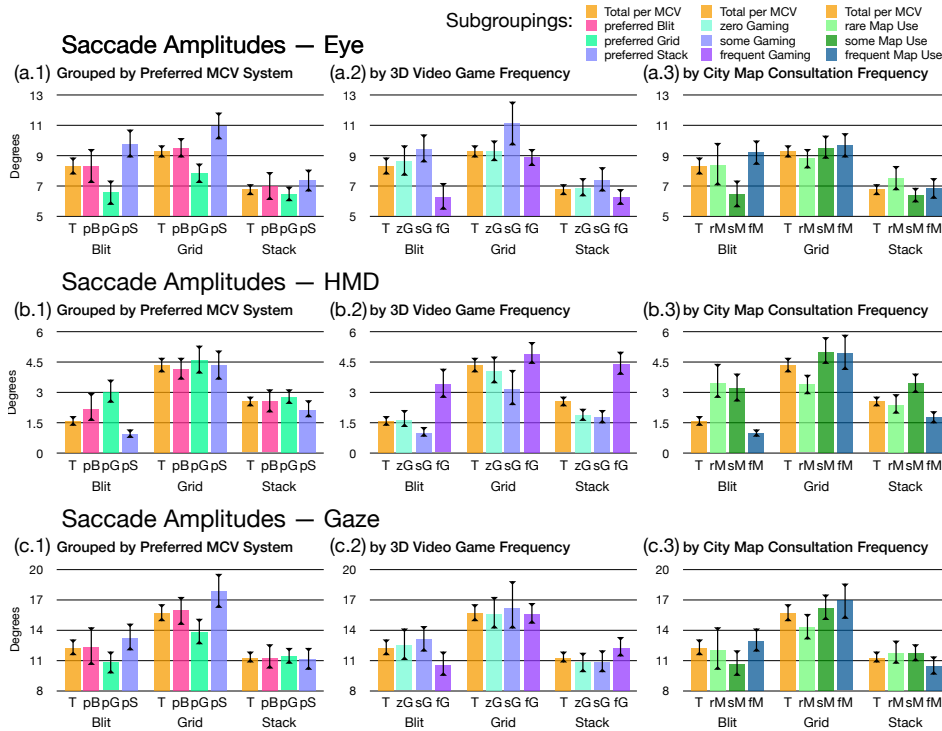
Figure 10. Average saccade amplitude measurements for each MCV system (blit/grid/stack): split by the eye (or viewport) component (**a**), the head movement (or HMD rotation) component (**b**), and taken together to determine the overall gaze (**c**). Further subgrouped by the participants' preference for an MCV system regarding its legibility (1), the frequency with which they play video games (2), and the frequency with which they consult city maps (3).
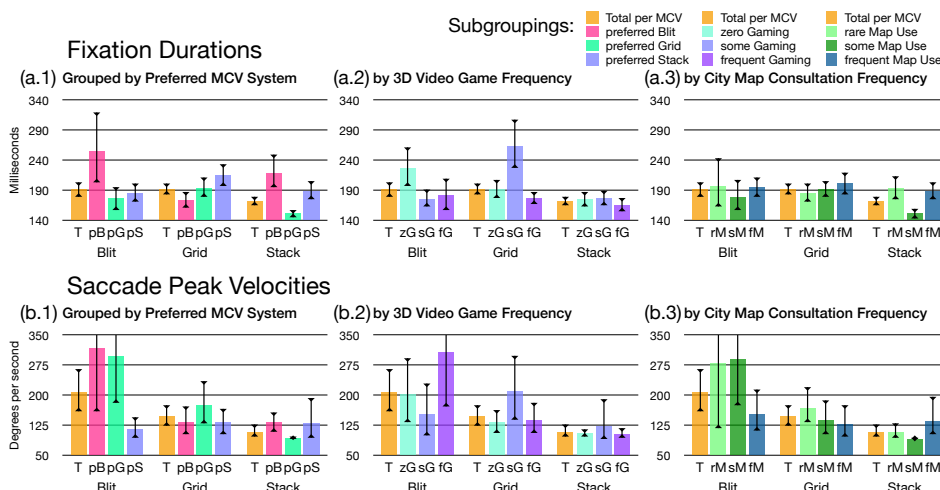
Figure 11. Further visuo-motor behavior measurements for each MCV system (blit/grid/stack): the average fixation durations between saccades (**a**), and the average peak velocities reached during saccades (**b**) (upper confidence intervals cut to increase legibility of smaller values). Further subgrouped by the participants' preference for an MCV system regarding its legibility (1), the frequency with which they play video games (2), and the frequency with which they consult city maps (3).

3.4. System Interactions and Performance

We deemed the measurements of how quickly they completed each scenario (Figure 8a), how many times they shifted their gaze from one map to another during each scenario (Figure 8b), and how much they panned and zoomed the maps (Figure 9) the most salient indicators in order to investigate how the participants interacted with each MCV system.

When comparing the scenario completion times across the systems (Figure 8a) using the (T)otal values from all participants, there are no significant differences between MCV systems. The preference subgroups here (Figure 8(a.1)) however point toward a tendency of those who prefer the stack system (*pS*) to be significantly faster in each system than the others, especially in spatial multiplexing systems. Those who prefer the blit system (*pB*) appear generally slower across all systems, albeit with a relatively high variability.

Similarly, we can see a differentiation when grouping by video game frequency (Figure 8(a.2)). Significantly faster were those who play video games with a medium frequency (*sG*), while the frequent players (*fG*) were as comparatively slow as those who never play (*zG*), although with a slight tendency to be faster in the stack system. A possible interpretation here could be that those who never play felt less pressure to complete the scenarios as quickly as possible, whereas habitual gamers could have seen a quick completion as a challenge. Those playing the most may have seen the challenge not in speed, but in accuracy of their answers or, alternatively, they could have enjoyed their time in the virtual environment more than the other groups, not feeling the need to exit it as soon.

A very similar distribution among those two subgroupings can be seen for the number of map layer gaze changes (Figure 8(b.1,b.2)), but there the system itself factors significantly more in that measurement. The most gaze changes happened in the stack system, and the least in the blit system—where each gaze change had to be accompanied by a physical interaction with a controller. In the blit system, the difference between the preference subgroups is quite large: those who prefer it change their gaze in it the most, and those who prefer the stack system do so the least. This can be interpreted in the way that those who prefer the stack rely more on their peripheral vision for comparisons, rather than directly shifting their gaze, which is easier to do in the stack system, where the adjacent layers are much closer together, hence their preference for that one and their dislike for the other systems.

The map consultation frequency subgrouping (Figure 8(a.3,b.3)) does not impact either measure as much as the others, only showing a slight trend for those who read city maps with the lowest

frequency (*rM*) to spend less time in each scenario, especially in the blit system, and for the most frequent map readers (*fM*) to make less gaze changes in the grid system.

This subgrouping is more significant in its differentiation for the pan and zoom movements (Figure 9(a.3,b.3)). Generally, the participants were navigating around the maps much more in both spatial multiplexing systems than in the larger, temporal multiplexing blit system, however the most frequent map users were clearly moving the maps much more, particularly by zooming in and out (Figure 9(b.3))—reflecting perhaps on their experience with navigating maps. They also had a higher confidence interval in this metric, which could be attributed to a more differentiated handling of different scenarios. Otherwise, the other subgroups track with their completion times—the more time they spend in a scenario, the more they navigate.

3.5. Visuo-Motor Behavior

More substantial differences between the MCV systems and the participant subgroups than those visible in the interaction data (Section 3.4) could be obtained by investigating their average saccade amplitudes (Figure 10), which consist of the eye movement component (a) and the head movement, as tracked by the HMD's position (b), to obtain insight into the movement of the participants' gaze (c) when interacting with the systems. This visuo-motor behavior is made even more legible by the analysis of the average durations of the fixations between those saccadic movements, and the average peak velocities reached during these saccades, as shown in Figure 11.

3.5.1. Saccade Amplitudes

Both head and eye movements, as expressed by saccade amplitudes (Figure 10a,b), were significantly the highest for the grid system, and lowest for the stack system, which corresponds to the distances between the layers in both systems. The blit system is in the middle, as there is only one map view, but it is quite large itself. All of the subgroups show significant differences within the systems for the eye movement, except within the stack MCV (Figure 10(a.1,b.1)): the closeness of the layers appears to have led most participants to similar behavior, whereas the other systems led to differentiation in accordance to the subgroup. Those who prefer the stack have far larger amplitudes in their eye movements in the other two systems—they seem to use their strategy that works for them in the stack, which could be a frequent comparison of far-apart regions or different layers to one another, in the other systems as well, where it requires much larger eye movements. This additional distance may be a factor in their choice of ranking those larger systems as worse in terms of legibility.

Looking at the saccade amplitudes as subgrouped by video game playing frequency (Figure 10(a.2,b.2,c.3)), shows a differentiation between this grouping and that of the preferences in (Figure 10(a.1,b.1,c.1)). The most frequent gamers move their heads far more in larger motions than those who play less (Figure 10(b.2)), and inversely move their eyes far less than those who play less (Figure 10(a.2)). Those differences cancel each other mostly out, and the overall gaze by the most frequent gamers is much more similar to the other participants (Figure 10(c.2)), save for a tendency to have smaller amplitudes when using the blit system, and larger ones in the stack than the average.

The saccadic movement as subgrouped by preference (Figure 10(a.1,b.1,c.1)) is however almost perfectly mirrored in the subgrouping by map reading frequency (Figure 10(a.3,b.3,c.3))—interesting, considering how these groups do not overlap, and how much they differ in their interaction patterns (Section 3.4). Here, those who consult city maps most frequently trend toward the extremes in terms of their gaze amplitudes: they have the lowest ones with the stack map, and the highest when using the grid, when compared to less frequent map readers. They appear to be more strongly adapting their behavior to the system they are using, being probably already more familiar with how to read maps.

3.5.2. Saccade Directions

Saccades and their components are also visualized in Figures A3–A5 in the Appendix B as polar plots of the density probability distributions, using the same groupings as the bar charts discussed in

this section. In addition to confirming the above findings visually, they also include information on the direction the saccadic movements take—the layout of the MCV systems become apparent in them.

Polar plots inform about saccade behaviour for the eye, the head, and the overall gaze movements (the combination of head and eye movement). The distance to the center encodes amplitudes (in degree of field of view), while the angular dimension communicates about two different aspects of saccades. One is absolute angle: the direction of the saccade relative to a horizontal plane (in our 3D case: the longitudinal axis). This information tells the reader how a participant moved their gaze relative to the environment: did observers make more saccades vertically or horizontally? On the other hand, the second set of data corresponds to relative direction of saccades [44]: the angle between two saccades.

Relative angles inform about short-term temporal dynamics of saccades: did observers plan a lot of saccades in the same directions, and are there signs of back-and-forth motion between several regions of interest? Forward saccades (located in the East quadrant) show saccades going farther towards the same direction, possibly to reach a new region of interest; it often represents a shift in attentional target. Backward saccades show a behavior where a new saccade is made toward the starting position of the preceding saccade (located in the West quadrant), possibly going back to it to analyze further.

Through absolute angles we see in the blit map system that participants produced mostly horizontal saccades to explore the map. Less head motions show in this condition, because there is no need for longer saccades to fixate another map. The grid system resulted in an horizontal bias as well as a smaller, but interesting bias showing saccades directed in all cardinal directions reaching amplitudes of approximately 20 deg. We posit that we observe the same horizontal bias as in the blit system, the cardinal bias probably indicated shifting of attention between maps with large horizontal and vertical saccades. Finally, the stack system elicited a substantial amount of vertical movements, by eye and head alike; this behavior is necessary to shift from one map to another located below or above.

When considering relative saccade directions, eye and overall gaze data show a clear bias towards backward saccades for the blit system. It is indicative of movements between regions of interest to analyze the content of the map in help of making a decision according to the task. The forward bias is minor here, as there is little need to make several saccades in the same direction as one would to orient one's attention away from the map; additionally, the maps are not so large in the field of view that they require large shifts of attention. In the grid and stack systems, we observe strong backward and forward biases. We believe that the backward bias is demonstrative of a fine analyses of a map region, while the forward bias accounts for map exploration. In all three map systems, the head rotates mostly further in the same directions (forward relative motions), it probably does not participate significantly in the fine analysis of a map's contents.

3.5.3. Fixation Durations and Saccade Velocities

When comparing the fixation durations between saccades across the MCV systems (Figure 11a), the stack system has significantly lower ones than the comparable blit or grid systems—the closeness of the layers may encourage viewers to linger less on any point and switch their gaze around more, as also seen in the higher number of gaze changes between map layers in that system (Figure 9b). The differences among the subgroup by system preference show a tendency of those who prefer the stack to linger longer on points than those who prefer the grid, and those preferring the blit system having a higher variance overall and lingering much longer on their preferred system. Frequent gamers (Figure 11(a.2)) show the least variation from system to system, with generally low fixation durations—they seem to adapt their search behaviors to each system.

The peak saccade velocities are similarly the lowest in the stack MCV, and the blit system having the highest ones, along with the largest variability, as measured by the confidence interval and the differentiation the different subgroupings show there. Those who prefer the stack are also very consistent in their peak velocities across the systems (Figure 11(b.1)), as are those who consult city maps

with the highest frequency (Figure 11(b.3)), whereas the other members of subgroups display a much higher variance in this regard—no subgroups of gamers show this kind of stability (Figure 11(b.2)) at all. These two groups (the most frequent map readers, and those who prefer the stack) do not exhibit this stability with other measurements, where their values vary greatly, such as saccade amplitudes, or the number of their gaze changes between layers. This could point to a higher confidence in the map reading behavior in these two groups, regardless of which system they use.

3.6. User Feedback

We collected a range of commentary from our participants after their task sessions in the free-form feedback phase, with each system eliciting both positive and negative remarks. The blit system received complaints regarding the amount of interaction required to compare layers, which made those comparisons take more time and be more difficult. Its bigger map view was however welcomed as making analysis of any single layer easier than with the other systems.

The simple arrangement of the grid system was mostly positively remarked on for providing a good overview over the layers. However, the size of the views caused problems in two ways: there were complaints about the smaller size of the views, as well as about the large head movements that are required to compare them.

In contrast to the comments about the other two systems, the stack mostly received suggestions for improving particular parts of it, such as the specific arrangement of the layers regarding their size, tilt, and position; with different and often conflicting suggestions from different users. Multiple participants pointed out their wish to be able to rearrange the layers in a different order. More general feedback included it offering a "practical ensemble view" of all layers, and it being the system that allowed the fastest comparisons between them.

Interestingly, the preferences of our participants, as expressed through their pairwise comparisons in the questionnaire, were not as correlated to their free-form feedback as could have been expected. Some used that text field to name their favorite systems, and it did not always correspond to its rank in the comparison matrices. A connection could be found however between the 3D video game playing habits of participants and the amount and detail of their feedback—those who played the most were generally more likely to offer constructive criticisms, often detailed. The frequent gamers also exhibited an inclination for positive comments on the stack system, or at least on ways for how to improve it, having shown promise to them—they were most interested in it.

3.7. User Comfort

We used the classic Simulator Sickness Questionnaire [47] to look for indications of cybersickness to evaluate the effects of the virtual environment and the systems within them on the well-being of the users. As a tool developed to assess sickness symptoms in flight simulators, it has also been valued for that purpose in VR research. However, it has recently (after the time of this study) been shown to be lacking in its psychometric qualities for measuring cybersickness in commercial VR applications, and updated methods are now being proposed [48,49].

We administered said questionnaire at the beginning of each session, and regularly after task completions. Most participants reported no significant symptoms. Most frequently, "none" was given for all symptoms, occasionally "slight". Only one participants complained of "severe" symptoms during the test, and these lessened as the session went on. Cybersickness typically manifests itself from the vergence-accommodation conflict, or a mismatch between visual and vestibular stimuli—when the camera, i.e., the user is shown to be moving while actually remaining seated or standing. This is not the case in our system, as the only movements of the camera or the surrounding room are direct results of the participant's movements, registered through tracking of their positions. The other movement, that of the stack layers to keep their viewing angle constant was very slight, as the participants generally did not change from their seated position. None remarked this movement in their free-form feedback.

3.8. Urbanism Use-Case Results

Requiring a free-form explanation of their choice of candidate immediately after each scenario from the users, we could not only get a distribution of these selections (Figure 12), but also gain deeper insights into why these choices were made. In a separate work [11], we conducted an analysis of the written answers, classifying each one by which map layer was referenced, and if the choice and answer made sense in the context of their scenario from the standpoint of an expert urbanist.
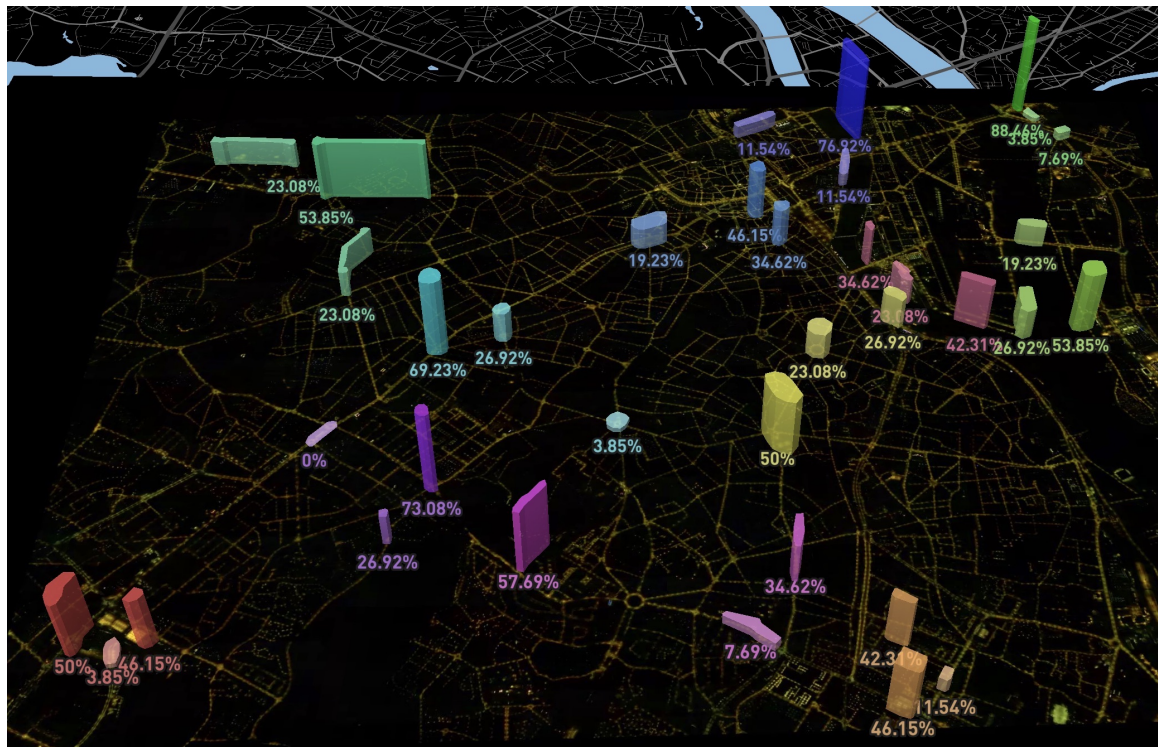
Figure 12. An overview over the 12 scenarios (color-coded) and the distribution of choices participants made between the three candidates in each.

While the characteristics of the participants, such as regarding how much time they spend outside after sunset, and which map layers they spent more time looking at were identified as having an influence on their choices, we could find no link between the quality or quantity of the answers given and the system that they used to read the maps and make their choice, or their preferences for said systems.

4. Discussion

Our study participants, as well as the experts we consulted during the implementation and pre-test phase have both expressed their enthusiasm for and voiced complaints about each of the three tested systems. By the metrics of our study, the stack MCV system prevailed in terms of user acceptance if its visual design is taken into account, and it is mostly comparable to the grid system in terms of map layer legibility and ease of use—both leaving the only temporal multiplexing system behind.

However, these subjective preferences do not paint a clear enough picture to allow us to declare a clear "winner" among the MCV systems. Neither do the objective eye-tracking metrics, as their meaning in this context remains open for interpretation—the speculations in the previous section remain attempts at just that.

Although there are significant differences in visuo-motor behavior, fixation durations or saccade amplitudes do not directly translate to quality of the system. An exception could be the amount of head movement, as plotted in Figure 10b—more movement of the still quite heavy HMD will certainly lead to more strain and potential exhaustion, if continued over longer periods [50]. In this instance,

the stack system clearly wins over the grid system in this regard—most likely due to its more compact layout, as we hypothesized. This layout is also likely to have caused the higher number of recorded changes in gaze from one map layer to another in the stack—to switch attention from the bottom layer to the one on top, the intermediary ones need to be crossed, inflating this number over that of the grid system. Even so, the number of gaze changes while using a system does not directly speak to the quality of it—higher numbers could indicate both a deeper process of comparison, as well as a higher state of confusion or disorientation, while lower ones could be indicative of more focus, or of a higher inconvenience of switching to different maps, as is definitely true for the blit system, which requires an additional controller input to switch.

Task completion times were highly variable and show no significant differentiation across the systems, which can be attributed to the differing difficulties and complexities of the scenarios, and to the fact that we did not ask the participants to complete the tasks as quickly as possible, but rather provide their best responses, even if that takes more time. As further analysed in [11], those choices and explanations were consistently sensible, with the MCV system having no statistical influence over which candidate areas were chosen in which scenario, or which map layers the participants referenced when explaining their choices.

This last point can serve as an affirmation of our design goal: implement the three systems as functionally equal as possible. The inherent differences in the three MCV systems had no impact on the quality of the insights attained through them—only on the comfort and performance of the participants while using them. This point also addresses our contribution to conducting studies on multi-layer map comparisons in VR: a setup, as reported in this work, is shown to be suitable, even insofar as providing valuable insights for research in urbanism that is only tangentially related to investigating visualization methods. With all systems performing equally in this regard, our contribution of evaluating the novel stack MCV system thus needs to be more nuanced than declaring it "better" or "worse" than the other systems we compared it against.

Our hypothesis concerning the benefits of its layout—the closer proximity of the map layers while maintaining map legibility—does bear out with the above analysis of our recorded measurements, but it appears not to be the only factor in determining its performance or acceptance among participants. Indeed, we have found the participants themselves to be the deciding factor in these matters: their characteristics, as seen through either questionnaires about their habits, or as measured by their visuo-motor behavior are significantly influential in which system they prefer or how they interact with it.

To provide an optimal experience for similar map comparison tasks, our results indicate that just optimizing and improving one system would not work universally. Different groups of the population appear to prefer opposing aspects of the different systems, depending on their experiences and behavior. A clear example for this were participants who frequently play 3D video games, and thus have probably greater experience with navigation in virtual environments. They showed a higher inclination to interact more deeply with the blitting and the stack system—they are both more "advanced" than the grid, in either requiring the use of controls to switch layers, or are arranged in a more unusual way than the simple grid. The grid, conversely, was by far the most popular system with participants whose video game playing frequency was the least, or nonexistent.

Limitations

We designed this study to evaluate each MCV concept on its own within different scenarios, without combining them, so as to better isolate their effects on user behavior and performance. Furthermore, we omitted a comparison between MCVs and a single map view containing all data in itself—such a comparison would fall strictly in the domain of cartography, and to how much information can be put into one map. We assume a case where this is not practical, and data layers have to be separated one way or another and, instead, focus on comparing different methods of such separation.

In the scenarios we employed, we also kept the number of data layers to four, in order to maintain an "even playing field" for all systems. This number provides an ideal case for both blit and grid systems: we could map all layers to the cardinal directions for switching between them, and put them in a the corners of a rectangle for a clean grid view. Higher numbers of layers would complicate such arrangements in both these systems, and potentially necessitating smaller views in the grid, while having no such obvious downsides for the stack system other than requiring a slightly stronger tilt.

Following our reasoning, stacking could accommodate a larger number of layers in the same visual space, or present the same number of layers closer to each other—even moderate inclination of the stack layers results in large savings of vertical field of view, thus having a clearer advantage over the other systems, e.g., by having more layers in the same field of view, or increasing the ease of comparison. Tilting the layers can increase the amount of horizontal space that they take up, depending on where that relative inclination occurs, by bringing the lower edges closer to the viewer. Appendix A illustrates space savings of stacks with different inclinations over coplanar arrangements and shows possible hybrid versions.

The limiting factor of how much vertical space can be saved with the stack system is the decreasing legibility of the views' content with increasing inclination. Figures A1 and A2 show different kinds of geovisualizations at different inclinations: while colored dots appear as legible at even 25 degrees, as they are at the near 90 degrees afforded by an upright grid, text labels and icons begin to lose legibility with higher distortion. In our implementation that is also a factor of the display resolution, but, in any case, further research is warranted into directly comparing the legibility of different kinds of information across inclinations and distortions, and into exploiting the virtual nature of these map views to e.g., elevate and billboard labels and icons for greater visual clarity.

Similar to limiting the number of map layers for the study, we also limited interaction with the MCV systems, especially the technically very customizable stack system. The parameters dictating the shapes and sizes of the layers, their positions and viewing angles were all kept static apart from the tracking of the stack system, in order to ensure consistent and comparable conditions for all participants. The choice of where each map layer is situated within a system may thus lead to less than optimal positioning of them, depending on the scenario and the priorities of the participants—through the free-form feedback, we received multiple wishes for flexibility in this regard.

5. Conclusions

We conducted an investigation into extending the MCV concept into VR, with a focus on geospatial visualization, and using them for comparisons of more than just two layers of data in a map. Thereby, we also have contributed to advancing the field of immersive analytics.

Basing our arguments and design choices on previous work regarding composite visualization schemes, we introduced the stack system—a novel form of MCV that takes advantage of the virtual and interactive environment current HMD technology offers. Using a task design that acted simultaneously as a participatory study in urbanism, we evaluated it in comparison to more traditional MCV paradigms. This task was carefully designed to also act as a citizen participation study in urban research, thus being directly relevant to tasks actually encountered in related domains. Using eye-tracking technology, we could analyze each user's indvidual behavior in how they read maps. Correlating their fine-grained performance and their expressed judgements, we could draw conclusions as to which kinds of MCV system are better suited for different types of users and why.

Furthermore, our analysis of a number of aspects, such as the characteristics of our participants, and how they performed with and rated each system led us to believe that no singular such measurement can sufficiently allow meaningful comparisons between such systems.

Consequently, while we did confirm our hypothesis that the novel stack MCV system has certain advantages over comparable, more traditional methods for multi-layered map comparisons, we could not arrive at a determination as to which of the three tested systems was, overall, a better choice for this kind of task.

All three systems yielded equally useful data in terms of the answers that participants provided through them to the urbanism study, but were rated very differently by the participants, and the behavior of those participants was found to have significant differences in all of the measurements we took. We identified subgroups of users, through which we could find better understandings as to which factors impact which people in what way.

In conclusion, we have discovered how preferences and priorities among users can strongly differ and oppose one another, and how their performances with the different systems can vary, based, in part, on the users' characteristics. Thus, prioritizing the development or the use of only one of the evaluated systems could certainly improve some user's experience with it, but also neglect the needs and preferences of others.

Instead, we recommend to implement a variety of different systems to give users in this kind of participatory study a choice—we found the reactions to the systems to be highly polarized, even in expert circles outside the study. An optimal user experience appears to necessitate a fulfillment of expectations and the support of existing habits, rather than education about why a certain system might technically be better than another.

With this work having shown the practicality of the stack system, and demonstrating its capacity for handling more data layers than other methods, future work could expand on that concept and go deeper into investigating more parameters about it. Features that were requested in the feedback we received during our implementation and the user study could be implemented and researched, as well as functionality adapted from other recent work, such as the "shelf" concept that was investigated in [28] (which can be approximated by creating multiple columns of our stack system) to create a hybrid between a grid and the dynamic stack system. Such a hybrid could aim at combining the advantages of all systems presented here, and be a better fit for data that can be grouped into columns, like seasonal data throughout multiple years.

The constraints on interactivity imposed during our study can be lifted for future studies more focused on interaction with the system itself, where we would allow participants to freely rearrange the layers in any order, or adjust the other parameters of the stack system, like the dimensions of the map views, their viewing angles, and distances between layers. Such future studies could also entail a different number of layers for different scenarios/tasks. With this enabled interactivity, and perhaps with even the choice of system (blit/grid/stack/hybrid) being open during each scenario, could provide insights into which configurations are better suited for which kinds of scenarios; depending on the kind and complexity of the geospatial information shown and the number of layers of data.

Author Contributions: Conceptualization, M.S. and V.T.; methodology, M.S.; software, M.S. and E.D.; validation, G.M., P.L.C.; formal analysis, M.S. and E.D.; writing—original draft preparation, M.S.; writing—review and editing, M.S., E.D. and V.T.; supervision, G.M., P.L.C.; project administration, V.T.; funding acquisition, V.T. and G.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by research programs in the Pays de la Loire region: RFI AtlanSTIC2020 and RFI OIC.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

HMD	Head-Mounted Display
MCV	Multiple and Coordinated Views
PCM	Pairwise Comparison Matrix
VR	Virtual Reality

Appendix A. Stack Layout Views and Comparisons

Figure A1 shows how the stack system compares against a flat grid of map views regarding distortion and their use of vertical space. Three views of identical size are used to illustrate the visual differences when the views are (a) arranged vertically and coplanar to each other; (b) equidistant from the viewer, and each perpendicular (90 degrees) to the line extending from the viewer to the center of each view (viewline); and (c) arranged in a stack similar to (b), but the views are titled by 40 degrees to the viewline.

As seen by the distances between viewer and points on the map views in Figure A1, perspective distortion can be unevenly distributed across views in a coplanar arrangement (1), as each has a different distance and tilt toward the viewer. Using the stack system in an extreme setting to keep each layer perpendicular and equidistant (2), perspective distortions are minimal and each layer appears in the same shape to the viewer in VR—the distortions in the camera views are due to projection to a flat screen image. This does however take up more of the field of view; it can be seen by the top layer being cut off in B2. A more moderate setting of the stack system, with the layers being tilted from perpendicularity (3) reduces the visual space all layers take up, while adding more perspective distortion uniformly to all layers.
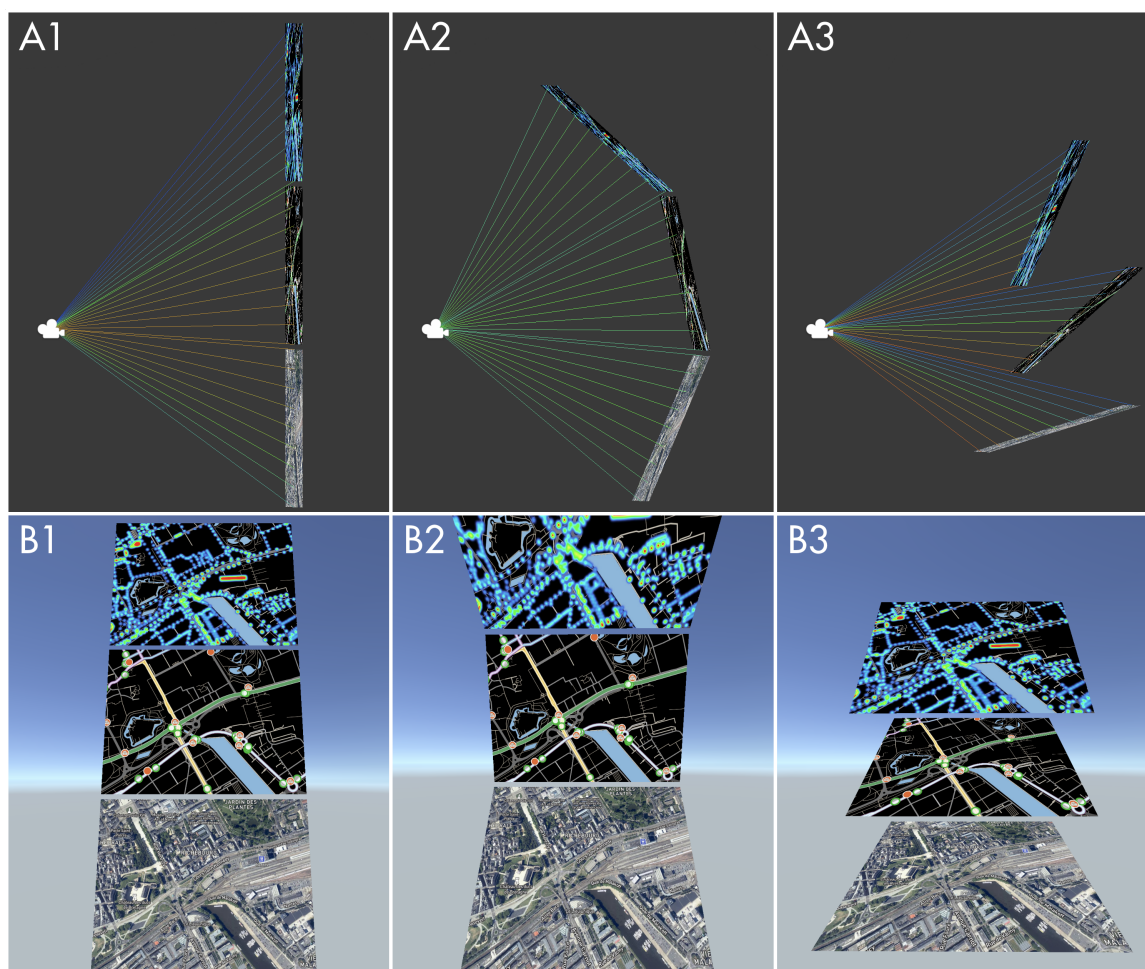
Figure A1. (**A**)bove (**A1**–**A3**): Isometric views from the side of a flat configuration of three layers (**A1**), a stack with each layer perpendicular to the viewline (**A2**), and a stack with each layer tilted at 40 degrees (**A3**). Colored lines between illustrate differences in distance between camera and equidistant points on the layers: green is the average distance, a shift to red signifies less distance, a shift to blue more. (**B**)elow (**B1**–**B3**): the corresponding camera views of the same number in (**A**), with equal fields of view (tilted up in (**B1**)). All layers equal in size.

Figure A2 shows comparisons between a grid and different stack arrangements for six layers, similarly to Figure A1. A 2-by-3 grid (1) takes up similar vertical space as a the same layers all arranged vertically in a stack of tilted layers (2), but more horizontal space. Multiple stacks can be arranged side-by-side to form a hybrid between grid and stack—similar to the "shelves" proposed in [28], but with the stack system keeping each layer tilted to the same angle. (4) shows the full functionality of the two stacks side-by-side, keeping a constant distance and view angle for all layers, and (3) a reduction in functionality that constrains the layers to all face in the same direction. (3) takes up less horizontal space than (4), at the cost of distorting each layer differently, while (4) keeps the stack system's full functionality of equal distortion of each layer at the cost of requiring more horizontal space and more distance between the stacks to avoid overlap.
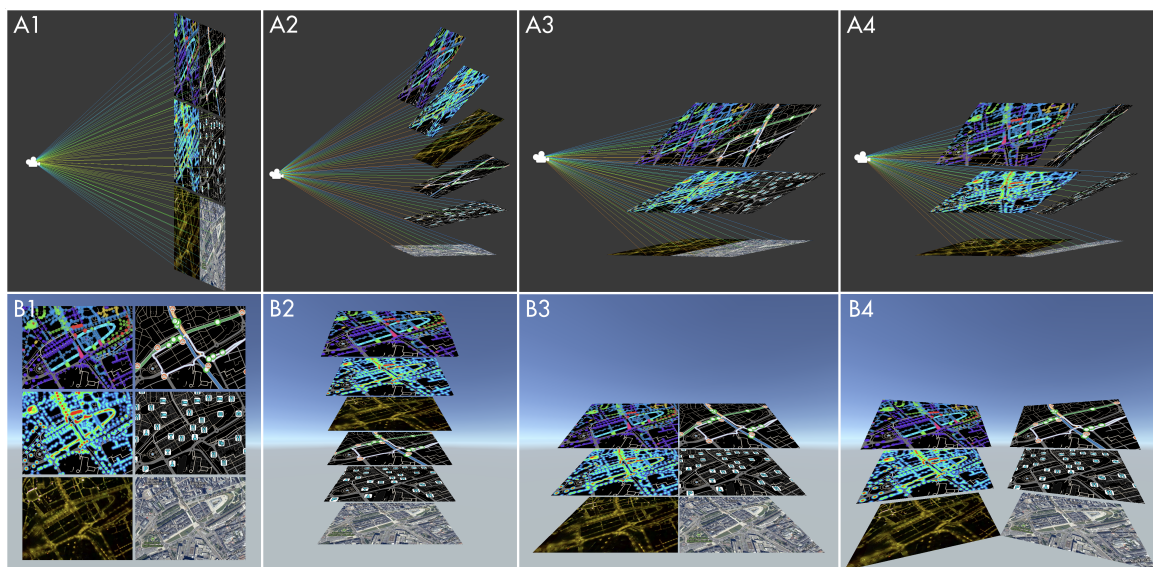
Figure A2. (**A**)bove (**A1**–**A4**): Isometric views from the side of a flat configuration of six layers (**A1**), one stack with each layer tilted at 25 degrees (**A2**), two stacks with each layer tilted at 35 degrees (**A3**), and two stacks as in (**A3**) but rotated to face the viewer (**A4**). Colored lines illustrate distances as in Figure A1. (**B**)elow (**B1**–**B4**): the corresponding camera views of the same number in (**A**), with equal fields of view (cropped sides in (**B1**,**B2**)). All layers equal in size.

Appendix B. Saccade Polar Plots
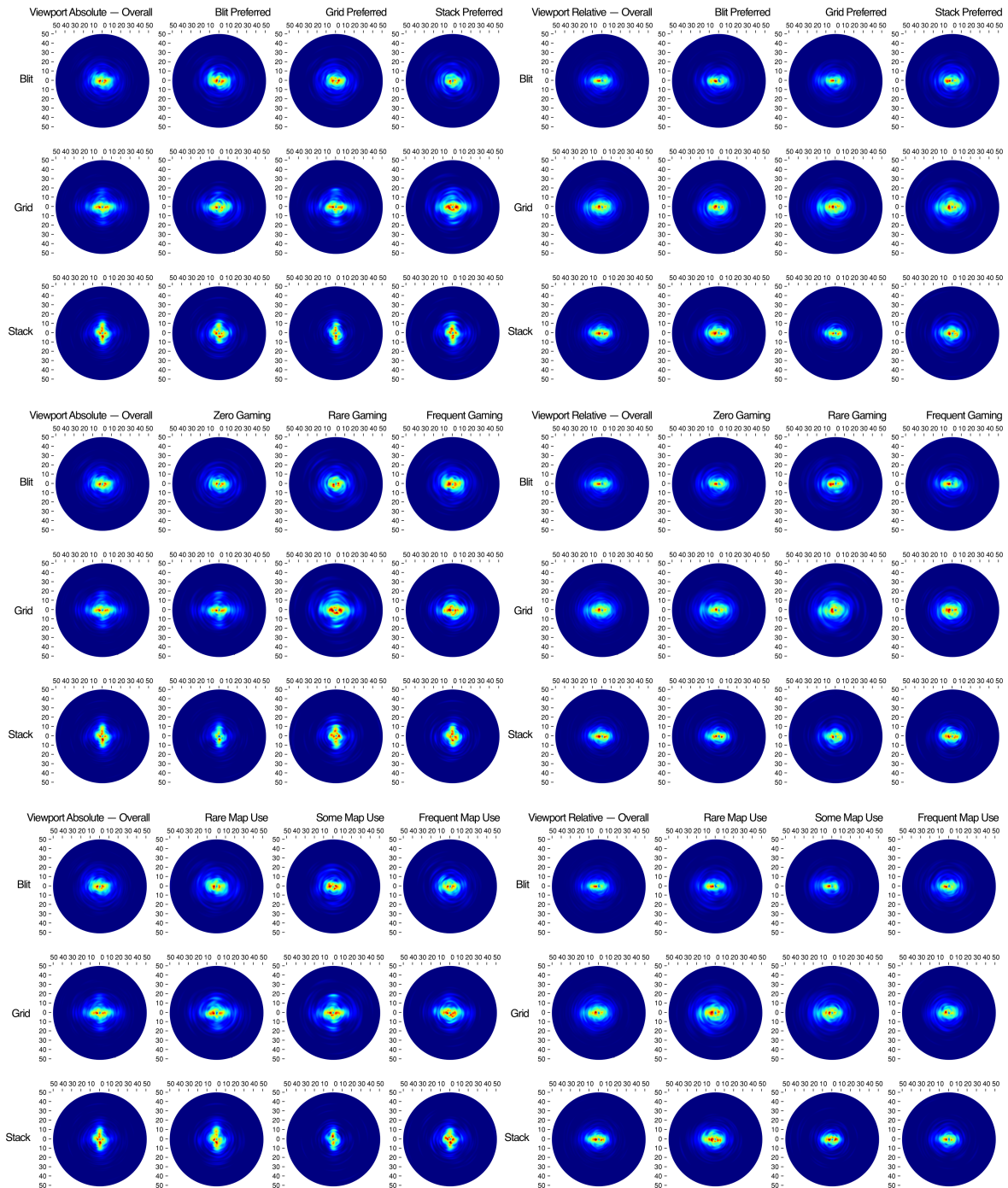
Figure A3. Polar plots showing the density probability distributions of absolute and relative saccades in the viewport—the eye movement component—in degrees, represented as polar plots, for each MCV system (blit/grid/stack) and subgrouped by the participants' preference for an MCV system regarding its legibility (**top**), the frequency with which they play video games (**middle**), and the frequency with which they consult city maps (**bottom**).
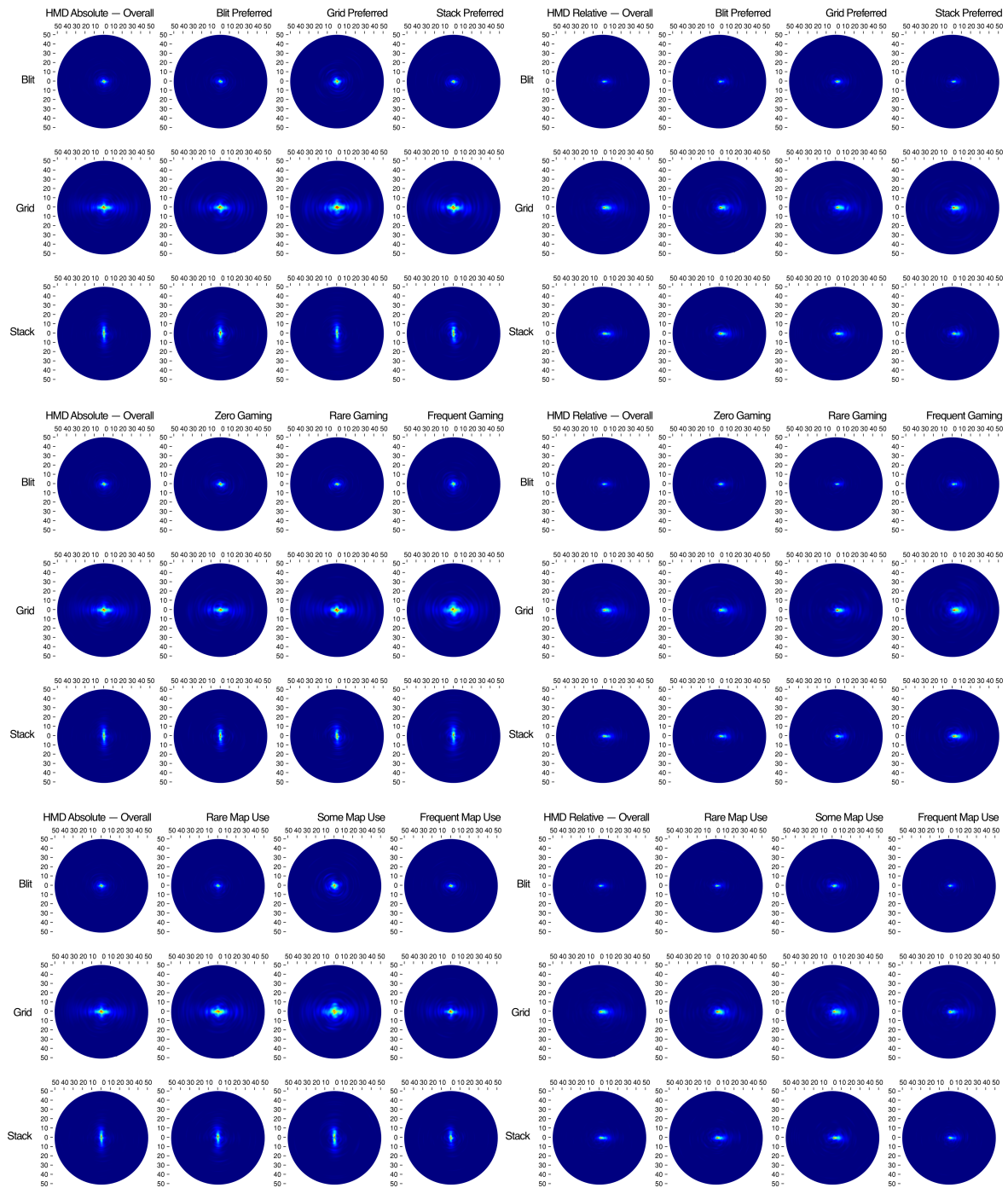
Figure A4. Polar plots showing the density probability distributions of absolute and relative saccades of the head—the HMD movement component—in degrees, represented as polar plots, for each MCV system (blit/grid/stack) and subgrouped by the participants' preference for an MCV system regarding its legibility (**top**), the frequency with which they play video games (**middle**), and the frequency with which they consult city maps (**bottom**).
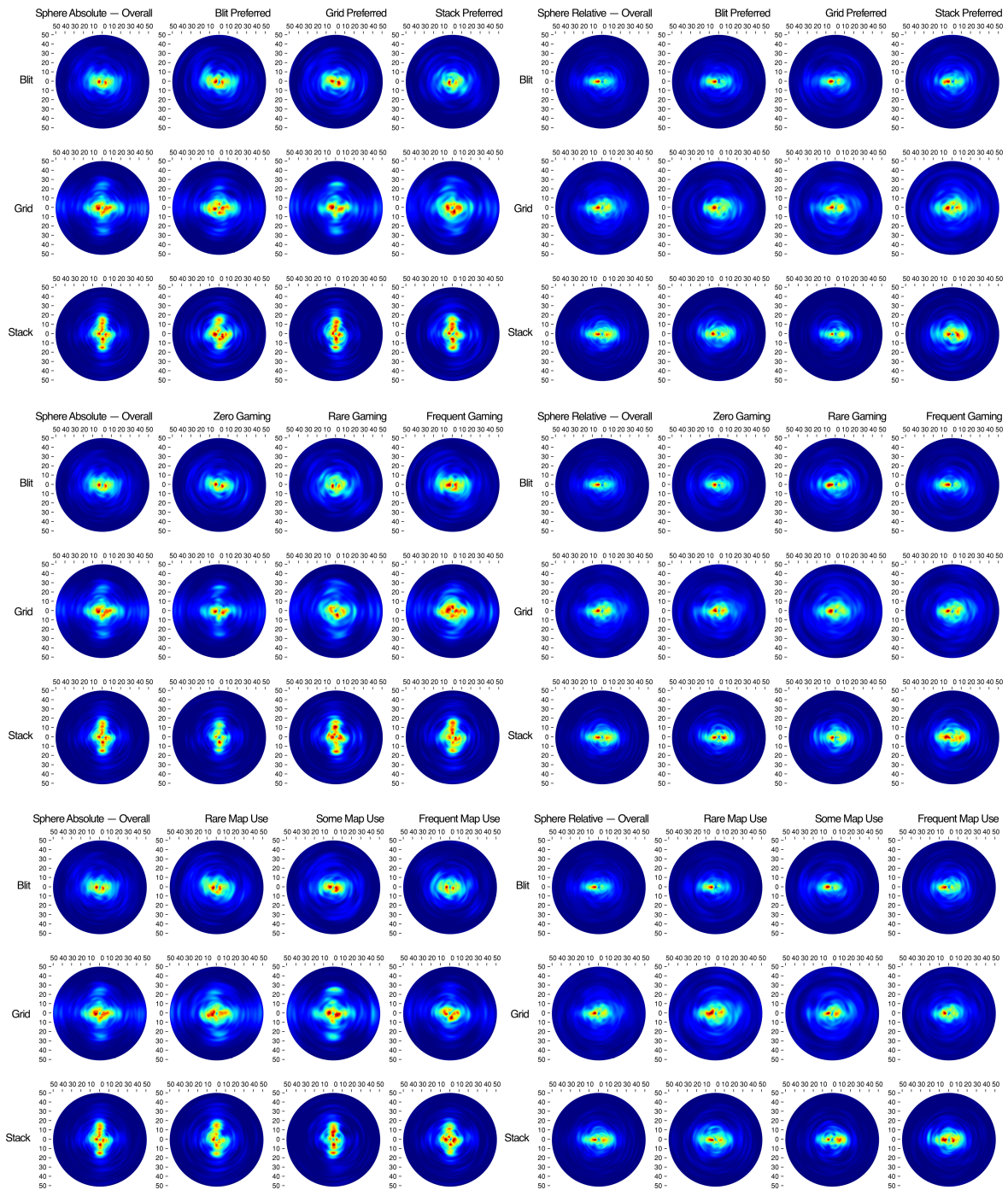
Figure A5. Polar plots showing the density probability distributions of absolute and relative total saccades—the gaze, as projected on a sphere—in degrees, represented as polar plots, for each MCV system (blit/grid/stack) and subgrouped by the participants' preference for an MCV system regarding its legibility (**top**), the frequency with which they play video games (**middle**), and the frequency with which they consult city maps (**bottom**).

References

1. Bertin, J. *Sémiologie Graphique: Les Diagrammes-Les Réseaux-Les cartes*; Gauthier-Villars: Paris, France, 1967.
2. Ware, C. *Information Visualization: Perception for Design*; Morgan Kaufmann: Waltham, MA, USA, 2012.
3. Munzner, T. *Visualization Analysis and Design*; AK Peters/CRC Press: Boca Raton, FL, USA, 2014.

4. Andrienko, G.; Andrienko, N.; Jankowski, P.; Keim, D.; Kraak, M.; MacEachren, A.; Wrobel, S. Geovisual analytics for spatial decision support: Setting the research agenda. *Int. J. Geogr. Inf. Sci.* **2007**, *21*, 839–857. [CrossRef]

5. Lobo, M.J.; Pietriga, E.; Appert, C. An evaluation of interactive map comparison techniques. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems—CHI '15, Seoul, Korea, 18–23 April 2015; ACM Press: New York, NY, USA, 2015; pp. 3573–3582. [CrossRef]

6. Roberts, J.C. State of the art: Coordinated & multiple views in exploratory visualization. In Proceedings of the Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007), Zurich, Switzerland, 2 July 2007; IEEE: Piscataway, NJ, USA, 2007; pp. 61–71. [CrossRef]

7. Goodwin, S.; Dykes, J.; Slingsby, A.; Turkay, C. Visualizing Multiple Variables Across Scale and Geography. *IEEE Trans. Vis. Comput. Graph.* **2016**, *22*, 599–608. [CrossRef]

8. Wang Baldonado, M.Q.; Woodruff, A.; Kuchinsky, A. Guidelines for Using Multiple Views in Information Visualization. In Proceedings of the Working Conference on Advanced Visual Interfaces AVI '00, Palermo, Italy, 23–26 May 2000; Association for Computing Machinery: New York, NY, USA, 2000; pp. 110–119. [CrossRef]

9. Healey, C.; Enns, J. Attention and visual memory in visualization and computer graphics. *IEEE Trans. Vis. Comput. Graph.* **2011**, *18*, 1170–1188. [CrossRef]

10. Spur, M.; Tourre, V.; Coppin, J. Virtually physical presentation of data layers for spatiotemporal urban data visualization. In Proceedings of the 2017 23rd International Conference on Virtual System & Multimedia (VSMM), Dublin, Ireland, 31 October–2 November 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–8. [CrossRef]

11. Spur, M.; Houel, N.; Tourre, V. Visualizing Multilayered Geospatial Data in Virtual Reality to Assess Public Lighting. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *XLIII-B4-2020*, 623–630. [CrossRef]

12. Javed, W.; Elmqvist, N. Exploring the design space of composite visualization. In Proceedings of the 2012 IEEE Pacific Visualization Symposium, Songdo, Korea, 28 February–2 March 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1–8. [CrossRef]

13. Zheng, Y.; Wu, W.; Chen, Y.; Qu, H.; Ni, L.M. Visual Analytics in Urban Computing: An Overview. *IEEE Trans. Big Data* **2016**, *2*, 276–296. [CrossRef]

14. Keim, D.; Andrienko, G.; Fekete, J.D.; Görg, C.; Kohlhammer, J.; Melançon, G. Visual analytics: Definition, process, and challenges. In *Information Visualization*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 154–175.

15. Dwyer, T.; Marriott, K.; Isenberg, T.; Klein, K.; Riche, N.; Schreiber, F.; Stuerzlinger, W.; Thomas, B.H. Immersive analytics: An introduction. In *Immersive Analytics*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 1–23.

16. Ferreira, N.; Poco, J.; Vo, H.T.; Freire, J.; Silva, C.T. Visual Exploration of Big Spatio-Temporal Urban Data: A Study of New York City Taxi Trips. *IEEE Trans. Vis. Comput. Graph.* **2013**, *19*, 2149–2158. [CrossRef]

17. Liu, H.; Gao, Y.; Lu, L.; Liu, S.; Qu, H.; Ni, L.M. Visual analysis of route diversity. In Proceedings of the 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), Providence, RI, USA, 23–28 October 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 171–180.

18. Ferreira, N.; Lage, M.; Doraiswamy, H.; Vo, H.; Wilson, L.; Werner, H.; Park, M.; Silva, C. Urbane: A 3D framework to support data driven decision making in urban development. In Proceedings of the 2015 IEEE Conference on Visual Analytics Science and Technology (VAST), Chicago, IL, USA, 25–30 October 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 97–104. [CrossRef]

19. Chen, Z.; Qu, H.; Wu, Y. Immersive Urban Analytics through Exploded Views. In Proceedings of the IEEE VIS Workshop on Immersive Analytics: Exploring Future Visualization and Interaction Technologies for Data Analytics, Phoenix, AZ, USA, 2017. Available online: http://groups.inf.ed.ac.uk/vishub/immersiveanalytics/papers/IA_1052-paper.pdf (accessed on 31 August 2020).

20. Edler, D.; Dickmann, F. Elevating Streets in Urban Topographic Maps Improves the Speed of Map-Reading. *Cartogr. Int. J. Geogr. Inf. Geovis.* **2015**, *50*, 217–223. [CrossRef]

21. Fonnet, A.; Prié, Y. Survey of Immersive Analytics. *IEEE Trans. Vis. Comput. Graph.* **2019**, 1. [CrossRef] [PubMed]

22. Chandler, T.; Cordeil, M.; Czauderna, T.; Dwyer, T.; Glowacki, J.; Goncu, C.; Klapperstueck, M.; Klein, K.; Marriott, K.; Schreiber, F.; et al. Immersive analytics. In Proceedings of the 2015 Big Data Visual Analytics (BDVA), Hobart, Australia, 22–25 September 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–8. [CrossRef]

23. Yang, Y.; Jenny, B.; Dwyer, T.; Marriott, K.; Chen, H.; Cordeil, M. Maps and globes in virtual reality. In *Computer Graphics Forum*; Wiley Online Library: Hoboken, NJ, USA, 2018; Volume 37, pp. 427–438.

24. Yang, Y.; Dwyer, T.; Jenny, B.; Marriott, K.; Cordeil, M.; Chen, H. Origin-Destination Flow Maps in Immersive Environments. *IEEE Trans. Vis. Comput. Graph.* **2018**, *25*, 693–703. [CrossRef]

25. Hurter, C.; Riche, N.H.; Drucker, S.M.; Cordeil, M.; Alligier, R.; Vuillemot, R. Fiberclay: Sculpting three dimensional trajectories to reveal structural insights. *IEEE Trans. Vis. Comput. Graph.* **2018**, *25*, 704–714. [CrossRef]

26. Chen, Z.; Wang, Y.; Sun, T.; Gao, X.; Chen, W.; Pan, Z.; Qu, H.; Wu, Y. Exploring the Design Space of Immersive Urban Analytics. *Vis. Inform.* **2017**, *1*, 132–142. [CrossRef]

27. Filho, J.A.W.; Stuerzlinger, W.; Nedel, L. Evaluating an Immersive Space-Time Cube Geovisualization for Intuitive Trajectory Data Exploration. *IEEE Trans. Vis. Comput. Graph.* **2019**, *26*, 514–524. [CrossRef]

28. Liu, J.; Prouzeau, A.; Ens, B.; Dwyer, T. Design and evaluation of interactive small multiples data visualisation in immersive spaces. In Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Atlanta, GA, USA, 22–26 March 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 588–597.

29. Knudsen, S.; Carpendale, S. Multiple views in immersive analytics. In Proceedings of the IEEEVIS 2017 Immersive Analytics (IEEEVIS), Phoenix, AZ, USA, 1 October 2017.

30. Lobo, M.J.; Appert, C.; Pietriga, E. MapMosaic: Dynamic layer compositing for interactive geovisualization. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 1818–1845. [CrossRef]

31. Mahmood, T.; Butler, E.; Davis, N.; Huang, J.; Lu, A. Building multiple coordinated spaces for effective immersive analytics through distributed cognition. In Proceedings of the 4th International Symposium on Big Data Visual and Immersive Analytics, Konstanz, Germany, 17–19 October 2018; pp. 119–128.

32. Trapp, M.; Döllner, J. Interactive close-up rendering for detail+ overview visualization of 3D digital terrain models. In Proceedings of the 2019 23rd International Conference Information Visualisation (IV), Paris, France, 2–5 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 275–280.

33. North, C.; Shneiderman, B. *A Taxonomy of Multiple Window Coordination*; Technical Report; University of Maryland: College Park, MD, USA, 1997.

34. Gleicher, M.; Albers, D.; Walker, R.; Jusufi, I.; Hansen, C.D.; Roberts, J.C. Visual comparison for information visualization. *Inf. Vis.* **2011**, *10*, 289–309. [CrossRef]

35. Gleicher, M. Considerations for Visualizing Comparison. *IEEE Trans. Vis. Comput. Graph.* **2018**, *24*, 413–423. [CrossRef] [PubMed]

36. Vishwanath, D.; Girshick, A.R.; Banks, M.S. Why pictures look right when viewed from the wrong place. *Nat. Neurosci.* **2005**, *8*, 1401. [CrossRef] [PubMed]

37. Amini, F.; Rufiange, S.; Hossain, Z.; Ventura, Q.; Irani, P.; McGuffin, M.J. The impact of interactivity on comprehending 2D and 3D visualizations of movement data. *IEEE Trans. Vis. Comput. Graph.* **2014**, *21*, 122–135. [CrossRef]

38. Ware, C.; Mitchell, P. Reevaluating stereo and motion cues for visualizing graphs in three dimensions. In Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization, A Coruña, Spain, 26–28 August 2005; ACM: New York, NY, USA, 2005; pp. 51–58.

39. Schulz, H.J.; Nocke, T.; Heitzler, M.; Schumann, H. A Design Space of Visualization Tasks. *IEEE Trans. Vis. Comput. Graph.* **2013**, *19*, 2366–2375. [CrossRef]

40. Theuns, J. Visualising Origin-Destination Data with Virtual Reality: Functional Prototypes and a Framework for Continued VR Research at the ITC Faculty. Bachelor's Thesis, University of Twente, Enschede, The Netherlands, 2017.

41. Guse, D.; Orefice, H.R.; Reimers, G.; Hohlfeld, O. TheFragebogen: A Web Browser-based Questionnaire Framework for Scientific Research. In Proceedings of the 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), Berlin, Germany, 5–7 June 2019; pp. 1–3.

42. Kurzhals, K.; Fisher, B.; Burch, M.; Weiskopf, D. Eye tracking evaluation of visual analytics. *Inf. Vis.* **2016**, *15*, 340–358. [CrossRef]

43. David, E.J.; Gutiérrez, J.; Coutrot, A.; Da Silva, M.P.; Callet, P.L. A dataset of head and eye movements for 360 videos. In Proceedings of the 9th ACM Multimedia Systems Conference, Amsterdam, The Netherlands, 12–15 June 2018; ACM: New York, NY, USA, 2018; pp. 432–437.

44. David, E.J.; Lebranchu, P.; Da Silva, M.P.; Le Callet, P. Predicting artificial visual field losses: A gaze-based inference study. *J. Vis.* **2019**, *19*, 22. [CrossRef]

45. Salvucci, D.D.; Goldberg, J.H. Identifying fixations and saccades in eye-tracking protocols. In Proceedings of the 2000 Symposium on Eye Tracking Research & Applications, Palm Beach Gardens, FL, USA, 6–8 November 2000; ACM: New York, NY, USA, 2000; pp. 71–78.

46. Bradley, R.A. 14 paired comparisons: Some basic procedures and examples. *Handb. Stat.* **1984**, *4*, 299–326.

47. Kennedy, R.S.; Lane, N.E.; Berbaum, K.S.; Lilienthal, M.G. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *Int. J. Aviat. Psychol.* **1993**, *3*, 203–220. [CrossRef]

48. Sevinc, V.; Berkman, M.I. Psychometric evaluation of simulator sickness questionnaire and its variants as a measure of cybersickness in consumer virtual environments. *Appl. Ergon.* **2020**, *82*, 102958. [CrossRef] [PubMed]

49. Bimberg, P.; Weissker, T.; Kulik, A. On the usage of the simulator sickness questionnaire for virtual reality research. In Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Atlanta, GA, USA, 22–26 March 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 464–467.

50. Yan, Y.; Chen, K.; Xie, Y.; Song, Y.; Liu, Y. The effects of weight on comfort of virtual reality devices. In *International Conference on Applied Human Factors and Ergonomics*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 239–248.