Exploring the SenseMaking Process through Interactions and fNIRS in Immersive Visualization

Alexia Galati, Riley Schoppa, and Aidong Lu

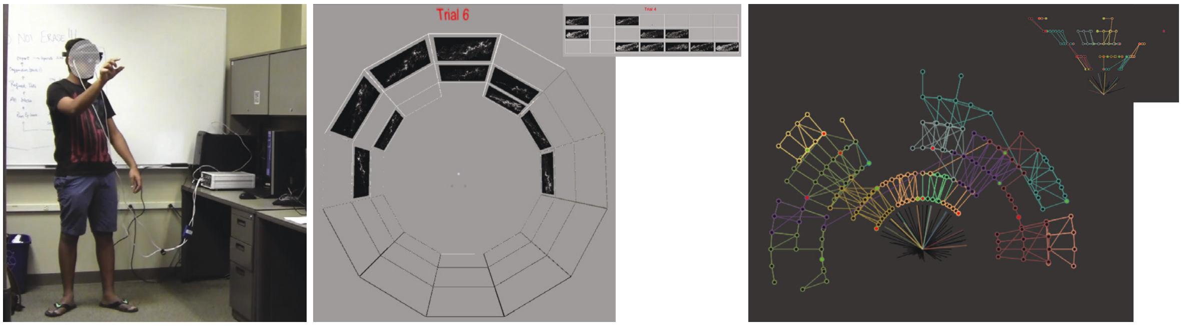


Fig. 1. This work examines user interactions during the sensemaking process in immersive visualization for spatial data clustering tasks. Snapshots show a user performing the task (left), two examples of the virtual views used as stimuli, one cylindrical and one planar layout (middle), and our visualization to explore data provenance of physical interactions in the sensemaking process (right).

Abstract— Theories of cognition inform our decisions when designing human-computer interfaces, and immersive systems enable us to examine these theories. This work explores the sensemaking process in an immersive environment through studying both internal and external user behaviors with a classical visualization problem: a visual comparison and clustering task. We developed an immersive system to perform a user study, collecting user behavior data from different channels: AR HMD for capturing external user interactions, functional near-infrared spectroscopy (fNIRS) for capturing internal neural sequences, and video for references. To examine sensemaking, we assessed how the layout of the interface (planar 2D vs. cylindrical 3D layout) and the challenge level of the task (low vs. high cognitive load) influenced the users' interactions, how these interactions changed over time, and how they influenced task performance. We also developed a visualization system to explore joint patterns among all the data channels. We found that increased interactions and cerebral hemodynamic responses were associated with more accurate performance, especially on cognitively demanding trials. The layout types did not reliably influence interactions or task performance. We discuss how these findings inform the design and evaluation of immersive systems, predict user performance and interaction, and offer theoretical insights about sensemaking from the perspective of embodied and distributed cognition.

Index Terms—Sensemaking, user behavior, immersive analytics, cognition load, mixed reality

1 Introduction

Previous work on sensemaking and distributed cognition has guided the interface design of many computer systems. As a high level mental process, sensemaking involves the analysis of facts and inferences, the integration of learning from different areas, creative thinking, and the evaluation and judgment of information [1, 15]. From the perspective of distributed cognition [39], sensemaking unfolds in a spatial environment. Within the environment, spatial arrangement serves different functions that support sensemaking, namely by simplifying choice, simplifying perception, and simplifying internal computation [27].

Our understanding of sensemaking and distributed cognition is essential to the design of computer systems, including immersive visualization systems that aim to assist the sensemaking of various information. Compared to traditional desktop settings, immersive systems built with VR/AR have several capabilities that could improve the sensemaking process. Immersive systems assist distributed cognition by presenting immersive environments designed appropriately and by allowing physical interaction through multi-sensory channels that are generally more intuitive for users.

The authors are with the University of North Carolina at Charlotte. E-mails: {agalati, rschoppa, aidong.lu}@uncc.edu

Manuscript received 9 Sept. 2020; revised 15 Dec. 2020; accepted 8 Jan. 2021. Date of publication 22 Mar. 2021; date of current version 7 Apr. 2021. Digital Object Identifier no. 10.1109/TVCG.2021.3067693

At the same time, immersive systems provide new methodological tools for studying the sensemaking process because they permit studying behavior at a fine-grained level that could not be reached without the latest tracking and analysis technology. Immersive technology now permits examining how different uses of space come into play during sensemaking, with data sampled frequently from both the spatial and temporal dimensions. Examining users' interactions (e.g., how they orient their visual attention in space) can illuminate their thinking process [1, 38]. Effective analysis of these data will help us study fundamental cognitive processes and develop advanced methods for creating personalized interaction models.

The present work explores the capability of immersive systems as a tool for unveiling the sensemaking process in immersive environments. Our primary goal (G1) was to investigate how user interactions depend on task constraints (cognitive load and layout type) in a visualization task in immersive AR, with neural activity as an additional index of cognitive load. Our secondary goal (G2) was to develop a visual analytics system that permits researchers not only to assess the users' interactions and the neural indices of their cognitive processing individually, but also to perform joint analysis for the intricate relationships among these sources of data.

Our specific research questions with respect to G1 were: (Q1). How do the users' interactions differ across different types of spatial arrangements and different levels of task difficulty? (Q2) How do the users' interactions change over the process of sensemaking? (Q3) How do the users' interactions predict sensemaking performance (accuracy on the task)?, and (Q4) How do the users' neural activity, which we take to indicate their cognitive processing, relate to their interactions in the task and task performance?

To answer these research questions, we designed and performed a user study involving visual correlation, a classical visualization problem [9]. Visual correlation relies on sensemaking because it requires evaluating data similarity through visual comparison and making decisions by clustering subsets of similar data. By examining visual correlation, we can gain insights about other visualization tasks that share common cognitive underpinnings (e.g., as evaluating data similarity, [40]). We developed an AR system (the *user's interface*) for performing clustering tasks with interactions. The user interactions we considered include direct, external interactions with the virtual artifacts in an immersive system (e.g., the user moving images) and internal behaviors that do not involve the direct manipulation of virtual artifacts (e.g., the user rotating their body; their gaze fixations on images). We manipulated two variables that would allow us to tap into embodied and distributed cognition: the cognitive load incurred by the task (two challenge levels with different numbers of images) and the type of spatial layout (two commonly adopted interfaces in immersive environments: planar 2D vs. cylindrical 3D). With respect to cognitive load, we reasoned that with more images the number of comparisons needed to be made to evaluate data similarity would increase, making it those clustering judgments more cognitively demanding.

To assess the users' processing under different cognitive demands, different layout types, and different time points in the trial (G1), we collected data from three different channels: AR HMD, video, and functional near-infrared spectroscopy (fNIRS). In statistical models, we addressed our research questions by examining how user interactions and measures from fNIRS were predicted by cognitive load and layout type, how they predicted user performance, and how they changed over the course of the trial.

To further support joint analysis of these rich datasets (G2), we developed a visualization system (the *researcher's interface*) that permitted exploring the relationship between user interactions, their neural activity, user performance across task features (e.g., cognitive load, layout time), and across task epochs (time periods in the trial or the task). By elucidating how interactions and neural signatures in immersive analytics aid sensemaking, this work can inform theories of embodied and distributed cognition [12, 23, 27] and demonstrate the potential of analyzing user interactions to evaluate and predict user performance in future immersive systems.

2 Related Work

We review related work on the effects of spatial arrangement (including layout type) on sensemaking in immersive analytics, the relevance of cognitive load to the sensemaking process, and use of fNIRS as indices of the sensemaking process.

2.1 The Spatial Arrangement of Information Influences Sensemaking

There is robust evidence that spatial arrangement of information in the task environment influences cognition. For example, in line with Kirsh's [27] insights about the functions of space, Andrews et al. [1] showed that analysts used space to encode the relationships between multiple elements, such as data, documents, display, and analyst. The space provided a semantic layer as a form of rapid access to external memory, which supported the analysis process. The distribution of information across space has also been shown to impact learning performance and the strategies used for learning from abstract data [16]. With distributed layouts, participants maintained better memory of the locations where information was presented. Similarly, Geymayer et al. [18] showed that the availability of sensemaking tools in the environment affected the use of display space, consistent with theories of distributed cognition. For immersive visualization, Batch et al. [4] studied space use and embodiment with 3D scatter plots in VR. The results showed that participants did not use the full available 3D space, which supported the most adopted design of immersive interface on 2D and 3D surfaces. Other work has further examined the use of space from a distributed cognition perspective: Mahmood et al. [35] explored the organization of subspaces, and Lee et al. [32] also examined the uses of space in collaborative visualization.

A consistent finding is that users' performance [40] and subjective satisfaction [1, 7] is improved when using large displays compared to small displays in analysis tasks. Large display users employ sophisticated strategies to exploit the available space for spatial cognition, such as dividing the space into focus and context areas [7, 19], placing application windows as reminders [22], and using windows for clustering or piling [1, 46]. Large displays can therefore be thought to act as externalized memory, as users employ the space to organize and memorize information [1]. Similar to large displays, immersive systems can utilize large physical spaces to create effective interfaces.

Despite evidence that immersive technologies improve performance and reduce cognitive load [6], it should be noted that some of the inherent restrictions of head-mounted displays (e.g. the limited field of view) can increase the users' cognitive load. It is therefore important to identify tools, affordances, and interaction strategies that can reduce that load in immersive technologies.

2.2 Sensemaking in Immersive Analytics

With the development of immersive technology, complex information processing has become a common application of immersive systems. Immersive analytics [11] extends the classical desktop visualization into a variety of immersive environments and enables new analytics capabilities. While still in its early stages, immersive analytics has attracted the interest of many researchers. This is demonstrated by a surge of recent work using virtual or physical 3D space (and interactions with that space) to explore cognitive processes during data analytic tasks [13,29,33,44,51,52]. Researchers have also started to evaluate the effectiveness of immersive visualization techniques. Recently, Lages and Bowman [31] identified desirable properties of adaptation-based interface techniques for augmented reality workspaces centered around walking. Whitlock et al. [49] examined how visual analytics tools can transform field practices by more deeply integrating data into these operations. Liu et al. [34] studied 2D small-multiples visualization in 3D spaces and suggested that flat layout or semi-circular layout over fully surrounding.

Interaction in immersive systems is often achieved through multisensory channels including gaze, voice, and gestures that are significantly different from desktop programs. The physical movements and interactions are often mixed and coordinated during the use of immersive systems, which could affect the users' performance. Immersive analytics demonstrates advantages, especially with stereoscopic techniques producing favorable results [13, 14, 29, 48]. For example, during immersive navigation [50] in VR and AR environments performance is significantly improved on tasks that include tracking, matching, searching, and ambushing objects of interest. Interaction logs have also been explored for evaluation and understanding analyst behaviors [20, 45]. In all, understanding how users apprehend information and how they deploy interactions in 3D is necessary for designing better immersive analytics systems.

2.3 Assessing Cognitive Load in Immersive Systems

Cognitive load refers to the working memory resources used in a task. These working memory demands stem from the amount of information that one has to hold in working memory (intrinsic cognitive load), the format or complexity of the instructed actions (extraneous cognitive load), and the work one puts into processing and representing that information for the long-term (germane cognitive load) [8]. Cognitive load has been examined in the context of information gathering and reasoning in complex virtual environments [41].

Cognitive load has also been examined in the context of examining the potential cognitive cost or benefit of using augmented reality displays [6]. There is evidence that immersive technologies can be employed to reduce extraneous cognitive load to make it easier to comprehend instructions, thus optimizing the germane load. In a study comparing three different augmented reality display technologies (spatial augmented reality, the optical see-through Microsoft HoloLens, and the video see-through Samsung Gear VR), spatial augmented reality increased performance and reduced cognitive load [6].

In this work, we examine whether different task features and user behaviors alleviate cognitive load and thus support sensemaking. We operationalize cognitive load in two different ways. First, we experimentally manipulate cognitive load by using trials with different numbers of items (i.e., manipulating intrinsic cognitive load). Second, we measure how cognitive load fluctuates in a given trial: we take the users' neural activity from fNIRS to indicate their cognitive demands over the course of the trial. In the next section, we explain in more detail how fNIRS can signal cognitive effort in a task.

2.4 fNIRS Data as Indices of Cognitive Load

As a non-invasive method of measuring brain activity, fNIRS has been used to study cognitive load in a variety of applications, such as real-time monitoring of mental workload of airline pilots [10] and the performance of participants during video-game like tasks [24]. Aydöre et al. [2] found that fNIRS from the prefrontal cortex (PFC) were highly dependent on cognitive load, with functional connectivity increasing with increasing cognitive load. Similarly, De et al. [17] found systematic relationships between fNIRS and cognitive load in symbol-meaning associative learning tasks: using an interval type-2 fuzzy classifier, the participants' fNIRS features could be reduced and classified to three levels of cognitive loads.

Because fNIR data are highly corrupted by measurement noise and physiology-based systemic interference [42], several statistical analysis methods have been used to extract neuronal activity-related signals from fNIRS [36, 42], including motion artifact correction, short source-detector separation correction, principal component analysis (PCA), independent component analysis (ICA), false discovery rate (FDR), serially-correlated errors, and inference techniques. More recently, machine learning methods have been applied to fNIRS analysis [26]. To resolve the noise issue, researchers often combine the analysis of fNIRS with video recordings and additional measurements [21, 37].

In this work, to explore the relationship between the users' interactions (obtained from Hololens) and their fluctuating cognitive load demands (captured through fNIRS) in immersive systems, we examine the distribution of signals from the two sources in spaces constructed through dimensionality reduction (PCA).

3 User Study of Data Clustering in AR

3.1 Study Design

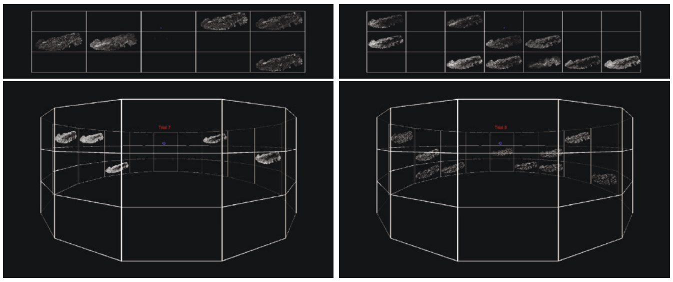
The main task of our study was a visual correlation task, where participants group, match, or cluster images based on visual features. The interactions we focused on were derived from the users' body movement and the spatial organization of their actions: how they moved images and how they oriented their attention in the displays.

To evaluate the cognitive impacts of different layouts, we manipulated two factors of interest across 8 trials. The first factor is layout type, plane (2D) vs. cylinder (3D), shown in Figure 2. The two types of layouts are commonly used as the information interfaces for immersive systems, as they are not affected by occlusion issues [4, 31]. Images were all placed at the center of each regular grid and filled the space of each cell while preserving the original image size ratios.

The second factor is cognitive load – half the trials of each type of layout (planar and cylindrical) involved low cognitive load and the other half high cognitive load. Each layout style was further divided into small and large grid size based on number columns. Displays with fewer columns (5 in planar or 12 in cylindrical layouts) were always used in low cognitive load trials (with 5 images), whereas displays with more columns (7 in planar or 16 in cylindrical layouts) were always used in high load trials (with 10 images). The sizes and numbers of grid cells were determined based on a pilot study to ensure that images would be large enough to observe without requiring body movement toward them.

3.2 Immersive Visualization System for Study

We developed an immersive analytics system for visualizing and interacting with image clusters through physical interactions (the *user's interface*). We used Microsoft HoloLens, which supports 3D augmented rendering and interaction with hand gestures. Figure 1 shows a user performing the study. Images were laid out surrounding the user, who was defined as the center location. Each spatial location was associated with no more than one image at a time (i.e., images did not overlap). Users could move around the space and rotate their view to any direction in the 3D space.



sparse (low cognition load) dense (high cognition load)

Fig. 2. Our study uses four types of layouts for arranging images: sparse planar layout, dense planar layout, sparse cylindrical layout, and dense cylindrical layout.

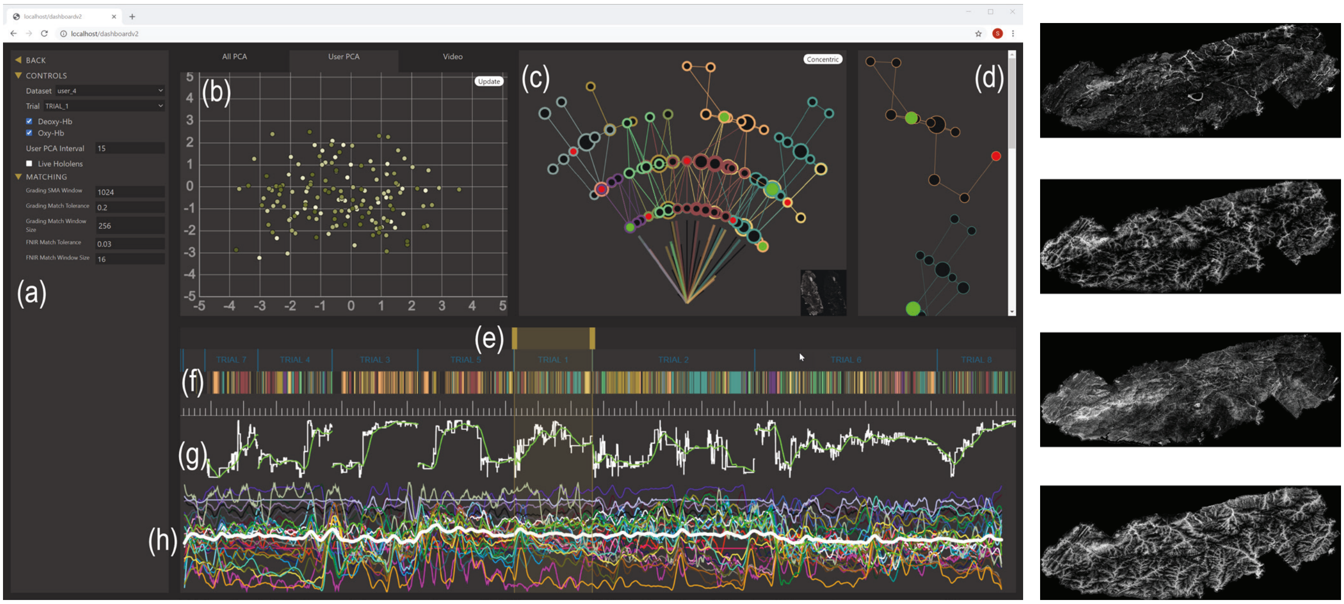
Layouts. We generated a 2D planar structure to simulate an environment of large display that permits 2D visualization and interaction (see Figure 2). The images were placed 1-meter away in front of the user on the grid structure. Similarly, we generated a 3D cylindrical structure by forming a circle with a radius of 1 meter, using the initial user position as the center. Images were displayed at the center of each cell, retaining their 2D shapes without distortion to ensure that they were visually identical in both layouts.

Multi-modal Interaction. We developed a set of multimodal interactions that support participants to perform tasks by combining voice, gaze, and hand gestures. Participants used voice commands to advance through the trials of the study: they started the session by saying "Start Trial" and advanced by saying "Next Trial", until the study was finished. Another interaction was to move images around the grid, simulating the regular "drag and drop" function. This was achieved by selecting an image with hand gesture "air tap", moving the image with the "air tap down" gesture, and placing the image by releasing the gesture. Upon completing the "drag and drop" gesture, a snap function positioned the image automatically at the center of the closest cell in the grid. To implement this we calculated the euclidean distance of the image with every cell in the grid and positioned the image to the center of the cell with the minimum distance. The images were also oriented toward the center automatically. The users' gaze icon changed from a colored sphere to a hollow sphere when they could interact with an image. The system was robust at recognizing the users' commands, and all the participants were given unlimited time to practice the interaction.

3.3 Materials

The images used in the user study were selected from around 1000 heat-maps of different species (plants and animals) in Great Smokey Mountain National Park. These heat-maps represent recorded observations of habitat distribution of the different species as 2D data distribution (see examples in Figure 3). All these images were similar, in that they concerned the same terrain, but had no shapes or patterns that may appear in real photographs. Thus, this rich set of images permitted us to pre-select manually images with different levels of similarity that could be clustered in several groups.

Each image was only used in one trial, thus minimizing effects of learning, memory, or interference from images from previous trials. The initial layout of each trial was predetermined, with its allocated set of images positioned on the same locations and the remaining cells being blank. Low cognitive load trials had only two or three clusters, with clear differences. In contrast, high cognitive load trials contained three to five clusters with more diverse images and with the differences among clusters being less clear.



system interface examples

Fig. 3. Our visualization interface has several components for analyzing a set of data collected in the study: (a) parameter panel, (b) PCA distribution of fNIRS sequences for overview and exploration of unique behaviors, (c) spatial visualization of user interaction from HoloLens, (d) patterns selected from (c) for search, (e) time selection by users, (f) time bar with a quick view of image interactions where each color represents an image (the same with colors in (c)), (g) our automatic evaluation of user results, and (h) fNIRS sequences. The right images are example data used in the study.

3.4 Participants

Participants (N = 14) were male (9) and female (5) college students between the ages of 20 - 28 years (the average age was 23).

3.5 Procedure and Tasks

The user study started with a practice trial, which used the same immersive visualization system but with a different set of images. Participants familiarized themselves with the system, practicing all the voice and gesture interactions, until they were ready to begin the user study.

They then proceeded to complete the 8 trials of the user study. The order of the trials was randomized for each participant; this was done to minimize practice effects, since participants might become more familiar with the image types and develop problem-solving strategies over time. As noted, different images were used across trials.

For both practice and experimental trials, participants were asked to cluster the images based on their visual similarity. Participants were told that their answers would be graded based on the accuracy and that they are expected to cluster the images as accurately as possible. They were not provided information about the number of clusters or possible clustering results. They were simply instructed to form any clusters they thought were appropriate while explaining their decisions out loud, so that their answers could be recorded on video.

3.6 Apparatus

Our study used Microsoft HoloLens and fNIR devices simultaneously. As shown in Figure 1, participants first wore the fNIR headband on their foreheads and then the HoloLens on top of it. The fNIR headband was connected to the fNIR device with a cable about 2 meters long, restricting the movement area of participants, which motivated our selection of large images for easy observation. Still, participants could turn to either side or to the back freely. The camera was set in front of the participants. All data were captured at fixed rates: 30 frames per second (fps) for HoloLens and video, and 2 fps for fNIRS.

3.7 Predictions

Based on theories of distributed and embodied cognition, we predicted that users would produce more interactions on the high (vs. low) cognitive load trials, as a way to offload cognitive demands to the environment. We also predicted that increased user interactions would benefit task accuracy, especially on the high cognitive load trials. Our predictions about the effect of layout type and the effect of time (beginning, middle, or end of the trial) were exploratory. Finally, we expected fNIRS to indicate cognitive load (i.e., differ across the high and low load conditions), and expected that fNIRS would exhibit systematic relationships with user interactions. However, the specific nature of these relationship was exploratory. These relationships would be explored with the visualization system we developed, described next.

4 Visual Analytics of the Sensemaking Process

4.1 Design Goals of Visualization system

To further analyze and examine the rich combined data from users' interactions and neural signals, we developed a specialized visualization system (the *researcher's interface*). As shown in Figure 3, our visualization system is composed of several panels for examining different aspects of the collected data. This system has the following key features:

- F1. Enable interactive exploration of various patterns of spatial sensemaking behaviors from a group of users performing a visual analysis task (e.g., see panel d in Figure 3).
- F2. Explore the relationship between the users' interactions, their fNIRS sequences, and task performance over the course of sensemaking (e.g., in a given trial; see panels f, g, h in Figure 3).
- F3. Identify important cognitive activities from all the data sequences, by combining HoloLens behavior data with neural signals from fNIRS (see panels b, c, and d in Figure 3).
- F4. Compare the interactions of different users and identify behaviors that were effective in terms of performance (see Figure 5).

4.2 Visualization of Interaction Sequences

We first describe the key components of our visualization system for studying individual users' datasets and then describe how the data of multiple users can be compared and explored in the section 4.3.

4.2.1 Visualizing Spatial Behavior from HoloLens

Using Microsoft HoloLens, we recorded all movements and interactions of the users. For any change in user position, orientation or head movement, we recorded the image the user was gazing at, the position of the image within the grid (in form of row and column number of the cell within the grid), the normalized position of the user in the world space, the forward vector of the user's head capturing their orientation, the up vector capturing changes in head movement, and the time of the interaction since the start of the study.

We first extracted an event list from the HoloLens data by removing continuous records with the same behaviors. This captured time durations where the user did not move. This step can significantly reduce the size of the remaining HoloLens data, and simplify the process of searching for similar interaction patterns. As shown in the Figure 5, visualizations are composed of multiple lines that mark the user's position and view direction. During our study, users rarely walked around, so the lines resemble a pie shape chart. The lengths of the lines are adjusted to the durations of the record, with longer lines indicating the directions at which the user spent more time.

Although the HoloLens data recorded 3D interactions, we designed a $3D->2D$ mapping to visualize the 3D record data in a 2D layout, without the need of 3D operations. Since all users performed the tasks in the standing position within a small region, we simply ignored the z value of the user's location. This also facilitates the comparison of the two layout types. The 3-level grids surrounding the user are therefore flattened. Each level is projected to a circle, with the lowest level (bottom row) closest to the center. This projection maintains the angles between the user and all positions on the grid. The positions of all images can be clearly shown in this fashion. The movements of an image on the grid are also captured, indicating the user's 3D interactions. The same mapping was applied to both planar and cylindrical layouts.

The system can be used to visualize the movement of individual images by hovering one's mouse over the representation of the grid. Alternatively, different time durations can be chosen to observe the user's interactions with all images in that period. The traces of individual images can also be selected and highlighted, shown in Figure 5.

In addition to its interactive visualization functions, the system permits automatically measuring several types of interactions that we have found to be relevant based on interactive exploration of user records. We automatically extracted several types of interactions for statistical analysis. Our metrics for each trial of each user included: the number of all images moved (consecutive moves defined by appeared less than 3 seconds are counted as one move), the duration of user gazing at any image (measured in seconds), the total length of all images moved, and the degrees of all body rotations. These measurements provide data for quantitative analysis we have performed in the section 5.

4.2.2 Measuring Cognitive Activity from fNIRS

The sequences of fNIR data are taken to index the fluctuations in cognitive load demands over the course of the trial, since they capture of the activity levels at corresponding brain regions [24]. Because fNIR data are often mixed with measurement noise from other factors, such as the user's movements, we applied the following preprocessing, feature extraction and dimension reduction steps.

The data obtained from the fNIR apparatus, recording the user's brain activity during the task, are represented as a time sequence of 32 signals. We separated the data into oxyhemoglobin (HbO) and deoxyhemoglobin (HbR) groups, each containing 16 fNIRS. The max and min of HbO and HbR were taken from all trials of all the users to ensure the features we measure were comparable among different users. We further preprocessed the fNIRS with a noise removal step, applying a digital filter of Savitzky-Golay with a pass band of (0.1-0.4 Hz) to eliminate the undesirable signals in that band.

From the 16 fNIRS pairs, we computed a set of target features capturing the central tendency, variability, and symmetry of the distribution of the data, previously used to indicate cognitive load [17]. Each fNIRS pair was measured with the following seven features (for a total of $16 \times 7 = 112$ features):

- F1: Mean values of oxyhemoglobin (HbO) concentration;

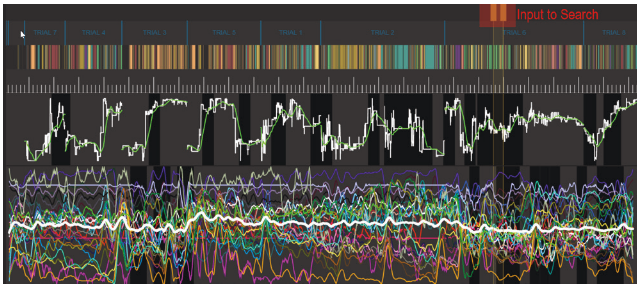


Fig. 4. System users can adjust the time range and search for durations with similar behaviors, which are shown with the black background on both evaluation curve and fNIRS.

- F2: Mean values of total hemoglobin (THb) concentration;
- F3: Mean values of oxygen demand (HbO-HbR) concentration;
- F4: Standard deviation of HbO concentration;
- F5: Standard deviation of THb concentration;
- F6: Skewness of HbO concentration;
- F7: Skewness of HbO-HbR concentration.

We then applied dimensionality reduction, through principle component analysis (PCA), to produce an interactive visualization of data distributions (see panel c in Figure 3). The system provides visualizations of all users' data or of individual users' data, permitting exploration of the changes of fNIRS across different time frames. In the system, one can choose one or multiple time durations, such as all planar (2D)/cylindrical (3D) trials or first half of all trials to explore different clusters of activities. Figure 9 shows an example of trial one from all users: one can use their mouse to hover over for detailed information or select a node to update the time range indicators and HoloLens views to further explore corresponding physical behaviors.

To explore the quantitative relationships between user interactions and cognitive activities, the system can be used to divide the PCA space into zones and measure those user interactions listed at the end of section 4.2.1 for each zone.

4.2.3 Quantifying Task Performance

The users' performance was automatically graded during the task. Compared to manual grading at the end of the study, the advantage of automatic evaluation of performance is that it can be computed for the entire duration of each user study, allowing us to explore the effects of user interactions on the sensemaking process.

The evaluation of user performance was based on the image positions on the grids and on image similarities. Since all the images in our study share the same structure (the same terrain shape), we estimated the image similarities with the pairwise pixel color differences. Specifically, for each pair of images i, j in the trial, the image similarity $s(i, j)$ was computed by the accumulated absolute differences of pixel values between two images. The distance $d(i, j)$ is the euclidean distance between two image center locations. The result of image layout is evaluated as the summation $\sum_{i,j} s(i, j) \times d(i, j)$ from all image pairs at each time stamp. Figure 4 shows an example of the estimation values – the white curve covering the entire duration of a user study. To balance the quick changes of evaluations, we use the averages of local values, rendered as the overlaying green curve, to identify the trends of evaluations and search for similar durations.

4.3 Interactive Exploration of Users' Behaviors

4.3.1 Interactive Exploration Functions

We developed interaction functions to perform interactive exploration for each type of data in the panels of HoloLens, fNIRS, and the PCA. Each panel allows overview of the entire time duration, for any selected subset of time ranges, subset of the data points (users or trials), or subset of channels (e.g., specific fNIRS features).

To enable joint analysis among the HoloLens, fNIRS, and evaluated performance records, we developed interactive functions for exploring similar durations based on selected data features (see Figure 4). Each visualization panel permits interaction functions for time windows of interest. Additionally, movement patterns of interest associated with a single image can be isolated, by clicking on the node for that image in the HoloLens panel.

Moreover, the system can be used to search for all other time durations with patterns with similar trends. The data features can be selected from any of the three panels and recorded on the search panel on the right of the interface (see Figure 3, panel d). For fNIRS or graded performance, we divide the time window to a number of small sections and search for other time windows with the same value distributions, such as the average value for each section. For HoloLens patterns, we use the relative positions/directions as the search criteria. We allow a threshold value to be adjusted for enlarging the range of similar patterns. As Figure 4 shows, the returned similar durations are highlighted on the timeline with a black background on the evaluation and fNIRS panels.

5 Analysis of Interactions in Immersive Analytics, Task Constraints, Task Performance, and FNIRS

To address our research questions, we first assessed the effect of layout type and cognitive load of the users' interactions in Hololens (Q1). We focused on the following behaviors: the number of times users moved images in each trial, the total distance images were moved, the number of times users switched their looking location in a trial, and the users' looking durations. We also assessed the length of the trial. Then, we explored how users' interactions changed over the course of the trial (Q2). Next, we examined the relationship between interactions in Hololens and their automatically evaluated performance (Q3). Finally, we explored the relationship between the users' neural signatures and their interactions (Q4) through qualitative and quantitative analyses within the visualization system.

5.1 Statistical models

To examine these questions, we built linear mixed effects models, fitted using the *lme4* package [5] in R [43]. Models included fixed effects for cognitive load (low vs. high), layout type (planar vs. cylindrical), and their interaction, and random effects for users. Trial order was included as a covariate in each model. All models started with the maximal random effect structure included intercepts for users, as well as random slopes for cognitive load, layout type, their interaction, and trial order, to account for between-participant variation for these effects [3]. For models with additional fixed effects (gaze location; performance range), we specify their structure below. If a model failed to converge, we simplified it by removing terms from the random effect structure in a theoretical motivated manner starting with the higher order term (the interaction of cognitive load and layout) and the fixed effects of least theoretical interest (e.g., trial order). The p-values were obtained from the *lmerTest* package [28] using the Satterthwaite's method. Captured variance of overall models is reported as Conditional R^2 variance explained by fixed and random factors together, computed using the *MuMIn* R statistical package [25]. The full output of the models is provided in tables in the accompanying supplement.

5.2 Effects of Cognitive Load and Layout on Interactions

We first assessed the effects of the cognitive load of the trial and the layout type on three types of interactions in models for: the number of times images were moved in each trial, the total distance all images moved in a given trial, and the users' looking durations. We had predicted higher levels of user interactions on high (vs. low) cognitive load trials, while our predictions about the effect of layout were exploratory given mixed results in the literature. As illustrated in Figure 6 (top two panels; see also Table 1 in the supplement), users moved images significantly more frequently and over a greater total distance in the high cognitive load condition (with 10 images) than the low load condition (with 5 images). Layout type did not affect the number of image moves. As shown in the top right panel, users moved images over a greater distance numerically in cylindrical layouts than planar layouts, but this difference was not statistically significant ($p = .06$). The effect of layout did not depend on cognitive load, for neither the number of image moves nor the total distance moved; that interaction was not significant in either model. Trial order was not a significant predictor of behavior either.

With respect to looking durations, we examined the amount of time that users looked at images vs. looked at other locations (not images). Looking location (image vs. non-image) and its interactions with cognitive load and layout type were entered as fixed effects in the model. Looking location, cognitive load, and their interaction were significant predictors of gaze durations, as shown in the bottom left panel Figure 6 (see also Table 2 in the supplement). Not surprisingly, users spent more time looking at images than non-images. Moreover, looking durations were longer in the high cognitive load condition than the low cognitive load condition. The effect of cognitive load depended on looking location: for low load trials looking durations at images and non-image areas were more comparable, but for high load trials users spent more than 1.5 times longer gazing at images than non-images. There was also a significant interaction between looking location and layout type: users spent a comparable amount of time looking non-images in the two layout types, but they spent more time looking at images in planar layouts compared to cylindrical layouts.

Finally, we considered trial completion times. As illustrated in Figure 6 (bottom right panel) completion times showed a significant effect of cognitive load (see Table 1 in the supplement). High load trials took about 1.5 times longer to complete than low load trials ($M = 187.76$ sec vs. $M = 125.11$ sec). Layout type did not influence completion times and neither did its interaction with cognitive load. Users tended to get faster at completing trials over the course of the study, but the effect of trial order was not significant ($p = .05$).

5.3 Changes in Interactions Over the Course of the Trial

Next, we examined how the users' interactions changed over the course of the trial. This was an exploratory analysis. We divided each trial in three equal segments and included trial segment as a fixed effect in the linear mixed effects models, along with its interactions with cognitive load and layout type. As shown in Figure 7, different user interactions exhibited distinct patterns over time.

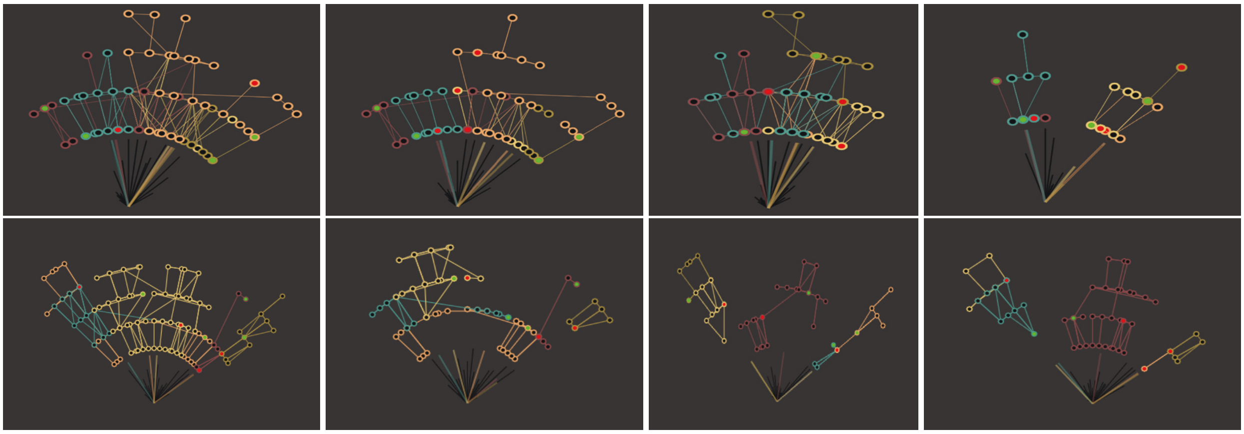
First, with respect to moving images (number of moves and distance images were moved), there was high activity for the beginning and middle segments of the trial and lower activity at the end of the trial (top two panels of Figure 7). Indeed, the difference between the first and second segments was not statistically significant for the number of moves and distance moved, but the difference between the second and third segments was (see Table 3 in supplement).

Second, looking durations (bottom left panel) exhibited an increase in the middle segment relative to the first and final segments. The difference between the first and second segments was statistically significant.

Finally, switches in looking orientation were most frequent in the first segment and dropped drastically after that (bottom right panel). This is captured by the fact that the difference between the first and second segments was statistically significant in the model, but that between the second and third segments was not.

5.4 Task Performance in Relation to User Interactions

Finally, we assessed the users' task performance as the trial unfolded relative to different types of interactions. We had hypothesized that increased user interactions would be associated with better performance, especially on high cognitive load trials. We considered 5 sequential ranges of the graded performance value for the task (0-.20, .20-.40, .40-.60, .60-.80, and .80-1), which reflected task accuracy. We identified the number of image moves, the distance that images were moved, looking durations, and the number of switches in looking orientation in each range. As shown in Figure 8, a consistent pattern was observed across all of these types of interactions. In line with our predictions, users were more accurate (i.e., fell into a higher grade range) when they moved images more frequently and over greater distances, when they looked at the display longer, and when they switched their gaze location more often.



Entire Trial Start Middle End

Fig. 5. Different sensemaking process shown with our spatial visualization. Each row shows the interactions from two users across different time-course of the same trial. Based on their sequences of physical interactions, we can observe clear but distinct sensemaking strategies: the first user moved images around and gradually formed 2 clusters, while the second user settled down to 3 clusters quickly and used the image represented by dark red to confirm the results. Both results are valid (two clusters from the second user is combined by the first user).

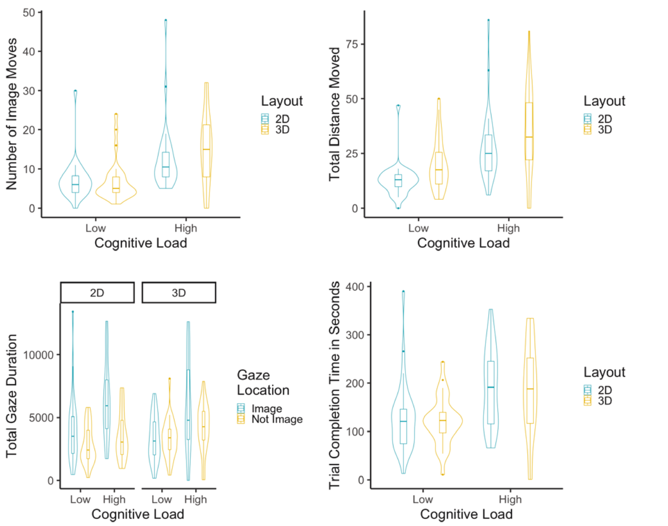


Fig. 6. Violin plots representing the distributions of the number of image moves (top left), total distance of all images moved (top right), gaze durations at images and non-images (bottom left), and trial completion times in a given trial (bottom right), according to the cognitive load (Low vs. High) and layout (planar/2D vs. cylindrical/3D) conditions. Boxplots represent the median and quartiles (Q1, Q3); dots indicate observations with values greater than Q3 plus 1.5 times the interquartile range. Cognitive load influenced these measures, but layout type did not.

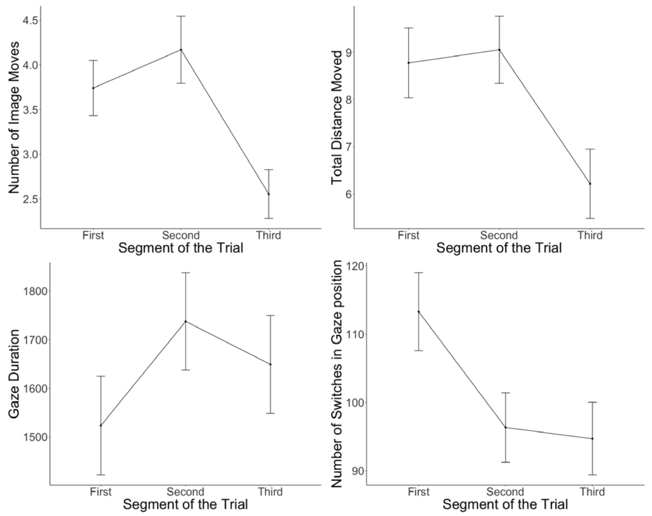


Fig. 7. Looking durations (bottom left) and gaze switches (bottom right) exhibit different patterns across the three segments of the trial than the number image moves (top left) and the total distance images were moved (top right). Error bars represent standard errors of the mean.

However, as indicated by the divergence between the blue and orange lines, this increase was modulated by cognitive load. For high cognitive load trials, the more users engaged in these interactions, the more accurate they were for the first 4 ranges of performance values (0-.80). Interestingly, users experienced an unexpected dip in these interactions for topmost range of performance (.80-1). In contrast, for low cognitive load trials, users engaged in these interactions relatively infrequently in the first four ranges of graded performance values (0-.80), but drastically more frequently at the topmost range of performance. The linear mixed effects models confirmed these observations.

As indicated by the levels of user interactions in Figure 8, some differences between high vs. low load trials at different performance ranges depended on cognitive load. The difference between moderately high (.60-.80) vs. low (0 - .20) performance differed significantly across cognitive load conditions as indicated by the significant interaction of this contrast for all user interactions (see Table 4 in supplement). The statistical interaction between cognitive load and performance range was also significant for the contrast between mid-level (.40-.60) vs. low (0 - .20) accuracy for image moves, looking durations, and gaze switches. Overall, the interaction of performance range and cognitive load was a significant predictor of these interactions, as it significantly improved model fit for image moves (χ^2 (11) = 34.68, $p < .001$), for distance moved (χ^2 (7) = 37.41, $p < .001$), looking durations (χ^2 (6) = 51.88, p < .001), and gaze switches (χ^2 (6) = 79.23, $p < .001$).

5.5 Relationship of Users' Behavioral and Neural Sequences

We used the visualization system to examine the relationship between the users' behaviors and fNIRS in different ways: first, by extracting and analyzing fNIRS from the system; second, through qualitative interactive analysis of spatial user interactions from the Hololens, and third through interactive and visual analysis of user variance from both aspects of fNIRS and spatial interactions.

Statistical Analyses of fNIRS. We examined linear mixed effects

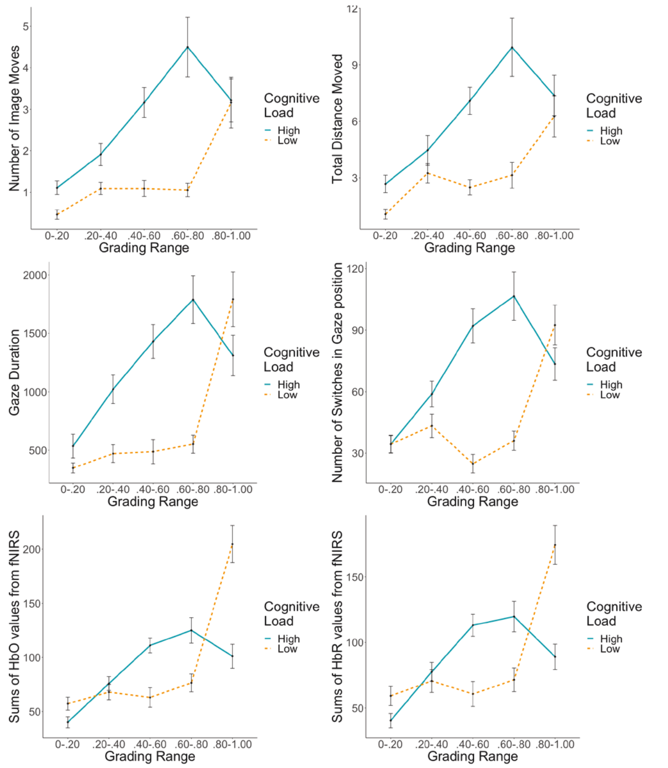


Fig. 8. Number of image moves (top left), total distance moved (top right), looking duration (middle left), number of switches in gaze position (middle right), sums of HbO values (bottom left), and HbR values (bottom right) from fNIRS across the 5 graded performance ranges and cognitive load conditions. Accuracy improves with increased user interactions and this effect is modulated by cognitive load. Error bars represent standard errors of the mean.

models of the same structure as in section 5.1, with the sums of oxy-hemoglobin (HbO) and deoxy-hemoglobin (HbR), normalized by user, as dependent measures. These cerebral hemodynamics measures exhibited the same patterns as the users' interactions at different ranges of task accuracy, as shown in Figure 8. For high load trials, HbO and HbR levels increased as accuracy increased, but dipped at the topmost level of performance. For low load trials, HbO and HbR levels rose suddenly at the topmost level of accuracy. That HbO and HbR levels exhibited the same pattern, when considering their aggregated sums in each trial, is consistent with evidence of their coupling [30]. We also examined whether HbO and HbR levels changes over the course of the trial; however, no differences were found across the three trial segments. (See Tables 5 and 6 in supplement for the full models.)

Interactive Analysis of Hololens Sequences. We used the visualization system to explore the relationship between user interactions and the other sources data (task performance and fNIRS). As a starting point, we used the spatial visualization from Hololens to identify users with distinct interaction styles for the same trial. Figure 5 illustrates a contrasting pair of users: the first user (top panels) started the trial by turning around to look at at all images, then moved images, and finished by focusing on two image clusters. The second user (bottom panels) spent most of the time comparing several images and ended with three image clusters. Although both users clustered the images with high accuracy in the end, one exhibited a divide-and-conquer strategy, whereby they actively moved items to constrain their search space whereas the other did not.

Exploration of Joint Data Using PCA. We also explored visualizations of the statistical distribution of fNIRS (as shown in Figure 9), since evaluating visually the raw data sequences on their own (as in Figure 4) can be difficult. Through dimensionality reduction, the PCA panel of the system presents an overview of the changes of one or multiple user and allows the evaluation of the distribution of nodes. Figure 9 displays all users, with each node representing one user during a time duration. Darker nodes represent earlier durations of the same user so that changes over time can be examined. If we treat the center region with dense nodes as the "common" stages of cognition, the nodes on the outside are more likely to represent "unique" behaviors. In addition, we treat users whose majority of the nodes overlaps with one another as "similar" and users whose majority of the nodes are on the outside as "different" (e.g., see the purple colors on the top and the red colors on the left bottom of the space). This exploration of the data suggests that the cognitive states of the same user can vary through the trial without an obvious pattern, but the distribution of cognitive states can reveal systematic patterns among a group of users.

Additionally, we used the system to examine the relationship between neural activity and user interactions by exploring the interaction types associated with nodes at different parts of the PCA space. Although the correlations were not always clear, we did find some consistent distributions: as shown in Figure 9, the nodes on the top of the PCA space are often linked with simpler interactions including looking at a small number of images within a small field of view, while the nodes on the bottom correspond to a variety of relatively complex interactions. The left bottom space is associated with interactions of body movements represented by long black lines, and the right bottom is associated with interactions of direct image selection and movement.

5.6 Summary and Discussion of Results

This study explored the sensemaking process in mixed reality with a visual correlation task, which is a common immersive visualization problem that has been used to study a variety of data, such as network [14,29], geospatial [34,51] and 3D neuron [44] data.

Our key findings were that the cognitive load of the trials influenced the users' interactions with the system, these interactions changed over the course of the trial, and were associated with task accuracy and neural activity in systematic ways. Consistent with our predictions, we found that compared to low load trials, on high load trials users moved images more frequently, over a greater distance, and spent more time gazing at images than non-images.

Layout type had less of an impact on users' interactions. Users exhibited comparable repertoires of interactions in planar and cylindrical layouts, suggesting that the sensemaking process is similar across the two formats of immersive visualization. They moved images equally frequently for both layouts and gazed at the displays a comparable amount of time. Users spent more time looking at images in planar layouts than cylindrical layouts, but this could be because 2D layouts, due to their smaller size, afforded more opportunities to obtain samples of the users gazing at images. While previous results on the effects of layout type on task performance [16,29,34] are mixed, our results showed that these two layouts are comparably effective as information interfaces in immersive systems when user interactions are not restricted. The two types of layouts were selected based on the current design of immersive systems, both without 3D overlapping, extending the pervasive use of 2D over 3D in information visualization [47]. We expect that these two layouts will continue to be popular interface designs in mixed reality, while we are also intrigued to explore 3D immersive visualization in the future.

Our exploratory analysis of the time-course of the users' interactions revealed that they changed over time. Interactions with artifacts involved consistent patterns: at the beginning and middle of the trial, users moved more images and over longer distances, but less so at the final segment of the trial. This suggests that they engaged actively in forming image clusters during the first 2/3rds of the task. These patterns in image interactions can contextualize the users' gaze switches, which showed a reduction over time. It is possible that as users placed images in more stable clusters over time, they switched their gaze locations less frequently. In terms of looking durations, users looked at images longer in the middle of the trial, which could indicate more deliberation and image comparison at that stage. Brain hemodynamics measures from fNIRS (HbO and HbR), did not change significantly over time when summed over long temporal scales (1/3 of the trial), which is

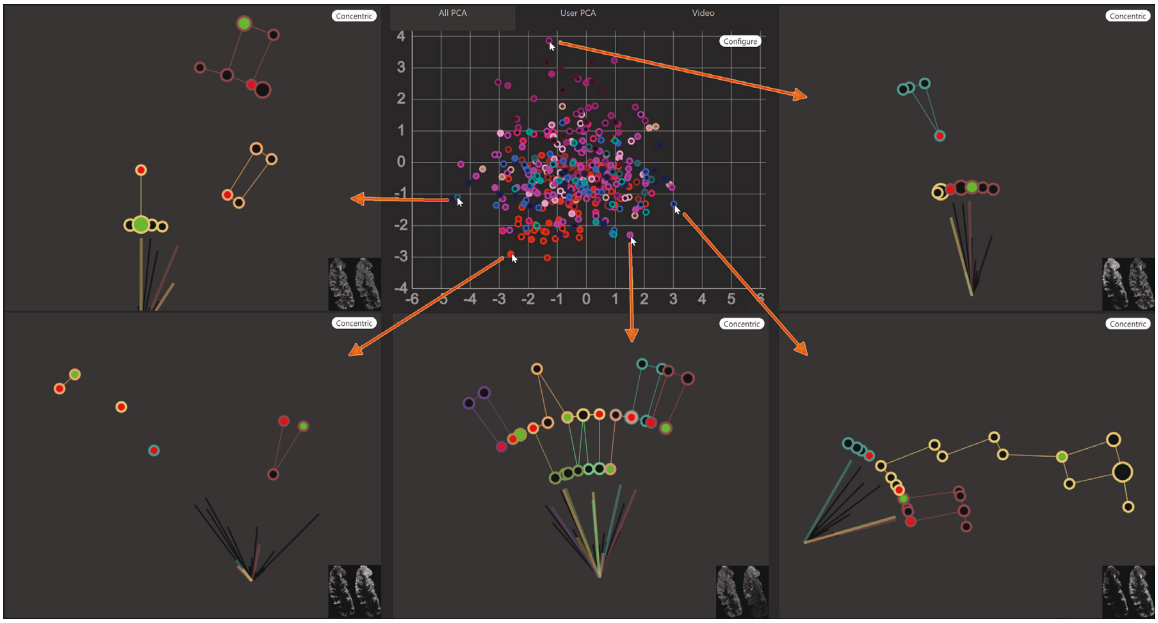


Fig. 9. The PCA panel assists us to explore different stages of cognition from the brain signals. While the PCA distribution is solely generated from fNIRS, this example shows that they correspond to the spatial interaction of users. The example nodes show that the top to bottom of the PCA space corresponds to low to high frequency of interactions, and the left to right corresponds to more body movements to more direct interactions on images.

not surprising. Altogether, these diverse patterns over time present a coherent picture about how users recruited different interactions to support their sensemaking in a visual correlation task.

With respect to task performance, consistent with our predictions, users' were more accurate when they engaged in more interactions, particularly for the high cognitive load trials. On those trials, both image moving behavior and looking behavior were associated with more accurate performance, except when the users were most accurate. For that most accurate performance, there was a dip in these interactions. In contrast, on low cognitive load trials, user interactions were relatively infrequent and not systematically associated with task accuracy, except for the top range of performance where they increased. Hemodynamic activity from fNIRS (HbO, HbR) showed a similar pattern, suggesting that the users' neural activity was coupled with their interactions.

The dip in user interactions and brain activity at the top range of performance on high load trials could be because users may have started moving the images to form clusters from the get-go, such that by then, during moments of high accuracy, they had settled on the position of the images: they did not need to move or look at these images more in those moments. Similarly the spike in activity at the top range of performance in low load trials could be because, initially, this smaller set of images was apprehended with less image movement and less looking around. As users finalized their decision about how form clusters, they moved images and looked around more, also exhibiting higher levels of brain activity and high accuracy. These findings have implications for theory because they suggest, under different task constraints, that the coupling between body movement, cognitive activity, and task state may change or may involve different time lags.

Additionally, we gained further insight into the relationship between patterns of neural activity and user interactions by using the visualization system to compare a PCA space analysis of fNIRS with the accompanying Hololens interactions. This permitted us to detect sensemaking strategies that are "common" vs. "unique" across users.

Consistent with theories of embodied and distributed cognition [12, 23, 27], we find that people perform actions that alter the world to offload their cognitive demands and support sensemaking. In immersive visualization, when cognitive demands are high, users increase the movement of images and of their own body. These interactions exhibit consistent patterns with the users' neural activity and task performance. Our findings show great potential for creating personal models of interaction patterns to predict user behaviors, performance, and sensemaking process during tasks. Our approach can inform the design of immersive interfaces by identifying the layouts, task constraints, and temporal phases of a sensemaking task that are more likely to involve complex interactions or unique behaviors.

6 Conclusion and Future Work

This work has explored the sensemaking process in immersive visualization. Our visualization system illuminated the users' strategies during sensemaking by permitting us to explore different user interactions and concurrent neural activity, both through quantitative analysis and interactive visual analysis. Specifically, using our system, we were able to extract and analyze data about the incidence of user interactions and levels of cerebral hemodynamic measures in each trial and to evaluate statistically the effects of layout, cognitive load, task performance, and task epoch. Our findings that user interactions increase on demanding trials and are associated with better performance, regardless of layout type, have implications for other visualization tasks that share common cognitive underpinnings, such as evaluating data similarity. By visually exploring, in the system, the statistical distribution of fNIRS subjected to dimensionality reduction (PCA), we were able identify common and unique patterns of interactions associated with different states of cognitive load. Our system–both the user's interface and the researcher's visualization interface–can be adapted to support and investigate a variety of complex decision-making and problem-solving tasks relevant to many real-world uses, including applications in teaching and learning. Beyond informing the design and evaluation of immersive systems, our system can be used to systematically examine sensemaking under different layouts, task constraints, task phases, user groups, and thus contribute to theoretical advances in embodied and distributed cognition.

In the future, we plan to extend the system to examine other sensemaking tasks and other complex cognitive processes. We are also interested developing machine learning models of user interaction to support personalized interfaces.

7 Acknowledgement

We are thankful for the reviewers' suggestions and early discussions with Tahir Mahmood and Dr. Nicholas Davis. This work was supported by NSF Awards 1564039, 1629913, 1661280, 1840080 and 1937010.

REFERENCES

[1] C. Andrews, A. Endert, and C. North. Space to think: Large high-resolution displays for sensemaking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pp. 55–64, 2010. doi: 10.1145/1753326.1753336

[2] S. Aydöre, M. K. Mihçak, K. Çiftçi, and A. Akin. On temporal connectivity of pfc via gauss - markov modeling of fnirs signals. *IEEE Transactions on Biomedical Engineering*, 57(3):761–768, 2010. doi: 10.1109/TBME.2009.2020792

[3] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3):255–278, 2013.

[4] A. Batch, A. Cunningham, M. Cordeil, N. Elmqvist, T. Dwyer, B. H. Thomas, and K. Marriott. There is no spoon: Evaluating performance, space use, and presence with expert domain users in immersive analytics. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):536–546, 2020. doi: 10.1109/TVCG.2019.2934803

[5] D. Bates, M. Maechler, B. Bolker, S. Walker, R. Haubo Bojesen Christensen, et al. lme4: Linear mixed-effects models using eigen and s4. r package version 1.1–7. 2014, 2015.

[6] J. Baumeister, S. Y. Ssin, N. A. M. ElSayed, J. Dorrian, D. P. Webb, J. A. Walsh, T. M. Simon, A. Irlitti, R. T. Smith, M. Kohler, and B. H. Thomas. Cognitive cost of using augmented reality displays. *IEEE Transactions on Visualization and Computer Graphics*, 23(11):2378–2388, 2017. doi: 10.1109/TVCG.2017.2735098

[7] X. Bi and R. Balakrishnan. Comparing usage of a large high-resolution display to single or dual desktop displays for daily work. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pp. 1005–1014, 2009. doi: 10.1145/1518701.1518855

[8] R. Brunken, J. L. Plass, and D. Leutner. Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, 38(1):53–61, 2003. doi: 10.1207/S15326985EP3801_7

[9] S. Butscher, S. Hubenschmid, J. Müller, J. Fuchs, and H. Reiterer. Clusters, trends, and outliers: How immersive technologies can facilitate the collaborative analysis of multidimensional data. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173664

[10] M. P. Çakir, M. Vural, S. O. Koç, and A. Toktaş. Real-time monitoring of cognitive workload of airline pilots in a flight simulator with fnir optical brain imaging technology. In *Proceedings, Part I, 10th International Conference on Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience - Volume 9743*, pp. 147–158, 2016. doi: 10.1007/978-3-319-39955-3_14

[11] T. Chandler, M. Cordeil, T. Czauderna, T. Dwyer, J. Glowacki, C. Goncu, M. Klapperstueck, K. Klein, K. Marriott, F. Schreiber, and E. Wilson. Immersive analytics. In *Big Data Visual Analytics (BDVA), 2015*, pp. 1–8, 2015.

[12] A. Clark and D. Chalmers. The extended mind. *Analysis*, 58(1):7–19, 1998.

[13] M. Cordeil, A. Cunningham, T. Dwyer, B. H. Thomas, and K. Marriott. Imaxes: Immersive axes as embodied affordances for interactive multivariate data visualisation. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, pp. 71–83. ACM, New York, NY, USA, 2017. doi: 10.1145/3126594.3126613

[14] M. Cordeil, T. Dwyer, K. Klein, B. Laha, K. Marriot, and B. H. Thomas. Immersive collaborative analysis of network connectivity: Cave-style or head-mounted display? *IEEE Transactions on Visualization and Computer Graphics*, PP(99):1–1, 2016.

[15] T. J. Crooks. *Assessing student performance*. Kensington, N.S.W., 1988.

[16] E. D. Ragan, A. Endert, D. Bowman, and F. Quek. The effects of spatial layout and view control on cognitive processing. In *Conference on Human Factors in Computing Systems*, pp. 2005–2010, 01 2011. doi: 10.1145/1979742.1979921

[17] A. De, A. Konar, A. Samanta, S. Biswas, A. L. Ralescu, and A. K. Nagar. Cognitive load classification in learning tasks from hemodynamic responses using type-2 fuzzy sets. In *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1–6, 2017. doi: 10.1109/FUZZ-IEEE.2017.8015659

[18] T. Geymayer, M. Waldner, A. Lex, and D. Schmalstieg. How sensemaking tools influence display space usage. In *EuroVis Workshop on Visual Analytics*, 2017.

[19] J. Grudin. Partitioning digital worlds: Focal and peripheral awareness in multiple monitor use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 458–465, 2001. doi: 10.1145/365024.365312

[20] H. Guo, S. R. Gomez, C. Ziemkiewicz, and D. H. Laidlaw. A case study using visualization interaction logs and insight metrics to understand how analysts arrive at insights. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):51–60, 2016. doi: 10.1109/TVCG.2015.2467613

[21] C. Herff, O. Fortmann, C. Tse, X. Cheng, F. Putze, D. Heger, and T. Schultz. Hybrid fnirs-eeg based discrimination of 5 levels of memory load. In *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 5–8, 2015. doi: 10.1109/NER.2015.7146546

[22] D. R. Hutchings and J. Stasko. Revisiting display space management: Understanding current practice to inform next-generation design. In *Proceedings of Graphics Interface 2004*, GI '04, pp. 127–134, 2004.

[23] E. Hutchins. How a cockpit remembers its speeds. *Cognitive science*, 19(3):265–288, 1995.

[24] K. Izzetoglu, S. Bunce, M. Izzetoglu, B. Onaral, and K. Pourrezaei. fnir spectroscopy as a measure of cognitive task load. In *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)*, vol. 4, pp. 3431–3434 Vol.4, 2003. doi: 10.1109/IEMBS.2003.1280883

[25] P. C. Johnson. Extension of nakagawa & schielzeth's r2glmm to random slopes models. *Methods in Ecology and Evolution*, 5(9):944–946, 2014.

[26] N. Karamzadeh, F. Amyot, K. Kenney, F. Chowdhry, A. Anderson, V. Chernomordik, E. Wegman, E. Wassermann, R. Diaz-Arrastia, and A. Gandjbakhche. A machine learning approach to identify functional biomarkers for traumatic brain injury (tbi) using functional near-infrared spectroscopy (fnirs). p. BTh4D.2, 01 2016. doi: 10.1364/BRAIN.2016.BTh4D.2

[27] D. Kirsh. The intelligent use of space. *Artif. Intell.*, 73(1-2):31–68, Feb. 1995. doi: 10.1016/0004-3702(94)00017-U

[28] A. Kuznetsova, P. B. Brockhoff, R. H. B. Christensen, et al. Tests in linear mixed effects models. *R package version*, 2:33, 2016.

[29] O. H. Kwon, C. Muelder, K. Lee, and K. L. Ma. A study of layout, rendering, and interaction methods for immersive graph visualization. *IEEE Transactions on Visualization and Computer Graphics*, 22(7):1802–1815, 2016.

[30] P. Lachert, D. Janusek, P. Pulawski, A. Liebert, D. Milej, and K. J. Blinowska. Coupling of oxy-and deoxyhemoglobin concentrations with eeg rhythms during motor task. *Scientific Reports*, 7(1):1–9, 2017.

[31] W. S. Lages and D. A. Bowman. Walking with adaptive augmented reality workspaces: Design and usage patterns. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, IUI '19, p. 356–366. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3301275.3302278

[32] B. Lee, X. Hu, M. Cordeil, A. Prouzeau, B. Jenny, and T. Dwyer. Shared surfaces and spaces: Collaborative data visualisation in a co-located immersive environment. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2020. doi: 10.1109/TVCG.2020.3030450

[33] B. Lee, P. Isenberg, N. H. Riche, and S. Carpendale. Beyond mouse and keyboard: Expanding design considerations for information visualization interactions. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2689–2698, 2012. doi: 10.1109/TVCG.2012.204

[34] J. Liu, A. Prouzeau, B. Ens, and T. Dwyer. Design and evaluation of interactive small multiples data visualisation in immersive spaces. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 588–597, 2020.

[35] T. Mahmood, E. Butler, N. Davis, J. Huang, and A. Lu. Building multiple coordinated spaces for effective immersive analytics through distributed cognition. In *2018 International Symposium on Big Data Visual and Immersive Analytics (BDVA)*, pp. 1–11, 2018. doi: 10.1109/BDVA.2018.8533893

[36] N. Naseer and K. shik Hong. fnirs-based brain-computer interfaces: a review. In *Front. Hum. Neurosci.*, 2015.

[37] F. Putze, S. Hesslinger, C.-Y. Tse, Y. Huang, C. Herff, C. Guan, and T. Schultz. Hybrid fnirs-eeg based classification of auditory and visual perception processes. *Frontiers in neuroscience*, 8:373, 11 2014. doi: 10.3389/fnins.2014.00373

[38] E. D. Ragan, A. Endert, D. A. Bowman, and F. Quek. How spatial layout, interactivity, and persistent visibility affect learning with large displays. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, pp. 91–98, 2012. doi: 10.1145/2254556.2254576

[39] Y. Rogers and J. Ellis. Distributed cognition: an alternative framework for

analysing and explaining collaborative working. *Journal of Information Technology*, 9(2):119–128, Jun 1994. doi: 10.1057/jit.1994.12

[40] M. Rønne Jakobsen and K. Hornbæk. Sizing up visualizations: Effects of display size in focus+context, overview+detail, and zooming interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 1451–1460, 2011. doi: 10.1145/1978942.1979156

[41] E. Suma, S. Finkelstein, M. Reid, S. Babu, A. Ulinski, and L. F. Hodges. Evaluation of the cognitive effects of travel technique in complex real and virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 16(4):690–702, 2010. doi: 10.1109/TVCG.2009.93

[42] S. Tak and J. C. Ye. Statistical analysis of fnirs data: A comprehensive review. *NeuroImage*, 85:72–91, 2014.

[43] R. C. Team. R: a language and environment for statistical computing. version 3.4. 1. vienna: R foundation for statistical computing, 2016.

[44] W. Usher, P. Klacansky, F. Federer, P. Bremer, A. Knoll, J. Yarch, A. Angelucci, and V. Pascucci. A virtual reality visualization tool for neuron tracing. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):994–1003, Jan 2018. doi: 10.1109/TVCG.2017.2744079

[45] R. Vuillemot, J. Boy, A. Tabard, C. Perin, and J.-D. Fekete. *LIVVIL: Logging Interactive Visualizations and Visualizing Interaction Logs*. 10 2016.

[46] M. Waldner, R. Grasset, M. Steinberger, and D. Schmalstieg. Display-adaptive window management for irregular surfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '11, pp. 222–231, 2011. doi: 10.1145/2076354.2076394

[47] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann, 2012.

[48] C. Ware and P. Mitchell. Visualizing graphs in three dimensions. *ACM Trans. Appl. Percept.*, 5(1):2:1–2:15, Jan. 2008. doi: 10.1145/1279640.1279642

[49] M. Whitlock, K. Wu, and D. A. Szafir. Designing for mobile and immersive visual analytics in the field. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):503–513, 2020.

[50] M. Wu and V. Popescu. Efficient vr and ar navigation through multiperspective occlusion management. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2018. doi: 10.1109/TVCG.2017.2778249

[51] Y. Yang, T. Dwyer, B. Jenny, K. Marriott, M. Cordeil, and H. Chen. Origin-destination flow maps in immersive environments. *IEEE Transactions on Visualization and Computer Graphics*, 2018. doi: 10.1109/TVCG.2018.2865192

[52] Y. Yang, B. Jenny, T. Dwyer, K. Marriott, H. Chen, and M. Cordeil. Maps and Globes in Virtual Reality. *Computer Graphics Forum*, 2018. doi: 10.1111/cgf.13431