Tensions between Preference and Performance: Designing for Visual Exploration of Multi-frequency Medical Network Data

Christian Knoll, Laura Koesten, Isotta Rigoni, Serge Vulliémoz, and Torsten Möller, *Senior Member, IEEE*

Abstract—The analysis of complex high-dimensional data is a common task in many domains, resulting in bespoke visual exploration tools. Expectations and practices of domain experts as users do not always align with visualization theory. In this paper, we report on a design study in the medical domain where we developed two high-fidelity prototypes encoding EEG-derived brain network data with different types of visualizations. We evaluate these prototypes regarding effectiveness, efficiency, and preference with two groups: participants with domain knowledge (domain experts in medical research) and those without domain knowledge, both groups having little or no visualization experience. A requirement analysis and study of low-fidelity prototypes revealed a strong preference for a novel and aesthetically pleasing visualization design, as opposed to a design that is considered more optimal based on visualization theory. Our study highlights the pros and cons of both approaches, discussing trade-offs between task-specific measurements and subjective preference. While the aesthetically pleasing and novel low-fidelity prototype was favored, the results of our evaluation show that, in most cases, this was not reflected in participants' performance or subjective preference for the high-fidelity prototypes.

Index Terms—Design study, EEG data, network analysis

I. MOTIVATION

THE visualization community developed many helpful design guidelines to greatly narrow the possible approaches for developing visual analysis interfaces, from facetting (and the general approach of interactive multi-view dashboards) to Shneiderman's Mantra [1]. However, all of these guidelines optimize the effectiveness of such visual analysis systems. When a new goal, such as aesthetics and engagement, is being formulated, it becomes again a challenge to narrow the design space properly. This paper reports on a design study to help clinical researchers explore and understand complex high-dimensional data from electroencephalography (EEG) and epilepsy patients. Specifically, they compute and analyze the specific network of connections between brain regions across different EEG frequencies (frequency bands) where simple visualization approaches have not helped the analysis. However, our domain experts very much favor novel and engaging interfaces.

One of the non-trivial challenges of this process is to develop tools that (a) embed properly in an existing workflow and (b) respect certain expectations of the domain experts. From the beginning, our domain experts communicated to us their expectations of engaging visual encodings that would be novel within their particular domain. Due to their aesthetics, they favored particular visual encodings (circular design) early

on in the prototyping process. Thus, we were interested in the challenge of developing the best prototype under the constraint of an engaging circular layout and comparing it to the best prototype based on alternative designs, following a traditional functional design approach. We are unaware of other design studies that have dealt explicitly with these competing design goals.

Because of our domain experts' bias toward novelty and aesthetics, it was important for us to test and collect feedback on our prototypes from people outside of the specific domain as well. While there are many definitions for domain novices [2], we refer to them as lay people without domain expertise having little or no visualization experience. The drawback was to find a way to give them tasks to be done with our prototypes that they could understand without domain knowledge. The advantage would be that they have likely not dealt with similar visualization problems before and might be less biased in either direction (aesthetics vs. functionality). At the same time, however, it was essential for our evaluation to not just measure overall usability (and functionality) as in a typical design study but to measure perceived aesthetics and engagement specifically.

We developed two high-fidelity prototypes encoding multi-frequency EEG data. We evaluate these prototypes regarding effectiveness, efficiency, and subjective preference with participants with domain knowledge (domain experts) and without domain knowledge (lay people), both groups having little or no visualization experience. The primary contributions of this paper include:

- a characterization of the domain problem and tasks through an in-depth requirement analysis with diverse medical researchers working on multi-frequency network data (section IV).
- two high-fidelity prototypes with different visual encodings both optimized for functionality. However, one was additionally constrained by an aesthetically pleasing visual encoding (section V).
- a user study comparing four domain experts' and twelve lay peoples' subjective preferences and objective task performance (section VI).

II. DOMAIN BACKGROUND

Recently, EEG-based network analyses have been increasingly applied to investigate brain networks in physiological conditions or brain disorders such as epilepsy [3]–[5].

Promising results have been found for both the diagnosis and prognosis of epilepsy patients. At the core of the analysis are novel approaches for estimating and characterizing the brain network of a particular patient. These networks give insight into the brain's function (or dysfunction). However, insufficient validation of these brain networks limits the applicability of analysis techniques for clinical practice [6].

Among issues linked to the multiple ways of estimating connectivity and characterizing functional graphs, the visualization of graphs and measures within them is complex. Simplified visualizations, as used commonly in the domain, further hinder the interpretation of these networks. Typically, for brain networks derived from high-density EEG (HD-EEG), the number of (cortical) regions investigated ranges from 64 [7] to 72 to even more regions, depending on specific research needs [8], [9]. Furthermore, EEG data is usually analyzed in several frequency bands to reflect specific activity patterns related to cognition and pathological alterations [10]. Graph measures can be computed at the global level (e.g., to describe the global integration or segregation of the graph), at the hemispheric level, or even at the nodal one [11]. Significant differences between nodal measures (e.g., across patients, vigilance state, before and after medication, etc.) are much harder to visualize due to the high number of nodes and frequency bands in the brain network, not to mention the temporal dimension in some studies.

Therefore, we developed a tool that provides a descriptive visual view of nodal graph measures (e.g., the clustering coefficient) and visualizes them across frequency bands or as part of different subnetworks. The goal of this tool is, on the one hand, to enable researchers to gain a descriptive view of the network before statistical testing. It should allow them to formulate hypotheses based, for example, on perceptual similarities across frequency bands for a specific cluster of nodal regions of interest (ROI). This would enable exploratory visual analyses of the data without investigating each brain region separately, leading to more powerful targeted analyses. It eliminates the need to adjust significance levels for multiple testing [12]. On the other hand, it could, for example, enable clinicians to explore different network patterns arising from removing some nodes to simulate a lesion or a surgical procedure, facilitating the incorporation of research data into clinical discussions and promoting interactive multidisciplinary meetings.

III. RELATED WORK

In this section, we review existing literature on the effectiveness of visualization encodings, different factors influencing a visualization design, and visual analysis tools for EEG data.

A. Visualization Marks and Channels

To create effective visualizations, designers must choose the optimal encoding for their data. Visual encodings consist of marks and channels. Marks are graphical elements in visual encodings, and their appearance is controlled through channels [13]. The effectiveness of a channel depends on the data type. We primarily use the length of a bar and color saturation as channels in our prototypes—the former being the more effective encoding for our data based on visualization literature [13]–[16]. These studies show that spatial channels most effectively encode ordered data. A recent survey by Quadri and Rosen [17] provides an overview of existing perception-focused visualization studies, categorizing them by the task taxonomy of Amar et al. [18].

Bar length as a channel. There are several studies investigating the perception of bar charts. The visualization tasks discussed in our work commonly relate to visual comparison, e.g., of bars across five frequency bands and up to 72 ROIs, as detailed in subsection IV-B. For instance, the visual comparison of bars has been researched by Talbot et al. [19], who found that comparing non-adjacent bars is difficult, especially for short bars. A study by Nothelfer and Franconeri [20] looked at how relations between bars are perceived and found that data deltas were better processed visually when explicitly coded. This is also suggested by Srinivasan et al., who evaluated different bar chart variants for visual comparison and found that adding "difference overlays facilitate a wider range of comparison tasks" [21, p. 1]. While most of these studies focus on comparing a pair or a small number of bars, one of our prototypes displays up to 72 bars in each of the five bar charts. Gramazio et al. [22] showed that size, quantity, and grouping of marks influence user performance and found that search tasks can be solved faster when marks are spatially grouped (such as in bar charts) rather than randomly arranged. Additionally, linear arrangements of marks are usually more effective [23] and efficient [24] than radial arrangements – a comparison we also touch upon in our prototypes.

Color as a channel. Using color in visualizations has also been broadly researched. Szafir [25] conducted experiments to measure the perception of color difference in visualizations. She found that "elongated marks provide significantly greater discriminability for encoding designers" [25, p. 392]. Color often encodes continuous quantitative data (e.g., on maps), such as the EEG data we used in our design study. While rainbow color maps are usually considered a poor choice [26]–[28], Khairi et al. [29] observed that rainbow schemes can be effective for quantity estimation, while divergent colormaps facilitate the perception of high-frequency patterns. We used single-hue sequential colormaps in our prototypes informed by the findings of Karim et al. [30] where they compared different single- and multi-hue colormaps in network visualization. They found that participants in their study completed tasks significantly faster with a blue single-hue sequential colormap. There are also tools [31], [32] that enable finding effective color schemes for a visualization, which we used to find effective colormaps.

B. Influencing Factors for Visualization Design

Visualization design decisions are influenced by the appropriate visual encoding and several other factors (e.g., aesthetics, familiarity) that need to be considered when creating effective visualizations. Tory and Möller [33] argue that human factors (e.g., visual acuity, culture, previous experience) must be considered in designing and evaluating visual analysis tools

to add to the systems' usefulness for future users. While human-centered design approaches for visualization systems usually involve a requirement analysis to characterize the domain problem and prospective users, visualization designers often focus primarily on task-based objectives (e.g., conveying facts or insights), and only recently, there has also been a shift towards affective objectives by the academic visualization community. Lee-Robbins and Adar also emphasize paying attention to affective intents "that seek to influence or leverage the audience's opinions, attitudes, or values" [34, p. 1]. Pandey et al. [35], who examined persuasion of data visualizations, also reflected on affective intent. They found that people's attitudes about a topic influence the persuasive power of data visualization. In our study, we were especially interested in the influence of aesthetics of visualizations [36]–[38]. He et al. [39] introduced a novel scale called *BeauVis* for measuring and comparing the aesthetic pleasure of visual representations. Another important factor to consider is the familiarity of future tool users with the visual medium [40], [41].

C. Visualization Tools for EEG Data

Visualization is widely used for displaying and analyzing multi-channel EEG data. While all of the following papers use visualization to facilitate EEG analysis, to the best of our knowledge, there are no bespoke tools tailored to domain experts' needs for exploratory EEG data analysis. Slaybeck et al. [42] report on novel visualization methods for EEG signals using a virtual reality system and a physical head model. A parallel coordinate method is proposed and evaluated by ten Caat et al. [43]. Since visualizing connectivity in EEG data as a graph layout can result in visual clutter, ten Caat et al. [44] propose a graph layout based on functional units (data-driven region of interest). We alleviated the issue of visual connectivity clutter by allowing users to filter the connections. Wulandari et al. [45] developed a system to visualize the brain wave signals of epilepsy patients. Fang et al. [46] created a visualization dashboard for comparative analysis. There is also literature on using 3D representations for analyzing EEG data [47]–[49]. FCLAB [50] is a plugin-based environment that visualizes brain functional connectivity networks with local and global measures. While this system enables network analysis for each frequency band at the electrode level, we focus on source-reconstructed EEG at the ROI level. Additionally, our prototypes allow for comparing metrics across frequency bands and for user-defined subnetworks.

IV. DESIGN PROCESS

We realized a three-step design process that follows the design study methodology framework [51]. In the first phase, we interacted with medical domain experts from diverse fields (medical doctors, biomedical engineers, and computer scientists) to identify the challenges they face in their data analysis and to gather ideas about how data visualization can help facilitate this process. We conducted an in-depth requirement analysis with domain experts on-site in the second phase. This requirement analysis was the basis for three digital pen-and-paper prototypes. We used the collected knowledge and iterative feedback cycles in the third phase to implement two interactive high-fidelity prototypes.

A. Ideation

To generate and explore promising ideas that support the analysis of our collaborators through visualization, we conducted a creative visualization-opportunities (CVO) workshop [52] early on in the project. Nine medical domain experts (five male, four female), including clinicians and neuroscientists, participated in a full-day workshop to explore data visualization opportunities. The workshop was composed of four guided activities: After a short introduction of the topic and participants at the beginning of the workshop, the aim of *wishful thinking* was to generate a broad spectrum of ideas mainly focusing on the domain experts' information needs and the potentially related data visualization tasks. We gathered the answers on sticky notes and clustered them into common themes. The results showed the participants' interest in different aspects of EEG data, focusing on, among others, the diagnosis, prediction, and nature of epilepsy in patients.

In the second part of the CVO workshop, the participants were shown existing medical or related data visualizations to inspire their thinking and prepare them for the last activity. The workshop concluded with *video prototyping* [53], where participants were asked to develop a low-fidelity prototype in small groups for one of the themes that were found earlier. The CVO workshop helped characterize the domain experts' analysis challenges, gather ideas on how visualization could facilitate the analysis process, and produce the first prototypes that showcase how a visual analysis interface could look. We also identified collaborators who were excited to work with us on this project.

B. Requirement Analysis

In the second phase, we analyzed the artifacts produced in the initial workshop and used this knowledge to plan an in-depth requirement analysis with our collaborators. We spent one week working closely with the domain experts at their hospital to define the requirements for a visual analysis tool to facilitate their analysis of multi-frequency EEG data. Interviews with four domain experts (medical doctors, neuroscientists, and engineers) about their research, focusing specifically on the data they analyze and the context of data use, gave us an in-depth understanding of their analysis processes.

Data. The domain experts mainly operate in a clinical setting and look at multidimensional EEG data (see Figure 1) of epilepsy patients over time. EEG is a non-invasive technique used to measure the electrical activity in the brain using electrodes placed on the scalp. The data consists of a time series of electrical voltages reconstructed for different areas of the brain (regions of interest, ROI). By quantifying the level of similarity between time series belonging to different ROIs, the domain experts build a connectivity matrix that describes statistical dependencies between different brain regions. As different frequency components have cognitive and pathological correlates, domain experts analyze the brain network across frequency bands. In other words, the final
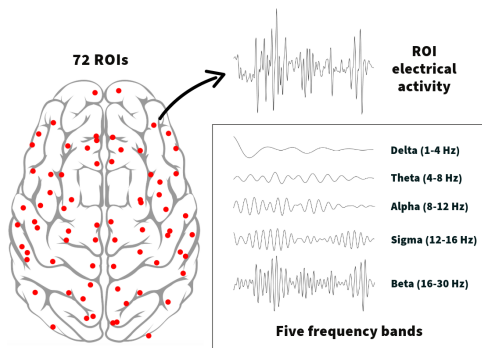
Fig. 1. Schematic representation of the EEG data. The grey matter (outer layer of the brain) was parceled into 72 ROIs (red dots), and each ROI's electrical activity was reconstructed from the EEG signal. The frequency content of the ROI signal is displayed for each frequency band in the box. Once the connectivity between each ROI is estimated, a brain network (ROI x ROI) is built for each frequency band. Then, nodal network metrics – such as the clustering coefficient – are computed for each ROI.

EEG connectivity data is a multidimensional matrix (ROI x ROI x frequency), where ROI's are the nodes and their relations are the network's edges. Finally, graph metrics are computed to characterize the brain graph at different frequencies, for example, by describing its segregation. More specifically, metrics such as the clustering coefficient (CC) are computed for each ROI and frequency band and describe the local level of clusterization between neighboring nodes, reflecting functional specialization. More details about EEG data collection, preprocessing, and connectivity analyses are provided as supplemental material. We used the knowledge gathered in the first part of the week to develop three low-fidelity prototypes in a digital pen-and-paper setting, which will be described in detail in subsection V-A.

Tasks. The domain experts focus their analyses on the whole brain network and the connections between specific ROI, and they extract global and nodal graph measures to describe these interactions. Whether this is done in broadband or for different frequency bands depends on the research focus. As mentioned, our data set consisted of five connectivity matrices, one for each frequency band (delta, theta, alpha, sigma, and beta), representing a typical data set for domain experts. The interview results show that the domain experts are especially interested in exploring nodal graph measures across frequency bands, as global measures are usually easier to display (typically in a boxplot). To summarize, the domain experts need a tool for the visual exploration of EEG data that facilitates the following tasks:

- *Overview:* getting an overview of ROIs and their connectivity for different frequency bands
- *Cluster:* grouping of ROIs in subnetworks
- *Locate:* locating ROIs and connections with high activity
- *Compare:* comparing ROIs and connections within and across frequency bands

C. Implementation

The last phase of our process focused on the prototype transition from low-fidelity to high-fidelity. Based on the results of the evaluation of the low-fidelity prototypes through interviews with the domain experts, we decided on two high-fidelity designs (see subsection V-B) for further implementation. Two interactive visualization dashboards were developed while conducting regular feedback cycles with the domain experts. Each dashboard consists of a JavaScript frontend that uses D3.js [54] for visualizing the data.

V. DESIGN DECISIONS

In this section, we will detail our design decisions during prototyping. First, we discuss our low-fidelity designs, where we created multiple digital pen-and-paper prototypes in parallel to explore different design options. Next, we elaborate on interviews that we conducted with our domain experts to assess these low-fidelity prototypes. We conclude this section by reporting on how we arrived at our high-fidelity prototypes.

A. Low-fidelity Designs

From the in-depth requirement analysis conducted on-site at our collaborators' hospital, we identified the researchers' need for a tool that facilitates the visual investigation of EEG-derived brain network data. While they primarily use mathematical analysis to explore the data, inspection of results through visualization occurs only as a final step of their analysis due to difficulties linked to visualizing complex multidimensional data. Finding an intuitive visual encoding of the data is necessary for several reasons, including but not limited to summarizing, validating, and communicating their findings. Based on this knowledge, we developed three low-fidelity prototypes (see Figure 2) that use different visual representations to facilitate analysis. Detailed information about the low-fidelity prototypes can be found in the supplemental material. Four domain experts were interviewed to evaluate the three low-fidelity prototypes. A ranking of the designs revealed a preference for low-fidelity prototype 3 (see Figure 2c) as it was ranked first in three of the four interviews. The participants found this design the most intriguing and innovative since they had never worked with this type of aggregated visualization before. They especially liked the idea of layering multiple frequency bands in segments on top of each other in a single circular layout, allowing for comparison across the frequencies. However, one participant found the design less intuitive than the other two, and the limited space for the connections inside the circle was pointed out as a weakness by another participant.

Low-fidelity prototype 1 (see Figure 2a) was ranked first by one participant and ranked in second place (out of three) by two other participants. Participants described the bar charts as intuitive, and one added that they allow for identifying trends. Grouping the bars by frequency band was identified as helpful by one of the participants since they can compare values within a frequency and between frequencies. The primary issue that three of the four participants identified was related to the tool's scalability. The number of ROIs is usually relatively high (from 72 ROIs and above), so showing all the regions in the bar chart might not effectively scale. As a follow-up, we showed them an equivalent stacked bar chart (the five
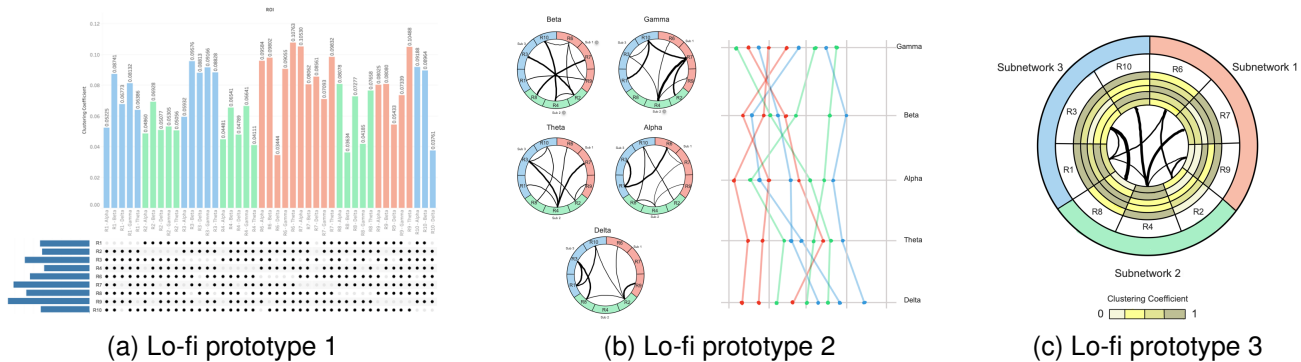
Fig. 2. The three low-fidelity prototypes: (a) using bar charts (metric) and a dot plot (connectivity), (b) using parallel coordinates (metric) and doughnut charts (connectivity), and (c) using a layered doughnut chart (metric and connectivity). *Remark: user interface elements (e.g., menu for selecting ROIs) have been excluded to focus on the visual encoding of the data.*

frequencies stacked for each region) to understand whether this would help the scalability of the visualization. The feedback suggested that this might be helpful for comparisons between the frequency bands. Still, three of the four participants did not find the stacked bar chart helpful since comparing values for specific frequencies across different regions was harder. All participants liked the dot plot encoding the connections between the regions. However, most reported that they would expect a heatmap to be more informative and intuitive since they are used to working with heatmaps. The horizontal bar chart left to the dot plot was introduced as an optional feature for users to display data attributes of their choice. The participants found this option intriguing but could not imagine a usage scenario.

The least preferred low-fidelity prototype was design 2 (see Figure 2b). Two participants pointed out that circular plots are commonly used for plotting connections between entities in the medical research literature, giving them a sense of familiarity. Most participants also liked the idea of having individual circles for each frequency. However, one participant raised concerns about edge clutter with increasing connections and suggested letting users switch between a node-link and an adjacency matrix representation. The parallel coordinates plot did not add any value for three of the four participants since they reported it to be less intuitive than the bar charts. The other participant, however, found it helpful to follow the value changes across frequencies.

B. High-fidelity Designs

The interview study showed that low-fidelity prototype 3 was the most favored of our three initial designs. As pointed out by several participants, combining different data features into a single chart was unconventional and new to them. Even though prototype 1 received good feedback, participants seemed to prefer the originality of prototype 3 over conventional chart types. Since their feedback was based on looking at the digital low-fidelity prototypes without any interactivity, we implemented prototype 1 and prototype 3 as testbeds for further evaluation. Prototype 2 was disregarded from further implementation as the interviews showed that the parallel coordinates plot did not add much value to the

analysis of our domain experts. We iteratively improved the implemented prototypes with feedback from the collaborating domain experts throughout the implementation phase.

Prototype A *(layered doughnut chart)* (see Figure 3a) was implemented based on low-fidelity prototype 3. The final testbed consists of four main views. The **brain view (1)** *(Overview, Locate)* shows a schematic representation of a patient's brain from three perspectives (from top to bottom): superior view, lateral view, and posterior view. The red dots represent the 72 nodal regions. The **ring view (2)** *(Overview, Locate, Compare)* contains two parts. The five rings show the values of a metric (in this case, the clustering coefficient) encoded with color. Each ring segment encodes the value for an ROI in different frequencies, and each of the five frequencies is depicted through a color (red, green, blue, orange, and purple). The color saturation of each segment encodes the metric value: brighter colors indicate lower values than darker shades. This allows users to get a first overview of the value distribution. Hovering over a segment, a tooltip details the exact value.

The area within the five rings encodes the connectivity between the different regions. Since the connectivity changes for every frequency, the user can select the frequency by clicking anywhere in the corresponding frequency ring. We used the color saturation of the lines to encode the connection strength to alleviate the problem that there might be up to 72x72 connections where it would be impossible to see individual lines. This should also help users visually identify potential connectivity patterns more easily.

The button in the upper left corner opens a popup window that lets the user select subnetworks *(Cluster)*. A user might want to focus only on specific groups of ROIs (here called subnetworks) rather than looking at the global network. Figure 4 shows three subnetworks containing five, three, and five ROIs, respectively. The lines outside the circle visually indicate the subnetworks.

The **metric histogram view (3)** *(Overview, Locate, Compare)* on the right shows the distribution of the metric values in each frequency (ring segments). The histograms have the same scale on the y-axis, which enables visual comparison
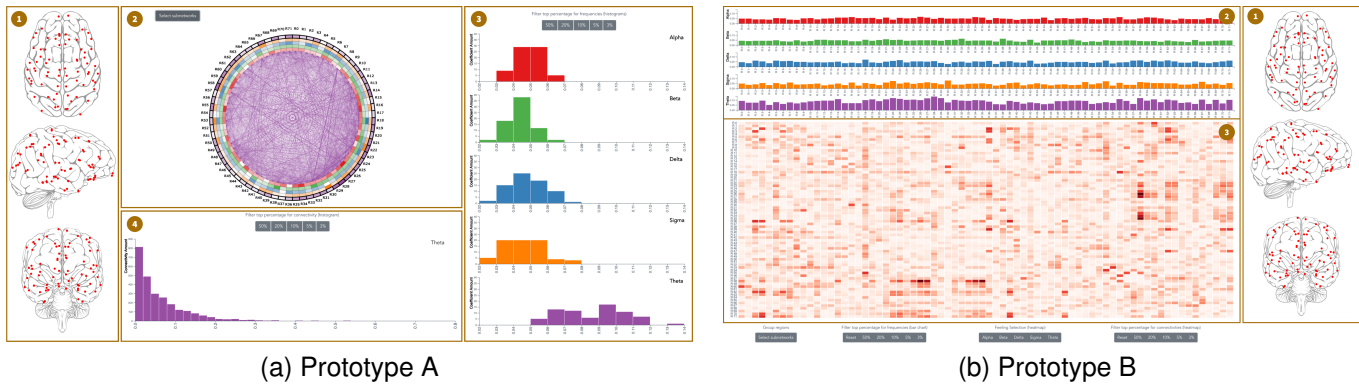
(a) Prototype A

(b) Prototype B

Fig. 3. High-fidelity prototypes with all 72 ROIs. Prototype A (based on low-fidelity prototype 3) shows a layered doughnut chart (metric and connectivity), histograms (metric and connectivity), and a brain plot with 72 ROIs. Prototype B (based on low-fidelity prototype 1) shows bar charts (metric), a heatmap (connectivity), and a brain plot with 72 ROIs.
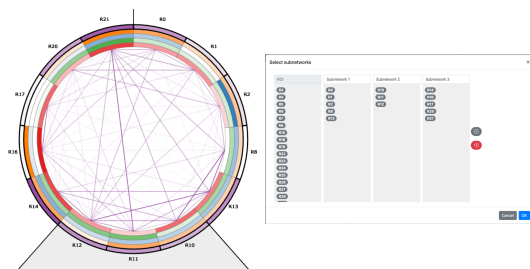

Fig. 4. High-fidelity prototype A with three subnetworks containing five, three, and five ROIs, respectively.

across the different frequencies. Users can select ranges in each histogram for filtering. When a range is selected, only the segments in the ring that are within this range stay colored. The other ones are visually moved into the background by gray grid lines. Each frequency can be filtered individually. In the interview study, we learned that the domain experts are often interested in the top [X]% of values. Hence, we implemented the buttons above the histograms that allow for a fast selection of those values.

The last **connectivity histogram view (4)** *(Overview, Locate, Compare)* at the bottom in the middle of the interface shows the distribution of the connection strengths (lines within the circle). The user can again use a range filter, and all connections not contained within this range are not displayed. We also implemented brushing and linking throughout the dashboard to identify the selected ROI(s) in all views.

Prototype B *(bar charts and heatmap)* (see Figure 3b) represents the design from low-fidelity prototype 1 and consists of three main views. The **brain view (1)** *(Overview, Locate)* is the same as in prototype A and represents the patient's brain with 72 ROIs. On top of the interface, five bar charts show the metric value of every ROI in each frequency. This **bar view (2)** *(Overview, Locate, Compare)* allows users to compare the values through the length of the bar rather

than by color saturation (as in prototype A). The bar charts have the same scale on the y-axis to visually compare the different bar charts. A tooltip showing the exact value appears while hovering over a bar.

The last part of this interface is the **heatmap view (3)** *(Overview, Locate, Compare)*. This view displays the connectivity between the ROIs for a selected frequency, and each column (ROI) is vertically aligned with the corresponding bars. Connectivity values are encoded through color saturation: brighter colors indicate lower values than darker shades. Hovering a cell allows the user to see the exact connectivity value. Additionally, we support navigation in the heatmap by highlighting the cell's row and column as well as the bars in the *bar view (2)*. On the bottom of this interface, the user has the same options as in prototype A: selecting subnetworks, selecting a frequency for the connectivity to be displayed in the heatmap, and filtering the top percent of values for the metric and connectivity.

VI. EVALUATION

We conducted a study where we used three methods to evaluate the effectiveness, efficiency, and subjective preference of the two high-fidelity prototypes:

- a task-based within-subject study where participants solve four tasks on each prototype while thinking out loud,
- a semi-structured interview after completion of the tasks, and
- a questionnaire aiming at usability and aesthetics.

A. Participants

We targeted two participant groups to evaluate the two prototypes: participants with domain knowledge in medical research (domain experts) and without domain knowledge (lay people), both having little or no visualization experience (see Table I). We recruited four participants between 27 and 35 years old and twelve participants between 20 and 55 for the expert and lay groups, respectively, through purposive sampling. While none of them have advanced visualization experience, they are familiar with basic chart types. Despite

TABLE I
DEMOGRAPHICS OF THE PARTICIPANTS, INCLUDING THEIR OVERALL PROTOTYPE PREFERENCE.

	ID	Gender	Age	Highest Degree	Job	Domain	Overall Preference
Lay people	P_1	Male	27	Master	PhD student	Computer science	A
	P_2	Male	29	Bachelor	Full-time employment	Computer science	B
	P_3	Male	20	Apprenticeship	Full-time employment	Relocation management	A
	P_4	Male	28	Master	Full-time employment	Software development	B
	P_5	Male	28	Master	Full-time employment	Civil engineering	B
	P_6	Female	27	Master	Full-time employment	Computer science	B
	P_7	Male	29	Bachelor	Part-time employment	Computer science	A
	P_8	Female	29	Master	PhD student	Computer science	A
	P_9	Male	28	Master	Part-time employment	Computer science	B
	P_{10}	Female	25	High school	Bachelor student	Statistics	B
	P_{11}	Female	51	High school	Full-time employment	Social work	B
	P_{12}	Female	55	High school	Full-time employment	Public service	B
Experts	E_1	Male	35	PhD	Postdoc researcher	Clinical neuroscience	B
	E_2	Male	34	PhD	Postdoc researcher	Neuroscience	A
	E_3	Male	27	Master	PhD student	Biomedical engineering	B
	E_4	Male	31	PhD	Postdoc researcher	Clinical neuroscience	B

our effort to aim for gender-balanced recruitment in both participant groups for the study, we, unfortunately, had no female participants in the domain expert group. This is because this specific data set is currently mainly investigated by male researchers in the research lab which we collaborated with. The study structure was mostly the same for both groups, with a minor difference in how the data was explained (see subsection VI-B). The tasks used in the lay-participant group did not require domain knowledge and focused solely on the visualizations. We believe the lay participants' results strengthen the user study as they can be used to compare and validate the domain experts' results. We conducted individual sessions with nine participants at our research lab, while the others were conducted through remote sessions. All participants gave written consent to participate in the user study and none of them reported any color vision deficiencies.

B. Task-based Within-subject Study

Our two interfaces use the Vue.js framework combined with D3.js [54] to draw the charts. They were optimized to be displayed in the Google Chrome web browser on a 27-inch monitor with a resolution of 2560 by 1440 pixels.

In the first part of the study, we showed participants the two dashboards, one at a time. To alleviate a potential priming bias, we alternated the prototype presented first in the study. The study started by introducing the data to the participants. While we used the original brain network data with the expert group, we changed the context for our lay participants. Our goal was to make the data more relatable so that no knowledge barrier might confuse the lay participants, and they could entirely focus on the visualizations. We replaced the medical terminology with more relatable terms. Specifically, the participants were told that the data about brain activity related to five different feelings (instead of frequency bands), which were supposed to be measured at the points showing the ROIs in the prototype. After the study, we told the participants about the EEG context, clarifying that the terminology was used to simplify and focus the task. In the next step, we introduced the first dashboard and explained its different parts,

allowing questions during this demonstration. Afterward, the participants were asked to complete four tasks while thinking aloud.

We based the tasks on the findings from the earlier requirement phase (see subsection IV-B). These showed that the domain experts might focus on the whole brain and specific parts of the brain and how these parts interact with each other. This is true for metrics on a nodal and a global level. Hence, for the first two tasks (task 1 and task 2), we asked the participants to consider all 72 ROIs (the whole brain) and only a subset for the two subsequent tasks (task 3 and task 4). We also found that the domain experts are especially interested in analyzing metrics across frequency bands, which is reflected in task 1 and task 3, and connectivity between ROIs, which is reflected in task 2 and task 4.

The following paragraphs show the task specification, which we provided as a printout to the participants so that they could revisit it during the study. Additionally, we explain how the tasks are solved with each prototype. This explanation was not provided to the participants during the study. We relate each evaluation task to the high-level tasks identified in subsection IV-B, as seen in parentheses.

Task 1 (*Overview, Locate, Compare*): "Which of the five frequencies has the highest metric value?" (domain experts) / "Which of the five feelings has the highest metric value?" (lay people)

- Prototype A: Select the ring segment with the highest color saturation of all five frequency bands/feelings in the **ring view (2)**.
- Prototype B: Select the highest bar of all five frequency bands/feelings in the **bar view (2)**.

Follow-up question: "Which region in this frequency has the highest metric value?" (domain experts) / "Which region in this feeling has the highest metric value?" (lay people)

- Prototype A: Identify the region's name with the highest color saturation in the previously found frequency band/feeling in the **ring view (2)**.
- Prototype B: Identify the name of the highest bar in the previously found frequency band/feeling in the **bar view**

(2).

Task 2 (*Overview, Locate, Compare*): "What is the strongest connection between these regions in each frequency band?" (domain experts) / "What is the strongest connection between these regions in each feeling?" (lay people)

- Prototype A: Identify the ROI-pair connected by the line with the highest color saturation in each frequency band/feeling in the **ring view (2)**.
- Prototype B: Identify the ROI-pair of the cell with the highest color saturation in each frequency band/feeling in the **heatmap view (3)**.

Before continuing with the next two tasks, we asked participants to create three subnetworks containing five, three, and five ROIs, respectively (see Figure 4).

Task 3 (*Cluster, Locate, Compare*): "In which frequency band does ROI 21 have the lowest metric value?" (domain experts) / "In which feeling does ROI 21 have the lowest metric value?" (lay people)

- Prototype A: Find the frequency band/feeling where ROI 21 has the lowest color saturation in **ring view (2)**.
- Prototype B: Find the frequency band/feeling where ROI 21 has the lowest bar in **bar view (2)**.

Task 4 (*Cluster, Locate, Compare*): "What is the strongest connection between these regions in each frequency band?" (domain experts) / "What is the strongest connection between these regions in each feeling?" (lay people)

- Prototype A: Identify the ROI-pair connected by the line with the highest color saturation in each frequency band/feeling in the **ring view (2)**.
- Prototype B: Identify the ROI-pair of the cell with the highest color saturation in each frequency band/feeling in the **heatmap view (3)**.

C. Semi-structured Interview

After completing the tasks, we conducted a semi-structured interview consisting of two parts. The first part aimed at general feedback (especially on effectiveness and efficiency) to both prototypes. In the second part, we asked the participants to compare the prototypes and let them rank one over the other for different use cases. Throughout the interview, participants had a screenshot of the two prototypes in front of them to avoid mixing them up. The complete interview outline can be found in the supplemental material.

D. Questionnaire

The study concluded with a questionnaire the participants completed without supervision right after the interview. The questionnaire included the system usability scale (SUS) [55] to assess the prototypes' usability, the BeauVis scale [39] to assess the aesthetic pleasure of the prototypes, an overall grading of the interfaces, and a demographic section. The questionnaire can be found in the supplemental material.

TABLE II
SUMMARY OF THE EVALUATION RESULTS FOR HIGH-FIDELITY PROTOTYPES A AND B. THE FIRST BLOCK (A) SHOWS THE PARTICIPANTS' PROTOTYPE PREFERENCE FOR DIFFERENT ASPECTS. THE AVERAGE SUS SCORE (0: WORST; 100: BEST), BEAUVIS SCORE (1: WORST; 7: BEST), AND GRADE (1: BEST, 5: WORST) ARE SHOWN IN (B). PART (C) SHOWS THE AVERAGE COMPLETION TIME PER TASK FOR EACH PROTOTYPE AND PARTICIPANT GROUP. THE LAST PART (D) SHOWS THE AVERAGE SCORE (1: VERY EASY; 5: NOT EASY AT ALL) OF HOW EASY PARTICIPANTS PERCEIVED SOLVING THE TASK WITH THE RESPECTIVE PROTOTYPE.

		Domain Experts (n=4)		Lay People (n=12)	
		Prot. A	Prot. B	Prot. A	Prot. B
(a) Preference	Overall	25%	**75%**	33.3%	**66.7%**
	Task 1	50%	50%	**66.7%**	33.3%
	Task 2	0%	**100%**	25%	**75%**
	Task 3	25%	**75%**	16.7%	**83.3%**
	Task 4	0%	**100%**	41.7%	**58.3%**
	Overview	25%	**75%**	**58.3%**	41.7%
	Value	25%	**75%**	25%	**75%**
	Aesthetic	**100%**	0%	**91.7%**	8.3%
(b) Avg. Score	SUS (0–100)	73.8 (σ=14.5)	**84.4** (σ=**3.2**)	63.3 (σ=24.9)	**85.4** (σ=**13.9**)
	BeauVis (1–7)	**5.7** (σ=**1**)	5.4 (σ=0.9)	**5.5** (σ=**1.3**)	5.3 (σ=0.9)
	Grade (1–5)	2.3 (σ=1.5)	2.3 (σ=0.5)	2.4 (σ=1)	**1.8** (σ=**0.6**)
(c) Avg. Time [sec]	Task 1	98 (σ=69)	**65** (σ=**31**)	104 (σ=58)	**58** (σ=**40**)
	Task 2	217 (σ=80)	**164** (σ=**41**)	213 (σ=76)	**208** (σ=**68**)
	Task 3	62 (σ=58)	**26** (σ=**13**)	53 (σ=34)	**32** (σ=**26**)
	Task 4	150 (σ=132)	**77** (σ=**34**)	80 (σ=14)	**75** (σ=**26**)
(d) Avg. Easiness	Task 1 (1–5)	**2** (σ=**1.4**)	2.5 (σ=1.3)	**1.3** (σ=**0.5**)	1.8 (σ=1)
	Task 2 (1–5)	2.3 (σ=1)	**1.3** (σ=**0.5**)	2.3 (σ=1.2)	**1.8** (σ=**0.9**)
	Task 3 (1–5)	2.5 (σ=1.7)	**1.8** (σ=**0.5**)	2.1 (σ=1)	**1.5** (σ=**0.7**)
	Task 4 (1–5)	1.5 (σ=0.6)	**1.3** (σ=**0.5**)	1.4 (σ=0.9)	**1.3** (σ=**0.7**)

VII. RESULTS

After the initial interviews about the low-fidelity prototypes with the domain experts (see subsection V-A), the combined ring view was preferred over conventional bar charts because of its aesthetics and originality in the medical domain. However, visualization literature [13]–[16] has shown that the visualization channels used to encode the data in prototype A (metric value encoded through color saturation) are usually less effective than the ones used in prototype B (metric value encoded as the length of the bar). Our evaluation results show that the initial preference changed: prototype B was the most effective and efficient choice. The results, therefore, concur with the visualization literature. The results are summarized in Table II.

In the following section, we present the results and findings of our user study. First, we discuss observations made during the task-based within-subject study, followed by a qualitative and quantitative analysis of the interview and questionnaire.

A. Observations

We conducted one-on-one sessions, each taking about 1 hour (average time: 60 minutes for domain experts, 58 minutes for lay people). The participants were asked to think aloud while solving the tasks. During the study, they could ask questions if something was unclear, and we observed the participants' interaction with the prototypes. Additionally, we measured the time participants needed to complete the tasks and recorded the number of wrong answers (error count).

On average, solving the tasks with prototype A took domain experts longer than with prototype B. While they needed between 5 and 13.4 minutes (μ = 8.8 min; σ = 4.4 min) with prototype A, they spent between 4.3 and 6.9 minutes (μ = 5.5 min; σ = 1.3 min) with prototype B. These results show that domain experts could solve the tasks more efficiently using prototype B. The lay participants, who spent between 3.9 and 11.4 minutes (μ = 7.5 min; σ = 2 min) with prototype A and between 3.5 and 9.4 minutes (μ = 6.2 min; σ = 1.9 min) with prototype B, were also more efficient using prototype B. Except for task 2, the individual completion times confirm this result (see Table IIc).

Task 1. Domain experts needed about two-thirds (66.3%), while lay people only needed about half (55.8%) of the time used for identifying the ring segment with the highest color saturation in the *ring view* (lay: 104s; experts: 98s) to find the overall highest bar in the *bar chart view* (lay: 58s; experts: 65s). Most participants filtered for the top [X] percentage of metric values before identifying the highest value visually. While lay participants did not give any wrong answers for the tasks using prototype B, four initially gave a wrong answer when working with prototype A. Domain experts made one error using prototype A. In contrast, they did not answer incorrectly when working with prototype B.

Task 2. Finding the ROI-pair with the highest connectivity value in each frequency in the *heatmap view* in prototype B (164s) took the domain experts only about three-fourths (75.6%) of the time compared to the *ring view* in prototype A (217s). This result might be influenced by the domain experts' familiarity with heatmaps since the time needed by the lay people was nearly the same for prototype B (208s) and prototype A (213s). Concerning the error rate, the domain experts gave one wrong answer per prototype for this task, while lay people made two wrong conclusions using prototype A and gave three wrong answers using prototype B.

Task 3. Solving this task with prototype B took domain experts only about 42% of the time needed with prototype A. Another speed-up was observed with lay participants: they solved the task with prototype B in 60.4% of the time needed with prototype A. Even though the participants had to compare bars on top of each other rather than being aligned on a common x-axis, they still needed less time than comparing the color saturation of the ring segments. While the solution of two domain experts was initially wrong using

prototype A, there were no errors using prototype B. All lay participants solved the tasks correctly with prototype B, while only one lay person gave a wrong answer with prototype A.

Task 4. The last task was similar to task 2, with the only difference being that the task had to be solved for a subset of 13 ROIs. Domain experts were almost twice as fast with prototype B (77s) than with prototype A (150s). The difference in completion times for lay participants is not as significant as for the domain experts: lay people only needed five seconds longer with prototype A (80s) compared to prototype B (75s). While domain experts did not give any wrong answers using both prototypes, only one lay participant gave a wrong answer using prototype A.

We also asked participants to rate how easy it was to solve each task with the prototypes (1: very easy; 5: not easy at all). While participants found solving most tasks easier using prototype B (see Table IId), therefore confirming the results above, task 1 was considered easier using prototype A. The participants mentioned this mainly because they could see the distribution of the metric values directly in the *metric histogram view*, not the *ring view*.

B. Interview

We recorded all interviews and took notes during the interview. For the analysis, we extracted themes from the gathered knowledge and clustered them for each participant group individually. In the following subsections, we present our interview findings.

1) Interviews with Domain Experts: While the layered doughnut chart (paper prototype 3) was favored by the domain experts in the low-fidelity interview study, three out of the four experts ($E_{1,3,4}$) in the final study preferred prototype B *(bar charts and heatmap)* in terms of effectiveness and efficiency. They found interface B especially helpful for quickly identifying connectivity patterns in the heatmap ($E_{1,4}$). This might be related to their familiarity and experience with this type of visualization. Another participant (E_3) mentioned that he enjoyed having separate bar charts for the metric in each frequency as this allowed him to get a global overview of all ROIs. He also noted that the space was used more efficiently than in prototype A *(layered doughnut chart)*, resulting in less whitespace overall. However, two experts ($E_{2,4}$) found it hard to compare bars across frequencies since the corresponding bar charts were on top of each other. One expert (E_3) also mentioned his concerns about prototype B not scaling well for more than 72 regions since the screen real estate is limited and comparing bars might get more difficult. Suggestions for improvement that were mentioned include using a dynamic or logarithmic color scale to facilitate the differentiation of the color saturation in the *heatmap view* ($E_{3,4}$), encoding metric values additionally through color saturation in the *bar view* (E_4), and implementing a feature to save and pre-load subnetworks (E_3). While three experts ($E_{2,3,4}$) saw the potential for the current version of prototype B to facilitate data analysis in their daily work, one expert (E_1) reported he

would need more customization features before he would use it daily.

Only one of the four participants (E_2) preferred prototype A. He liked the overall presentation of the connections and frequencies in the aggregated *ring view* and the filter option only showing connections of interest. Even though he initially found prototype B easier to use, he supports his choice for prototype A by saying, "[prototype B] being easier today does not mean that it is always easier – prototype A gets better over time." The main advantage of prototype A that was mentioned ($E_{1,2,4}$) is the compact *ring view* that allows for a good overview and presentation of the metric and connectivity values across all frequencies. E_1 said that "[prototype] B is better for analyzing the data; [prototype] A is better for presenting the data". E_3 commented that if he "had both tools, [he] would only use [prototype] B. [He is] not the biggest fan of A as it is not so easy to interpret".

These results underline our findings in the other two parts of the study, which state that prototype A is still preferred in terms of aesthetics. Still, prototype B is considered more effective and efficient for data analysis. When asked about potential use cases for prototype A over prototype B, our participants mentioned situations in which they must present the data (e.g., team meetings, research papers) and want to attract attention. The *histogram views* to filter data attributes were also well received by three experts ($E_{1,3,4}$) who suggested including them in prototype B. Two participants ($E_{1,4}$) mentioned that it is difficult to see exact values in prototype A since they are mainly encoded through color saturation. Related to that, one participant (E_3) found it particularly "difficult to compare the color saturation in the small ring segments". While all experts saw a potential benefit of prototype A for their daily work (especially for data presentation), most of them ($E_{1,3,4}$) still preferred prototype B for analyzing the data as it is "intuitive and easier to use than the other prototype".

2) Interviews with Lay People: The lay participants' responses mostly aligned with the domain experts' opinions. The results show that eight of them ($P_{2,4,5,6,9,10,11,12}$) preferred prototype B *(bar charts and heatmap)*. They based their decision on several reasons, one being that it is easier for them, especially in the *bar view*, to see data values more accurately than in prototype A *(layered doughnut chart)* ($P_{1,3,4,6,8,10,11,12}$). More concretely, four participants ($P_{3,5,7,9}$) explicitly pointed out that they preferred the bars over the encoding with color. The encoding of the connectivity values in the *heatmap view* through a heatmap was received well ($P_{7,9,10}$). They point out that the heatmap conveys more information visually than the *ring view* in prototype A ($P_{4,6,8}$) and that the color saturation is easier to distinguish in a horizontal and vertical grid than in a circular layout ($P_{2,10,12}$). Three participants ($P_{5,6,8}$) found prototype B more user-friendly and intuitive than the other prototype. The main weakness identified in prototype B was related to the scalability of the interface. Since the initial interface features all 72 ROIs, participants ($P_{1,8,9}$) found that there is an overload of visual elements on the screen when all regions are shown and P_1 added that comparing bars is difficult when there are so many of them. Therefore, the interface would not scale well as

more ROIs are added, making the analysis more difficult ($P_{1,3,5}$). Since the connectivity matrices in our data set are all symmetric, half of the participants ($P_{2,3,7,8,10,11}$) found it unnecessary to display the whole heatmap. This was an explicit design decision, as our requirement analysis showed that connectivity matrices in other data sets might not be symmetric. It is, however, a suggestion for improvement to allow the user only to display half of the heatmap.

Even though most participants preferred prototype B overall, almost half of them ($P_{1,3,4,5,8}$) enjoyed the clean and compact representation of the data in prototype A. $P_{1,5,6,8}$ liked the aggregation of the data attributes (metric and connectivity) in a single *ring view*. Additionally, all participants pointed out that the *metric histogram view*, as well as the *connectivity histogram view*, are very helpful, primarily since they can be used as a filter for the *ring view*. The main weaknesses identified in prototype A are that the usability is worse than in the other prototype ($P_{5,6,8,9,10}$) and that training ($P_{6,8,9}$) is needed before the tool can be used efficiently. $P_{5,7,9,10,11}$ mentioned that the *ring view* and its ring segments are too small for the displayed data, thus negatively influencing the tool's usability. Participants also had a hard time differentiating color saturation in the *ring view* for both data features: metric ($P_{2,5,7,8,10}$) and connectivity ($P_{8,9,10}$). This was underlined by P_1, who said that "[data feature] values are easier estimated through bars than color saturation"; P_3 found that "hovering for comparing segments with higher color saturation is tedious". All participants agreed that they find prototype A more aesthetic. One of the participants (P_6) commented, "I'm just not a big fan of the circle. However, if I had to visualize results and print them in a research paper, then I would go with [prototype] A". A few lay participants ($P_{2,4,9}$) also pointed out that in both prototypes the *brain view* did not add any value for them since the brain representation is for users with domain knowledge who can interpret the data.

3) Suggestions for Improvement: Both participant groups shared suggestions to improve the prototypes after completing a task or in the interview afterward.

Prototype A. Two participants (P_6, E_1) were interested in seeing the connectivity values in prototype A, so they suggested displaying the value when the user hovers over a connection. E_1 further mentioned that he would be interested in simultaneously seeing connections from multiple frequencies. However, displaying too many connections in the *ring view* could result in an unreadable mesh of lines. E_3 suggested adding line thickness as a second encoding channel for the connectivity value to make strong connections even more noticeable to the analysts. Concerning data filtering, some participants ($P_{1,2,4,9}$) wanted additional lower-level filter options (e.g., letting users input a custom percentage for the top percentages of values). One expert (E_3) mentioned that he had a hard time identifying bins that only contain a few elements in the *connectivity histogram view* when there were bins containing high numbers of elements. He suggested rescaling the histogram when the user selects a range to alleviate this problem.

Prototype B. Both histogram views (connectivity and metric)

benefitted all participants in prototype A, making it evident that including them in prototype B could improve it further. This is also indicated by participants' suggestions to add more filter options ($P_{1,2}$) and allow users to choose custom percentages for the top [X] values ($P_{2,4}$). Another feature request would allow users to sort bar charts in ascending/descending order ($P_{2,6}$) to identify maxima and minima quickly. Two participants (P_6, E_4) suggested adding color saturation as a second encoding channel for the metric value in the bar chart to help users identify higher values more easily. A dynamic color scheme updating whenever a filter or subnetwork selection is applied was suggested ($E_{1,4}$) for the *heatmap view* to distinguish data points easily. Loading pre-defined subnetworks and saving custom subnetwork selections could save the user time during the configuration (E_3).

C. Questionnaire

The quantitative analysis results mirrored the qualitative analysis's findings. The results of the questionnaire (see Table IIb) show that prototype B has a higher SUS score for domain experts (84.4) as well as lay people (85.4) than prototype A (experts: 73.8; lay people: 63.3). This almost corresponds to "excellent" usability for prototype B, whereas the usability of prototype A is rated between "ok" and "good" [56]. We used the BeauVis scale [39] in its recommended version: a 7-point Likert scale with five items (enjoyable, likable, pleasing, nice, and appealing) from "strongly disagree" (1) to "strongly agree" (7). Even though the usability score of prototype B is higher, the BeauVis rating of prototype B (experts: 5.4; lay people: 5.3) is close to the rating of prototype A (experts: 5.7; lay people 5.5). This shows that our effort for a clear and accessible design paid off by also getting high results in terms of aesthetics for prototype B. However, we did not reach the same level as prototype A, which is still better, and the result illustrates the participants' aesthetic preference for prototype A, as indicated clearly in the interviews.

We also asked participants to grade each prototype from 1 (best) to 5 (worst). While the average grade of domain experts is the same for both prototypes (2.3), lay people assigned better grades to prototype B (1.8) than to prototype A (2.4).

VIII. DISCUSSION

Quantitative and qualitative results show that we successfully developed effective and efficient prototypes for analyzing multi-frequency network data. The visual analysis tools facilitate the visual exploration of the data. Here we address important aspects of the design and evaluation process.

The importance of high-fidelity prototyping. The evaluation with domain experts in the early low-fidelity prototyping stages showed that their preference for a tool design was affected by its aesthetics and novelty. Another design proposal was less preferred even though it was considered more optimal based on visualization theory. However, the overall preference shifted once we implemented the low-fidelity designs as high-fidelity prototypes and allowed users to interact with the systems. This preference shift underlines the importance of low- and high-fidelity parallel prototyping. Visualization

designers must create multiple prototypes in parallel to explore the design space. Furthermore, our findings in this study show that it is not enough to stop at the low-fidelity phase since preferences can change in further iterations when new factors (e.g., interactivity) come into play. Hence, more and better rapid prototyping tools are needed to bridge this gap between low- and high-fidelity prototyping.

There is a lack of design guidelines for affective intent. Designing visualization goals solely based on their task-based objectives (e.g., efficiently conveying facts or insights) might not lead to the optimal solution. Other factors (e.g., aesthetics, engagement), which can be used to achieve affective intent [34], must also be considered in characterizing the goal at the early stages of the requirement analysis. While our results show that prototype B outperformed prototype A in terms of effectiveness and efficiency for analyzing the data, prototype A was considered well-suited, especially for presenting the data, as it was considered more engaging. While one participant explicitly said, "[prototype] B is better for analyzing the data; [prototype] A is better for presenting the data," other participants articulated similar thoughts agreeing with this statement. We argue that this trade-off between performance and aesthetics influences the design space significantly. The latter are often not explicit goals of expert visualization tools. Our results demonstrate that acknowledging these subjective types of goals changes the design. However, our community lacks design guidelines for affective intent. Our work shows that there is plenty of room to incorporate aesthetic aspects into prototype B. We believe creating engaging prototypes without sacrificing half the performance is possible, as was the case comparing our two prototypes.

Be aware of familiarity with visual representations. Similar to aesthetics, familiarity with specific visual representations could influence users' subjective preferences. Dasgupta [41] argues that "familiar visualizations and visual encodings within a domain are often in conflict with the optimality of visualization design, in terms of how well they support the data analysis or communication tasks.". This was not reflected in our findings in the initial interview study during the low-fidelity prototyping phase. Domain experts preferred the circular prototype to the one with bar charts and a heatmap regardless of their familiarity with heatmaps since they are prevalent in network analysis. However, this might have been because participants prioritized aesthetics and novelty over familiarity. We believe familiarity can play an important role in subjective preference and should be considered carefully while designing and evaluating a visual analysis tool.

Limitations. We aimed to qualitatively and quantitatively validate and compare the expert group's findings with participants from outside the medical domain. Our non-representative participant sample was, in the case of the lay group, not biased or primed by domain knowledge or constrained by familiar visualizations. Hence, they could focus primarily on usability and aesthetics rather than functionality. This approach strengthened our findings and validated the expert group's results.

As a result of this design study, we developed two high-fidelity prototypes that showed great potential to improve

the effectiveness and efficiency of data analysis for medical experts. However, to evaluate these factors of the systems in production, more implementation iterations are needed to cover all aspects of the data. A deployed productive system was out of scope for this project and is considered future work.

IX. CONCLUSION AND FUTURE WORK

In this paper, we reported on a design study for a visual analysis tool that facilitates exploratory data analysis of multi-frequency network data. We conducted an in-depth requirement analysis and developed two high-fidelity prototypes with domain experts in an iterative design process. We evaluated the prototypes regarding effectiveness, efficiency, and subjective preference both qualitatively and quantitatively. The evaluation results with domain experts and lay people, both with little or no visualization experience, showed that subjective preference and objective task performance do not always align. Furthermore, subjective preference can change throughout the design process, underlining the importance of high-fidelity prototyping, where users interact with their own data for the first time. We also found that subjective factors (e.g., aesthetics, familiarity) are important when defining design goals for bespoke visual analysis tools. However, as a community, we lack the ability to design for affective intent.

In the future, we envision further investigating the relationship between different subjective factors that influence the visualization goal and, therefore, also the tool design. Furthermore, we want to improve the prototypes based on the user study feedback and develop a consolidated version that can be deployed to production.

REFERENCES

[1] B. Shneiderman, "The eyes have it: a task by data type taxonomy for information visualizations," in *Proceedings 1996 IEEE Symposium on Visual Languages*, 1996, pp. 336–343.

[2] A. Burns, C. Lee, R. Chawla, E. Peck, and N. Mahyar, "Who do we mean when we talk about visualization novices?" in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23. New York, NY, USA: Association for Computing Machinery, 2023.

[3] F. Barollo, M. Hassan, H. Petersen, I. Rigoni, C. Ramon, P. Gargiulo, and A. Fratini, "Cortical pathways during postural control: New insights from functional eeg source connectivity," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 72–84, 2022.

[4] M. Carboni, P. De Stefano, B. J. Vorderwülbecke, S. Tourbier, E. Mullier, M. Rubega, S. Momjian, K. Schaller, P. Hagmann, M. Seeck, C. M. Michel, P. van Mierlo, and S. Vulliemoz, "Abnormal directed connectivity of resting state networks in focal epilepsy," *NeuroImage: Clinical*, vol. 27, p. 102336, 2020.

[5] J. Royer, B. C. Bernhardt, S. Larivière, E. Gleichgerrcht, B. J. Vorderwülbecke, S. Vulliémoz, and L. Bonilha, "Epilepsy and brain network hubs," *Epilepsia*, vol. 63, no. 3, pp. 537–550, 2022.

[6] P. van Mierlo, Y. Höller, N. K. Focke, and S. Vulliemoz, "Network perspectives on epilepsy using eeg/meg source connectivity," *Frontiers in Neurology*, vol. 10, 2019.

[7] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany, "An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest," *NeuroImage*, vol. 31, no. 3, pp. 968–980, 2006.

[8] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C. J. Honey, V. J. Wedeen, and O. Sporns, "Mapping the structural core of human cerebral cortex," *PLOS Biology*, vol. 6, no. 7, pp. 1–15, 07 2008.

[9] I. Rigoni, J. Rue Queralt, K. Glomb, M. Preti, N. Roehri, S. Tourbier, L. Spinelli, M. Seeck, D. Van De Ville, P. Hagmann, and S. Vulliemoz, "Structure-function coupling increases during interictal spikes in temporal lobe epilepsy: a graph signal processing study," *Clinical Neurophysiology*, 2023.

[10] M. X. Cohen, "Where does eeg come from and what does it mean?" *Trends in Neurosciences*, vol. 40, no. 4, pp. 208–218, 2017.

[11] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *NeuroImage*, vol. 52, no. 3, pp. 1059–1069, 2010, computational Models of the Brain.

[12] R. Bender and S. Lange, "Adjusting for multiple testing—when and how?" *Journal of Clinical Epidemiology*, vol. 54, no. 4, pp. 343–349, 2001.

[13] T. Munzner, *Visualization Analysis and Design*. A K Peters/CRC Press, 2014.

[14] W. S. Cleveland and R. McGill, "Graphical perception: Theory, experimentation, and application to the development of graphical methods," *Journal of the American Statistical Association*, vol. 79, no. 387, pp. 531–554, 1984.

[15] J. Mackinlay, "Automating the design of graphical presentations of relational information," *ACM Trans. Graph.*, vol. 5, no. 2, p. 110–141, apr 1986.

[16] J. Heer and M. Bostock, "Crowdsourcing graphical perception: Using mechanical turk to assess visualization design," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 203–212.

[17] G. J. Quadri and P. Rosen, "A survey of perception-based visualization studies by task," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 12, pp. 5026–5048, 2022.

[18] R. Amar, J. Eagan, and J. Stasko, "Low-level components of analytic activity in information visualization," in *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, 2005, pp. 111–117.

[19] J. Talbot, V. Setlur, and A. Anand, "Four experiments on the perception of bar charts," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2152–2160, 2014.

[20] C. Nothelfer and S. Franconeri, "Measures of the benefit of direct encoding of data deltas for data pair relation perception," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 1, pp. 311–320, 2020.

[21] A. Srinivasan, M. Brehmer, B. Lee, and S. M. Drucker, "What's the difference? evaluating variations of multi-series bar charts for visual comparison tasks," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–12.

[22] C. C. Gramazio, K. B. Schloss, and D. H. Laidlaw, "The relation between visualization size, grouping, and user performance," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1953–1962, 2014.

[23] M. Waldner, A. Diehl, D. Gračanin, R. Splechtna, C. Delrieux, and K. Matković, "A comparison of radial and linear charts for visualizing daily patterns," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 1, pp. 1033–1042, 2020.

[24] S. Diehl, F. Beck, and M. Burch, "Uncovering strengths and weaknesses of radial visualizations—an empirical approach," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 935–942, 2010.

[25] D. A. Szafir, "Modeling color difference for visualization design," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 392–401, 2018.

[26] D. Borland and R. M. Taylor Ii, "Rainbow color map (still) considered harmful," *IEEE Computer Graphics and Applications*, vol. 27, no. 2, pp. 14–17, 2007.

[27] K. Reda, "Rainbow colormaps: What are they good and bad for?" *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 12, pp. 5496–5510, 2023.

[28] I. M. Gołbiowska and A. Çöltekin, "Rainbow dash: Intuitiveness, interpretability and memorability of the rainbow color scheme in visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 7, pp. 2722–2733, 2022.

[29] K. Reda, P. Nalawade, and K. Ansah-Koi, "Graphical perception of continuous quantitative maps: The effects of spatial frequency and colormap design," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–12.

[30] R. M. Karim, O.-H. Kwon, C. Park, and K. Lee, "A study of colormaps in network visualization," *Applied Sciences*, vol. 9, no. 20, 2019.

[31] C. C. Gramazio, D. H. Laidlaw, and K. B. Schloss, "Colorgorical: Creating discriminable and preferable color palettes for information visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 521–530, 2017.

[32] M. Harrower and C. A. Brewer, "Colorbrewer.org: An online tool for selecting colour schemes for maps," *The Cartographic Journal*, vol. 40, no. 1, pp. 27–37, 2003.

[33] M. Tory and T. Moller, "Human factors in visualization research," *IEEE Transactions on Visualization and Computer Graphics*, vol. 10, no. 1, pp. 72–84, 2004.

[34] E. Lee-Robbins and E. Adar, "Affective learning objectives for communicative visualizations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 1, pp. 1–11, 2023.

[35] A. V. Pandey, A. Manivannan, O. Nov, M. Satterthwaite, and E. Bertini, "The persuasive power of data visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2211–2220, 2014.

[36] A. Lau and A. Vande Moere, "Towards a model of information aesthetics in information visualization," in *2007 11th International Conference Information Visualization (IV '07)*, 2007, pp. 87–92.

[37] D. Filonik and D. Baur, "Measuring aesthetics for information visualization," in *2009 13th International Conference Information Visualisation*, 2009, pp. 579–584.

[38] A. V. Moere and H. Purchase, "On the role of design in information visualization," *Information Visualization*, vol. 10, no. 4, pp. 356–371, 2011.

[39] T. He, P. Isenberg, R. Dachselt, and T. Isenberg, "Beauvis: A validated scale for measuring the aesthetic pleasure of visual representations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 1, pp. 363–373, 2023.

[40] A. Dasgupta, J.-Y. Lee, R. Wilson, R. A. Lafrance, N. Cramer, K. Cook, and S. Payne, "Familiarity vs trust: A comparative study of domain scientists' trust in visual analytics and conventional analysis methods," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 271–280, 2017.

[41] A. Dasgupta, *Experts' Familiarity Versus Optimality of Visualization Design: How Familiarity Affects Perceived and Objective Task Performance*. Springer International Publishing, 2018, pp. 75–86.

[42] D. Slayback, S. Abdali, J. Brooks, W. D. Hairston, and P. Groves, "Novel methods for eeg visualization and visualization," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2018, pp. 1–5.

[43] M. ten Caat, N. M. Maurits, and J. B. Roerdink, "Design and evaluation of tiled parallel coordinate visualization of multichannel eeg data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 1, pp. 70–79, 2007.

[44] ——, "Data-driven visualization and group analysis of multichannel eeg coherence with functional units," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 4, pp. 756–771, 2008.

[45] D. P. Wulandari, Y. K. Suprapto, A. I. Juniani, T. F. Elyantono, S. W. Purnami, and W. R. Islamiyah, "Visualization of epilepsy patient's brain condition based on spectral analysis of eeg signals using topographic mapping," in *2018 International Conference on Computer Engineering, Network and Intelligent Multimedia (CENIM)*, 2018, pp. 7–13.

[46] Z. Fang, L. Yao, and Z. Qian, "Research on multi-parameter visualization technology of brain function based on eeg," in *2019 International Conference on Medical Imaging Physics and Engineering (ICMIPE)*, 2019, pp. 1–4.

[47] M. Gavrilescu and F. Ungureanu, "Enhanced three-dimensional visualization of eeg signals," in *2015 E-Health and Bioengineering Conference (EHB)*, 2015, pp. 1–4.

[48] T. Mullen, C. Kothe, Y. M. Chi, A. Ojeda, T. Kerth, S. Makeig, G. Cauwenberghs, and T.-P. Jung, "Real-time modeling and 3d visualization of source dynamics and connectivity using wearable eeg," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 2184–2187.

[49] K. R. Christopher, A. Kapur, D. A. Carnegie, and G. M. Grimshaw, "Implementing 3d visualizations of eeg signals in artistic applications," in *2013 28th International Conference on Image and Vision Computing New Zealand (IVCNZ 2013)*, 2013, pp. 364–369.

[50] V. C. Pezoulas, A. Athanasiou, G. Nolte, M. Zervakis, A. Fratini, D. I. Fotiadis, and M. A. Klados, "Fclab: An eeglab module for performing functional connectivity analysis on single-subject eeg data," in *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, 2018, pp. 96–99.

[51] M. Sedlmair, M. Meyer, and T. Munzner, "Design study methodology: Reflections from the trenches and the stacks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2431–2440, 2012.

[52] E. Kerzner, S. Goodwin, J. Dykes, S. Jones, and M. Meyer, "A framework for creative visualization-opportunities workshops," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 748–758, 2019.

[53] C. Knoll, A. Cetin, T. Möller, and M. Meyer, "Extending recommendations for creative visualization-opportunities workshops," in *2020 IEEE Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV)*, 2020, pp. 81–88.

[54] M. Bostock, V. Ogievetsky, and J. Heer, "D³ data-driven documents," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2301–2309, 2011.

[55] J. Brooke, "SUS: A 'quick' and 'dirty' usability scale," in *Usability Evaluation in Industry*, P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland, Eds. Taylor and Francis, 1996, pp. 189–194.

[56] A. Bangor, P. T. Kortum, and J. T. Miller, "An empirical evaluation of the system usability scale," *International Journal of Human–Computer Interaction*, vol. 24, no. 6, pp. 574–594, 2008.