



An Interactive and Multimodal Virtual Mind Map for Future Workplace

David Kut'ák^{1*}, Milan Doležal¹, Bojan Kerous¹, Zdenek Eichler², Jiří Vašek² and Fotis Liarokapis¹

¹ Human Computer Interaction Laboratory, Faculty of Informatics, Masaryk University, Brno, Czechia, ² Konica Minolta Laboratory Europe, Brno, Czechia

Traditional types of mind maps involve means of visually organizing information. They can be created either using physical tools like paper or post-it notes or through the computer-mediated process. Although their utility is established, mind maps and associated methods usually have several shortcomings with regards to effective and intuitive interaction as well as effective collaboration. Latest developments in virtual reality demonstrate new capabilities of visual and interactive augmentation, and in this paper, we propose a multimodal virtual reality mind map that has the potential to transform the ways in which people interact, communicate, and share information. The shared virtual space allows users to be located virtually in the same meeting room and participate in an immersive experience. Users of the system can create, modify, and group notes in categories and intuitively interact with them. They can create or modify inputs using voice recognition, interact using virtual reality controllers, and then make posts on the virtual mind map. When a brainstorming session is finished, users are able to vote about the content and export it for later usage. A user evaluation with 32 participants assessed the effectiveness of the virtual mind map and its functionality. Results indicate that this technology has the potential to be adopted in practice in the future, but a comparative study needs to be performed to have a more general conclusion.

OPEN ACCESS

Edited by:

Francesca Odone,
University of Genoa, Italy

Reviewed by:

Yoshifumi Kitamura,
Tohoku University, Japan
Francesco Isgro,
University of Naples Federico II, Italy

*Correspondence:

David Kut'ák
433409@mail.muni.cz

Keywords: virtual reality, immersive environments, multimodal interaction, voice recognition, collaborative interfaces, mind maps

Specialty section:

This article was submitted to
Robot and Machine Vision,
a section of the journal
Frontiers in ICT

Received: 18 January 2019

Accepted: 04 July 2019

Published: 31 July 2019

Citation:

Kut'ák D, Doležal M, Kerous B,
Eichler Z, Vašek J and Liarokapis F
(2019) An Interactive and Multimodal
Virtual Mind Map for Future
Workplace. *Front. ICT* 6:14.
doi: 10.3389/fict.2019.00014

1. INTRODUCTION

Modern technologies offer new opportunities for users to communicate and interact with simulated environments, to quickly find information and knowledge when needed and also to learn anytime and anywhere (Sharples, 2000). When humans communicate, the dynamics of this social interaction are multimodal (Louwerse et al., 2012) and provide several different patterns, like entrainment of recurrent cycles of behavior between partners, suggesting that users are coordinated through synchronization and complementarity, i.e., mutual adjustments to each other resulting in corresponding changes in their behavior occurring during the interaction (Sadler et al., 2009). When users are engaging in a collaborative task, then synchronization takes place between them through multiple modalities. Some of these include gestures, facial expression, linguistic communication (Louwerse et al., 2012) or eye-movement patterns (Dale et al., 2011). By designing multimodal interfaces, it is possible to improve the accessibility and usability of mind mapping systems to achieve a natural and intuitive experience to the users.

Traditional types of mind maps involve some means of the visual organization of information. During the past few years, there have been some initial approaches to developing 3D mind maps to boost productivity. A number of human-computer interaction (HCI) technologies exist nowadays that address this topic, and some of them tend to work well in specific situations and environments. A virtual reality (VR) solution for shared space can be realized in several ways depending on the required level of immersion of the end-user in addition to the requirements of the application. One of the most accepted definitions states that immersion refers to the objective level of sensory fidelity a VR system provides whereas presence addresses user's subjective psychological response to a VR system (Slater, 2003). The level of immersion is directly interrelated with the end-user perception and promoted if tactile devices are used. VR, therefore, has the potential to augment processes of our everyday life and to mitigate difficult problems.

An open problem is a co-location in the office of the future environment. Meeting with people on site is costly because often people need to travel to the location from various cities or countries. The Internet enables to connect these people from the technical point of view while VR allows to achieve much more immersive and natural cooperation. Nowadays, several immersive virtual solutions that address co-location exist, such as collaborative applications for cave automatic virtual environment (CAVE) systems (Cruz-Neira et al., 1993), where users do not wear head-mounted displays (HMDs) and are able to see each other and interact directly due to their location in the same physical space. Nowadays, collaborative immersive VR allows users to be co-located in the same space or in different locations and achieve communication through internet (Dolezal et al., 2017). The availability of current HMDs allows for the easier creation of strongly immersive user experiences. Typical sub areas of shared spaces for VR include visualization, communication, interaction and collaboration, and VR-based mind map workflow overlaps and relies on all four aspects of the experience.

The main focus of this research is on multimodal VR collaborative interfaces that facilitate various types of intelligent ideation/brainstorming (or any other mostly creative activity). Participants can be located in different environments and have a common goal on a particular topic within a limited amount of time. Users can group (or ungroup) actions (i.e., notes belonging in a specific category) and intuitively interact with them using a combination of different modalities. Ideally, the multimodal interface should allow users to create actions (i.e., post-it note) and then post it on the virtual mind map using one or more intuitive methods, such as voice recognition, gesture recognition, and through other physiological or neurophysiological sources. When a task is finished, users should be able to access the content and assess it.

This paper presents a novel shared virtual space where users are immersed in the environment (i.e., same meeting room) and participate in a multimodal manner (through controllers and voice recognition). Emphasis is given on the (a) shared VR environment; (b) effective performance of the multimodal interface; and (c) assessment of the whole system as well as the

interaction techniques. The tasks are typically moderated by one or two individuals who facilitate the process, take care of the agenda, keep the schedule, and so on. Such ideation exercise can be used on various occasions but is typically associated with the creative process in the company where the output of the exercise is uncertain before it is executed. During a particular task users can create and manipulate shared nodes (equivalent to real-world sticky notes), modify their hierarchical or associative relationships and continuously categorize, cluster, generalize, comment, prioritize, and so on. Moderator's role is to guide the discussion and regulate the voting phase.

2. RELATED WORK

With the appearance of novel interfaces and mediums, such as VR and increasing presence of sensors and smart devices in our environment, it has become apparent that the typical way we interact with the computer is changing rapidly. Novel ways of achieving fluent interaction in an environment saturated with sources of useful behavioral or physiological data need to be explored to pave the way for new and improved interface designs. These interfaces of the future hold the promise of becoming more sophisticated, informative, and responsive by utilizing speech/gesture recognition, novel peripherals, eye-tracking, or affect recognition. The role of multimodal interfaces is to find ways to combine multiple sources of user input and meaningful ways of leveraging diverse sources of data in real time to promote usability. These sources can be combined in one of three levels, as outlined in Sharma et al. (1998), that depends on the level of integration (fusion) of distinct sources of data. There is a real opportunity to mitigate the difficulties of a single modality-based interface by combining other inputs.

Collaborative Virtual Environments (CVEs) may be considered as shared virtual environments operating over a computer network (Benford et al., 2001). They have different application domains ranging from health-care (McCloy and Stone, 2001; Rizzo et al., 2011), cultural heritage (White et al., 2007; Liarokapis et al., 2017), education (Redfern and Galway, 2002; Pan et al., 2006; Faiola et al., 2013; Papachristos et al., 2013) to psychology (Loomis et al., 1999), and neuroscience (Tarr and Warren, 2002). One of the main disadvantage of CVEs is that they do not support non-verbal communication cues (Redfern and Galway, 2002). The typical solution to overcome this problem is to include a representation of the participants in a form of avatars. Although this does not solve the problem, it allows for some form of limited non-verbal communication. As a result, participants of CVEs can interact with objects or issue commands while being observed by the virtually co-located collaborator.

The benefits, design, and evaluation in the field of designing speech and multimodal interactions for mobile and wearable applications were recently presented (Schaffer and Reithinger, 2016). Having a multimodal VR interface can be beneficial for several complex operations as well as new applications, ranging from automotive styling to museum exhibitions. The multimodality can also be achieved by providing different visual

representations of the same content so the user can choose the most suitable one (Liarokapis and Newman, 2007). The main principle of the concept of multimodality is that it allows users to switch between different types of interaction technologies. Multimodal interfaces can greatly expand the accessibility of computing to diverse and non-specialist users, for example, by offering traditional means of input like the keyboard and also some uncommon ones like specialized or simplified controllers. They can also be used to promote new forms of computing and improve the expressive power and efficiency of interfaces (Oviatt, 2003).

The flexibility of multimodal interfaces allows for the better alternation of input modalities, preventing overuse and physical damage arising from a repeated action during extended periods of use. Furthermore, multimodal interfaces can be used to provide customizable digital content and scenarios (White et al., 2007) while, on the other hand, they can bring improvements by combining information derived from audio and visual cues (Krahnstoever et al., 2002). Acquisition of knowledge is also augmented through the use of such multimodal MR interface compared to a traditional WIMP-based (Windows, Icons, Menu and Pointer) interface (Giraudeau and Hachet, 2017). In fact, one example implementation of a mind-map based system reported in Miyasugi et al. (2017) allows multiple users to edit a mind map by using hand gestures and voice input and share it through VR. Initial comparative experiments with representative mind map support software (iMindMap¹) found that the task completion time for creating and changing the key images was shorter than that of iMindMap. Currently, there are several software alternatives for mind map creation; XMind² and iMindMap being the most famous ones, but most of these solutions are aimed at a single user and support only traditional non-VR enabled interfaces. In the world of VR applications, Noda³ is one of the most progressive alternatives. Noda utilizes spatial mind maps with nodes being positioned anywhere in the three-dimensional space, while it does not offer collaboration possibilities.

Having a three-dimensional mind map presents some advantages like increased ability to exploit spatial thinking and theoretically infinite place for storing ideas. On the other hand, spatial placement might decrease the clarity of the mind map as some nodes might be hidden behind the user or other nodes. The one-to-one correspondence with traditional mind mapping software is lost as well, which makes it hard to export the results for later processing and review. This would decrease the usability of the outputs created inside the VR, and it is the reason why our approach works with two-dimensional (2D) mind maps.

Another alternative tool is Mind Map VR⁴ offering more or less same functionalities as Noda. An interesting feature of the

Mind Map VR is the ability to change the surroundings for a different looking one. When concerned about collaborative platforms, rumii⁵, CocoVerse (Greenwald et al., 2017), and MMVR⁶ (Miyasugi et al., 2017) are closely related to our work. In its core, all of these systems provide users a possibility to cooperate in VR. Rumii is, however, aimed mostly at conferencing and presentation, while CocoVerse is aimed at co-creation mainly via drawing so although mind mapping is theoretically possible, the application is not designed for this purpose.

As MMVR is focused on mind mapping and online collaboration in VR, it does have a similar purpose as our application. MMVR utilizes hand gestures to create mind maps with nodes positioned in three-dimensional space. On the contrary, in our system, VR controllers are used for the interaction, and the map canvas is two-dimensional. Similarly to Noda, authors of MMVR decided to take a slightly different approach than we did regarding the mind map representation. Besides already mentioned things, we tried to make the mind mapping process more related to the real-world one—VR controller acts as a laser pointer while 2D canvas is a virtual representation of a whiteboard. MMVR also excludes features related to brainstorming, such as voting.

3. SYSTEM ARCHITECTURE

The traditional way of brainstorming using post-it notes presents several drawbacks related to a reshuffling or modifying of notes during the whole process as post-it notes often fall from the wall and trying to do it multiple times makes them not staying on the wall anymore. Besides that, taking multiple notes from multiple places to move them to some other place is cumbersome. Mapping relationships between post-it notes is another difficult task, one needs to make lines among post-it notes and to label the lines if needed, but to do this one must often re-shuffle the post-it notes to make the relationships visible. Elaborating on a particular topic (for example deeper analysis requiring more post-it notes) in one part of the exercise is also difficult as all of the other post-it notes need to be reshuffled again to make space for the new exercise. It is challenging to draw on the post-it note when needed and then stick it on the wall. Finally, post-exercise analysis is difficult; it typically involves a photograph of the result and then manual transcription into a different format; for example, brainstorming “tree” and word document as meeting minutes. If it is necessary to perform a remote brainstorming, the disadvantages are much more significant, and there does not exist a flawless solution.

Our system is designed in such a way to try to take the best of both the interpersonal brainstorming and software approaches and merge it into one application. The main part of our system is a canvas with a mind map. The canvas serves as a virtual wall providing users space to place their ideas and share them with others. All nodes are positioned at this 2D wall to keep the process as close as possible to the real-world while also providing

¹iMindMap official website, <https://imindmap.com/> (accessed October 15, 2018).

²XMind official website, <https://www.xmind.net/> (accessed October 15, 2018).

³Noda—Steam Store, <https://store.steampowered.com/app/578060/Noda/> (accessed March 29, 2019).

⁴Mind Map VR—Steam Store, https://store.steampowered.com/app/885250/Mind_Map_VR_VR/ (accessed March 29, 2019).

⁵rumii official website, <https://www.rumii.net> (accessed October 15, 2018).

⁶MMVR—Full name of the application, i.e., not an acronym.

similar visual style as conventional mind mapping software tools. To simplify collective manipulations, our system introduces a simple gesture. The user draws a shape around the nodes he or she wishes to select and then simply moves them around using a cursor button which appears after the selection ends. This feature is described in more detail in section 3.2. One of the big challenges of VR technology lies in the interaction side. Existing speech-to-text tools were integrated into our system to allow users to use voice-based interaction.

When the process of brainstorming finishes, voting about the best ideas takes place. In real-world exercise, it is hard to make sure that all participants obey the voting rules. Some participants might distribute a different number of points than they should, or they can be influenced by other participants. Our tool provides a voting system, ensuring that the voting points are distributed correctly and without being influenced by other participants. Brainstorming exercise is usually performed with more people at the same place while one of them serves as a moderator. This might not be a problem for teams sharing a workspace, but when it is desired to collaborate with people being physically far away, things are much more complicated. Our tool provides a VR environment where all users can meet, although they might be located at different places in the world. The overview of the different parts of the system is shown in **Figure 1**. **Figure 2** shows a screenshot of the application while brainstorming is in progress.

The software was developed in Unity and C#. The networking system presented in Dolezal et al. (2017) was incorporated into the application. Interaction with the HMDs is possible thanks to Virtual Reality Toolkit (also VRTK) plugin. To run the application, SteamVR is required as it provides application interfaces for VR devices. Even if the application is designed to be operational with HMDs, it is also possible to use it just with personal computers like desktop or laptop—without HMD. In this case, the keyboard and mouse are required as input devices. If a microphone is present as well, speech recognition service can be still utilized as an input modality. Regarding the HMDs, the system is implemented to work with HTC Vive (Pro) and one controller. Our focus was on making the controls as simple as possible. For this reason, the user is required to work only with two of the controller's buttons (touchpad and trigger) to have complete control over the system features. When the user presses the touchpad, laser pointer is emitted from the VR controller. Trigger serves as an “action” button; when the user is pointing at some element, pressing the trigger initiates the appropriate action. Video showing some of the system functionality and interaction is in the **Supplementary Material**.

3.1. Map Nodes

Map nodes are the core component of the system. Each node is represented by a visual element having color and a label. Map nodes can be modified in several ways - they can be moved, deleted, modified, and being updated with new visual styles. It is also possible to make a relation between nodes represented by lines between appropriate nodes. Two types of relations were implemented - parent/child and general ones. Former ones represent a “strong” relation where each node can have only one parent and is dependent on its antecedents - when some of them

are moved or removed, this node is modified as well. Latter type of relations is there mainly for semantic purpose. Each node can have as many of these relations with other nodes as desired while persisting independency. Modifications of the nodes are done using radial menus shown while pointing at a node. This allows users to perform the most important actions while still being focused on the mind map. The content of the node's radial menu is shown in **Figure 3**. Blue buttons provide the functionality to add the aforementioned relations to other nodes. Red button removes nodes while the green button creates a new node as a child of the currently selected node. The last button allows users to record a text for this node.

3.2. Selection of Multiple Nodes

The multiple selection is handled in such a way that a user is required to draw a shape around the nodes he wishes to select. Selection shape is drawn using a controller by pressing touchpad and the trigger buttons at the same time while pointing at the canvas. When the selection is finished, the user can move the selected nodes or perform some actions provided by the appropriate radial menu. Thanks to this feature, selecting several nodes and changing their visual style, position, or relations is quite simple. In the background, this feature is based on a point-in-polygon test. The selection shape is visually represented as a long red line (technically a polyline) which is converted into a polygon (based on vertices of individual line segments of a polyline) after the drawing is finished. Then, for each node, it is computed whether its center lies in a resulting polygon.

3.3. Voice Recognition

Language technology is easier to accept for participants only if it is implemented in an intuitive and easy to use way (Schaffer and Reithinger, 2016). Text input is a big challenge for all VR applications as a traditional keyboard cannot be properly used due to not being visible. It also disallows the user to move freely. The most straightforward idea is to use a virtual keyboard, but this approach is not very effective, especially with only one controller. For this reason, we decided to use speech-to-text technology. Our system is using Wit.ai service to provide this functionality. The user uses an appropriate button in the radial menu to start the recording, says the desired input, and then ends the recording. The rest is handled by our system in cooperation with the service mentioned above. In the background, voice recognition operates in such a way that the user's recording is converted into an audio file which is uploaded to the Wit.ai servers. These servers process the audio and return a response containing the appropriate text. The whole process is running on a separate thread to not block the program while speech is transformed into the result.

3.4. Voting

Voting is a special state of the application during which nodes cannot be edited and which provides an updated user interface where each node is accompanied by plus and minus buttons and a text box with points. This allows participants to assign points easily. Voting consists of several rounds where during each round, one color to be voted about is chosen. Voting is led by a

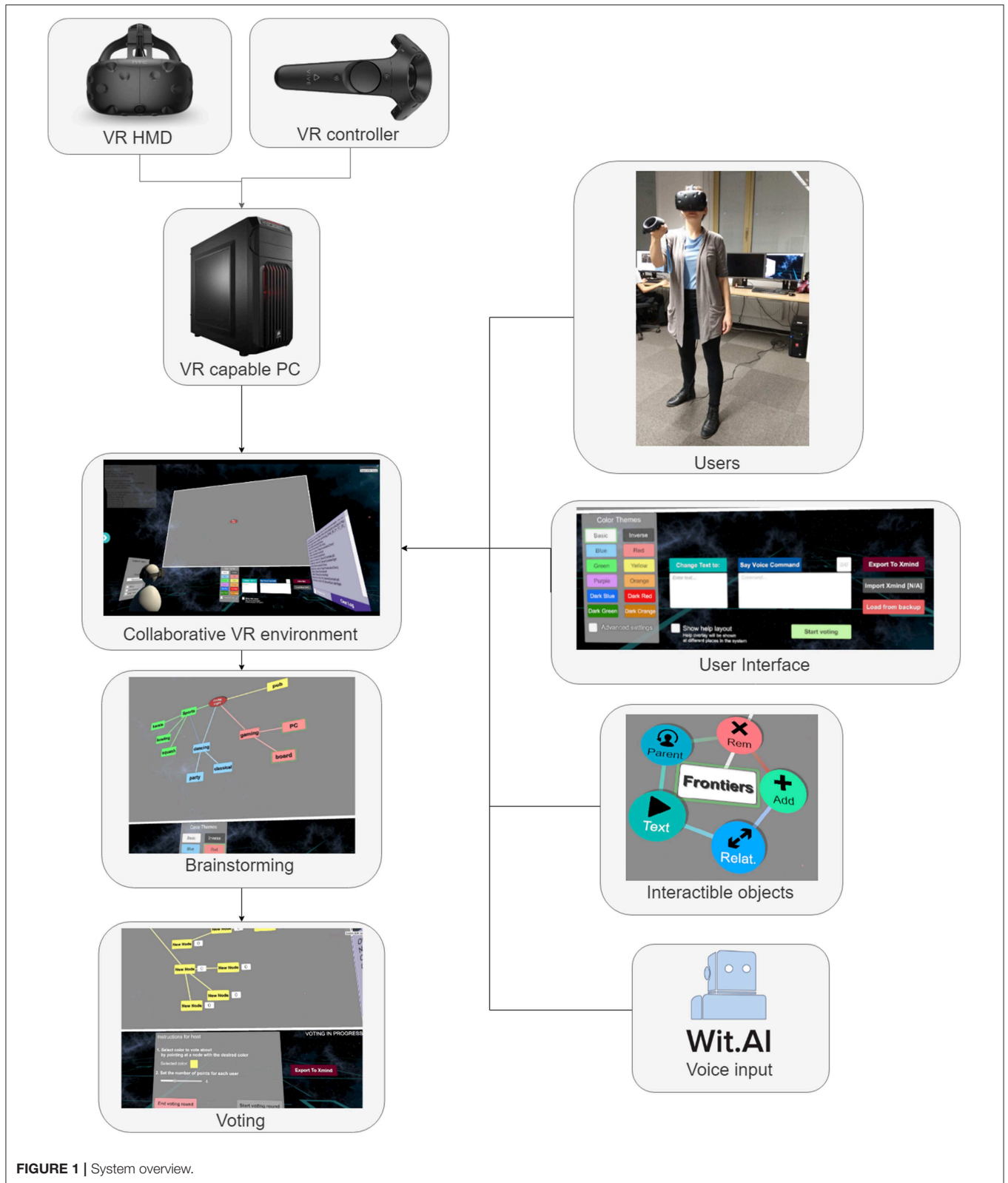
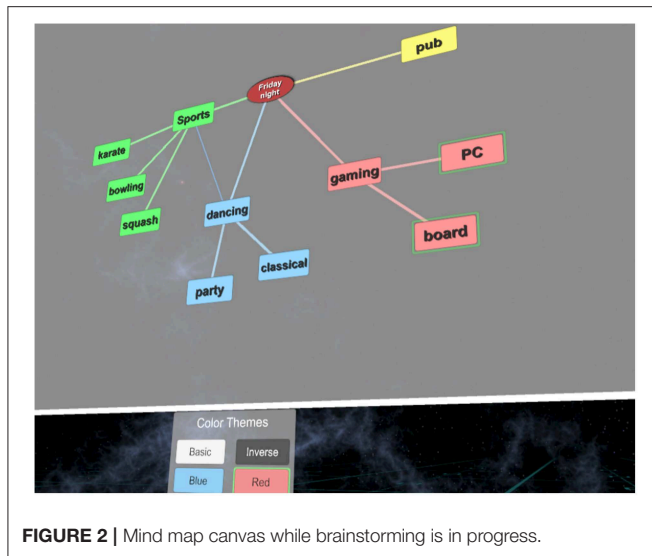


FIGURE 1 | System overview.

moderator of the brainstorming who decides about the colors to vote about and assigns the number of points to distribute between the voted ideas. For each such voting round, participants see only

the number of points they assigned, and they have to distribute all points. When a moderator tries to end the voting round, the system checks whether all participants distributed their points



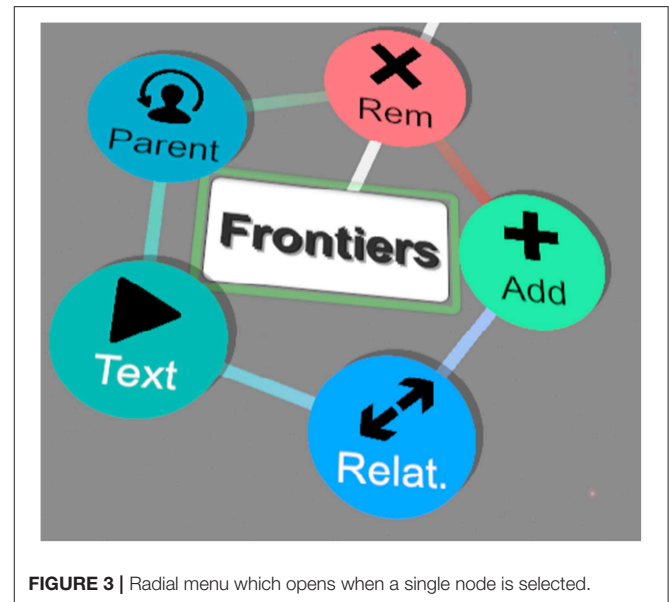
and if not, then the round cannot be closed. When the voting ends, all participants see the summary of points for each node. Winners in each category are made visually distinct.

3.5. Online Collaboration

The core of the network-related part of the system is Collaborative Virtual Environments (CVR) platform (Dolezal et al., 2017) utilizing Unity Networking (UNET) technology (its core structure is shown at **Figure 4**). The system works on a host/client model, in which one user is a server and a client at the same time while other users are just clients. Each user is represented as an abstract representation of a capsule (as shown in **Figure 5**) with HMD and VR controller in hands. Both the positions of the avatar and controller are synchronized over the network. Online collaboration also includes a system of node-locking, preventing users from modifying a node while another user is currently working with it, and controller-attached laser pointer which allows users to get immediate feedback about the place they or other user are pointing to. Regarding the node-locking, this functionality is based on the concept of node managers. When a client points at a node, system locally checks whether the node is locked for this client or not. If the node is already locked, it is not selected. Otherwise, the client sends a request to the server to lock this node. Server processes these requests sequentially and for each request verifies whether the claimed node is without a manager, otherwise denies the request. If the node has no manager yet, the server makes requesting user the manager of this node and sends an remote procedure call (RPC) to the rest of the clients that this node is locked. If a node is deselected, unlock message is sent to the server, which then propagates this information down to the clients.

3.6. Mind Map Export

At any time during the brainstorming, users of our system can export the mind map into the open-source XMind format. Possibilities of the format are fully utilized, therefore most of the important information like node visuals, relations, or points



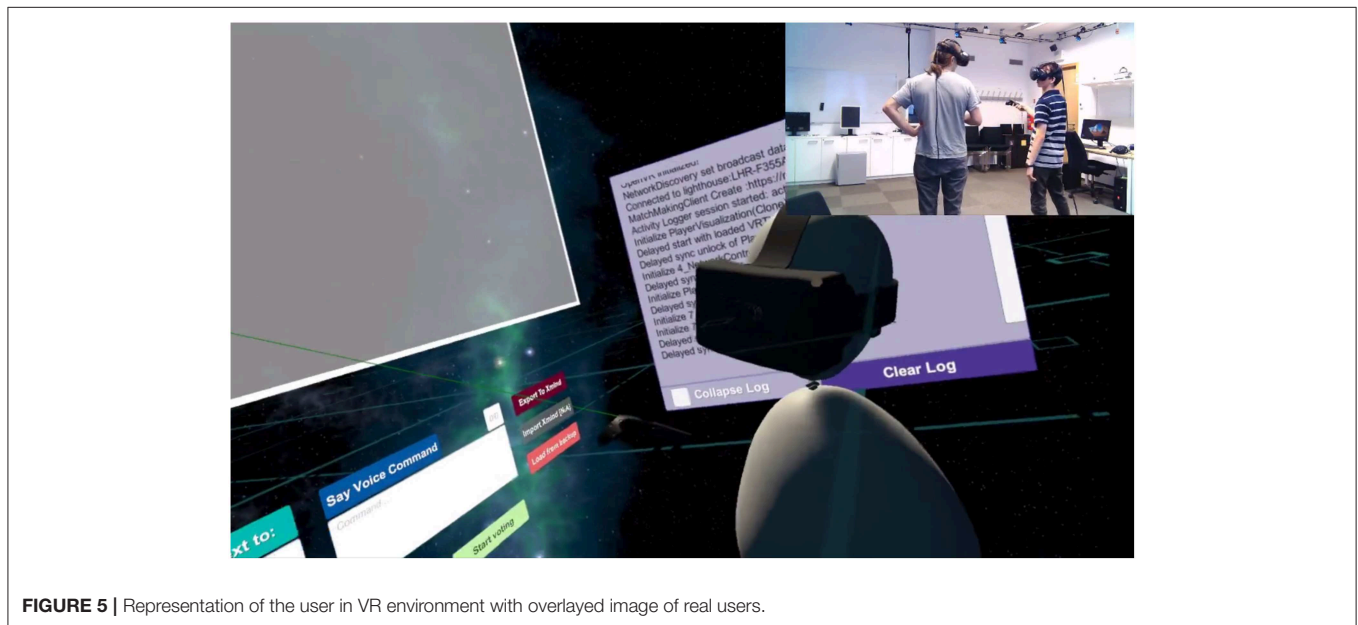
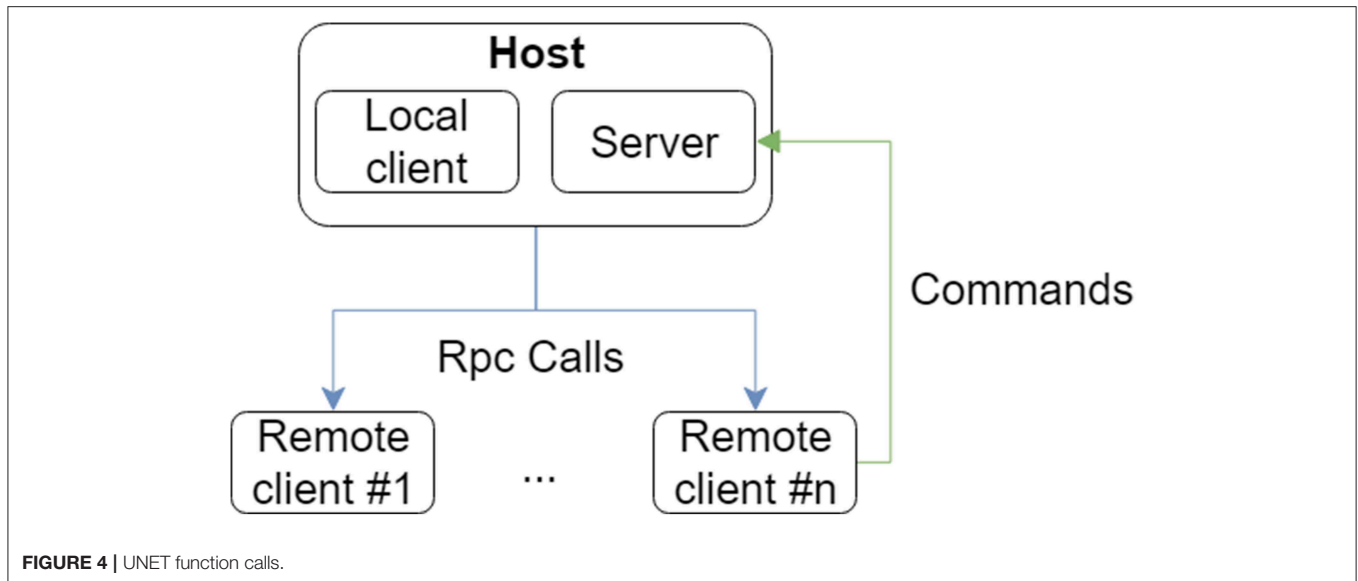
received during voting are exported. Support of mind map export provides an ability to access the brainstorming results later on or even modify them in other software tools. The mind map is also regularly saved to a backup file which is stored in a custom JavaScript Object Notation (JSON)-like format. This format was designed to be as simple and as fast as possible while still storing all the necessary information. The backup file can be loaded at any time during the brainstorming, making it therefore possible to restore mind mapping progress in case of some failure like lost internet connection.

4. METHODOLOGY

This section presents the methodology of the experiment performed for collecting information about the application.

4.1. Participants and Questionnaires

The study consisted of a total of 32 healthy participants (19 males, 13 females) and testing took place in pairs (16 individual groups). Participants were a voluntary sample, recruited based on their motivation to participate in the study. All subjects signed informed consent to participate in the study and to publish their anonymous data. They were aged from 18 to 33 years old, and all of them were regular computer users. They were rather inexperienced with mind maps and generally had some experience with remote collaboration. The very first step was to explain the workflow of the experiment to participants. Then, statistical and demographic data were collected. After the completion of the experiment, subjects were asked to fill in questionnaires related to the recent experience. Two questionnaires were used. The first one focused on measuring presence in VR (Witmer and Singer, 1998; Witmer et al., 2005). The second questionnaire aimed at assessing the cognitive workload and was based on the NASA Task Load Index (Hart,



2006). The subjects were also asked to fill in a free-form debriefing session questionnaire, where they provided qualitative feedback for the whole experiment.

4.2. Procedure

The procedure of user testing consisted of two main steps. Participants were located in different rooms, and during the first 10–15 min, depending on the skill of the individual user, each of them was alone in the virtual environment while being introduced to the system and presented with its features. While trying the system functionality, the participant's feedback was gathered. The second part of the evaluation consisted of participants trying to brainstorm on a scenario. To assess the

functionality of the system, a number of different brainstorming scenarios were designed. The topics that were chosen include: (a) How to cook an egg properly, (b) What is best to do on Friday night, (c) How will artificial intelligence take over the world, (d) Wine selection for dinner, and (e) Propose your own topic. The given topic for the experiment was “What is best to do on Friday night.” The process was directed by a moderator and contained the following steps:

1. Participants were asked to present possibilities how to spend Friday night using nodes on the wall together
2. Participants were asked to assign other specific properties to ideas from previous exercise and to use different color of nodes

3. Each participant was asked to select one idea and add nodes describing concrete proposal
4. Participants were asked to present to each other results of previous exercise
5. Participants ran a voting session. One of the participants took a role of a voting moderator, the second one was acting as a voting participant.

Time of completion for each of the steps was measured and the behavior of the participants was monitored in order to get another type of feedback.

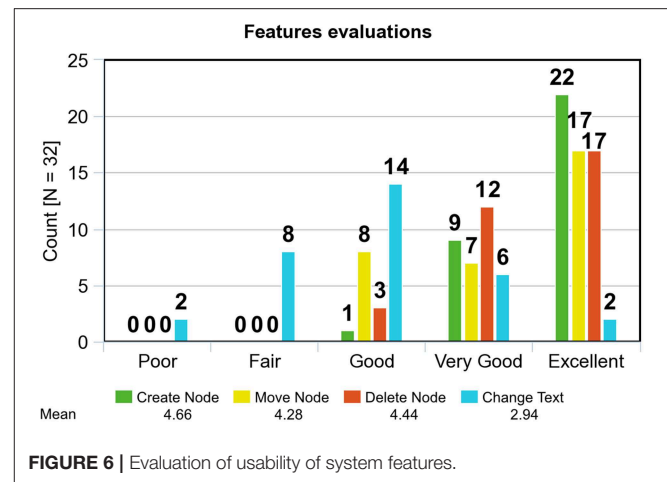
5. RESULTS

5.1. Qualitative Results

The participants provided us with valuable feedback necessary for further improvements. The feedback was gathered not only by direct communication with participants but also by watching their behavior during the actual scenario. Thanks to this approach, it was possible to collect useful information during the whole time of the testing. During the debriefing, we asked participants whether they know any other tools which can be used for remote brainstorming or collaboration and if they can find some (dis)advantages of our system in comparison to these tools. The mentioned tools included services like Skype, Google Docs/Hangouts, Slack, Facebook, Team Speak, IBM Sametime, and video conferencing platforms.

The most commonly mentioned advantage of our system was immersion. Quoting one of the users, "It makes you feel like you are brainstorming in the same room on a whiteboard (...)." Similarly, the ability to see what is going on was praised, mainly the fact that the users are represented as avatars with a laser pointer instead of abstract rectangles with names as is common in some applications. Another advantage, in comparison to other tools known to participants, was an absence of outside distractions. Also, it was mentioned several times, that our application is more fun than comparable tools. Regarding the disadvantages, the inability to see other peoples' faces was mentioned. Many users also pointed out the necessity to have appropriate hardware, i.e., that such an application requires more equipment and preparations than the tools they know. Another drawback was physical discomfort, mainly the requirement to wear a HMD. Some users mentioned that it takes more time to get familiar with the interface in comparison to common tools they know. Also, the speed with which the ideas can be generated was considered by some participants to be slower than in the case of conventional platforms.

At the end of the experiment, users gave us general feedback about the application. We expanded this feedback by insights we collected by observing their behavior. The most mentioned drawback of the system was the position of the mind map canvas. It was positioned too high, forcing users to look up all the time, which resulted in physical discomfort and decreased the readability of nodes which were positioned at the top of the canvas. Some users also had some remarks about the speed and reliability of the used speech-to-text service. The application itself was generally considered as responsive, although the user



interface has space for improvement. Mainly at the beginning, users tended to forget to stop the voice recording after they finished saying the desired text for a node. Also, the difference between parent-child relations and general relations was not clear enough. Regarding the environment, some participants spoke favorably about the space surroundings; on the other hand, one user mentioned that there exists a risk of motion sickness or nausea for some people. Others mentioned that the text at the top of the canvas is hardly readable. Unfortunately, the pixel density of HMD is not good enough at such distance, so it is necessary to consider this drawback when designing similar types of applications. We also noticed that the system of node locking sometimes slows down the work.

Participants also provided some ideas for further improvements. One mentioned that it would be good to have the possibility to decide whether to hide or show activity (including laser pointers) of other people during the voting. Another one pointed out that the current selection of color themes is not very visually pleasing and that it might be good to use some better color palette. One participant said that it might be useful to have more control over voice so you can mute yourself or others, for example, when saying a text for a node. Ability to change the size of the node's text would also be welcomed addition for some users. Overall, the application seemed to be quite immersive but for the price of increased physical demand and possibly slower pacing.

5.2. Quantitative Results

The first part of this section presents compound histogram summarizing participants' evaluations of core system features. Each user was assigning one (= poor) to five (= excellent) points to each feature.

Figure 6 confirms observed behavior which is that users had no major problems when trying to create or delete nodes. The delete might perform a bit worse because when a node is deleted its radial menu remains open until the user points elsewhere. Although the menu is not working anymore, it is a bit confusing that it is still present. This behavior is going to be addressed in the future to deliver a smoother user experience. Distribution of

yellow colored responses in **Figure 6** shows that the mechanism for moving nodes was not as user-friendly as desired for some participants. This might be caused by the fact that moving of nodes fails (i.e., the node returns to the previous state) when both of the controller buttons are released at the same time. This was a slight complication for some users. Red values, revealing evaluations of change text feature, have a distribution with a mean of 2.94, it can be therefore said that the speech recognition in its current state is acceptable. The question is, how would it perform if different scenario with more complicated words was used? Hence, although the performance is not entirely bad, there is a space for improvement in both the user interface and recognition quality. Then, it might be worth considering whether to stick to the current speech recognition solution or try something else. Another idea to think about is to utilize multimodality even in text input. It was not unusual that the user said a word which was recognized as a different one, but also very similar, to what he wanted, so the difference was just a few letters. It might come as handy to have a quick way of fixing these errors, either in the form of a virtual keyboard or some dictionary-like mechanism.

Table 1 presents the results obtained based on Spearman's correlation. An interesting point is the relation between stated physical demand and frustration. When users felt physical discomfort, caused, for example, by too highly placed canvas or weight of the HMD, they became more frustrated. Physical demand can be partly decreased by improving the application's interface, but as long as HMD is used, there will always be a certain level of discomfort. Another interesting output is the correlation between the TLX temporal demand and effort. Participants considering the pace of the task hurried felt that they have to work harder in order to accomplish the task. In this case, improvement of speech-to-text service might be helpful. There was also a strong correlation between answers on "How easy did you find the cooperation within the environment?" and "How quickly did you adjust to the VR environment?" A negative correlation was found between satisfaction with "change text" functionality and answers to TLX questions regarding the feeling of frustration and physical demand. Since this is a key feature from the system perspective, it is used a lot, and when the user does not feel comfortable with it, it might make him or her tired both physically and mentally. Finally, users who considered the visual display quality of HMD as distracting and unsatisfactory felt like the task was more physically demanding. This is partially

due to the technological limits of current HMDs but also certain design aspects could be improved. The idea is to improve the colors and sizes of UI elements to decrease the users' eye strain caused by the relatively low pixel density of HMDs.

5.3. Log Results

The activity of participants during the testing scenario was logged in order to get more information about their behavior as well as the effectiveness of our platform. Stored information contains general actions performed by the participant (e.g., node creation and deletion) and visualizations of mind map canvas interactions. The median of collaboration times for the scenario was 19 min and 5 s (excluding explanations of each step). Nodes were created only during the first three steps of the scenario, the median of these times is 14 min and 20 s. This accounts for an average speed of $\sim 1\text{--}2$ nodes per minute since the median of nodes created during the scenario was 24. It is worth mentioning that the speed, respectively duration, of the brainstorming depends on the creativity of the users. The fastest pair was able to create 3.3 nodes per minute on average while slowest one achieved the speed of nearly one node per minute. The relation between the

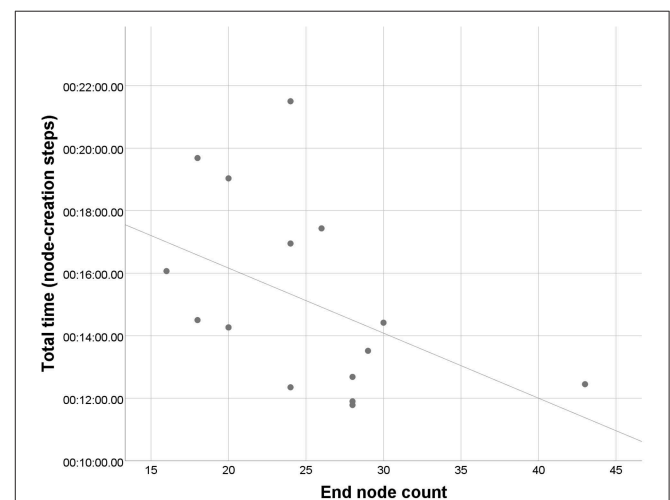


FIGURE 7 | Scatter plot of collaboration times and number of nodes (two testing pairs had exactly the same time and node count so there are only 15 visible points).

TABLE 1 | Outputs of selected Spearman's correlation coefficients.

First element	Second element	Correlation coefficient	Sig. (2-tailed)	N
TLX—physical demand	TLX—frustration	0.479	0.006	32
TLX—temporal demand	TLX—effort	0.747	<0.001	32
Ease of VR cooperation	Adjustment speed	0.528	0.002	32
Ease of VR cooperation	Use for next meeting	0.470	0.007	32
Change text satisfaction	TLX—frustration	−0.481	0.005	32
Change text satisfaction	TLX—physical demand	−0.538	0.002	32
HMD display quality dissatisfaction	TLX—physical demand	0.537	0.002	32

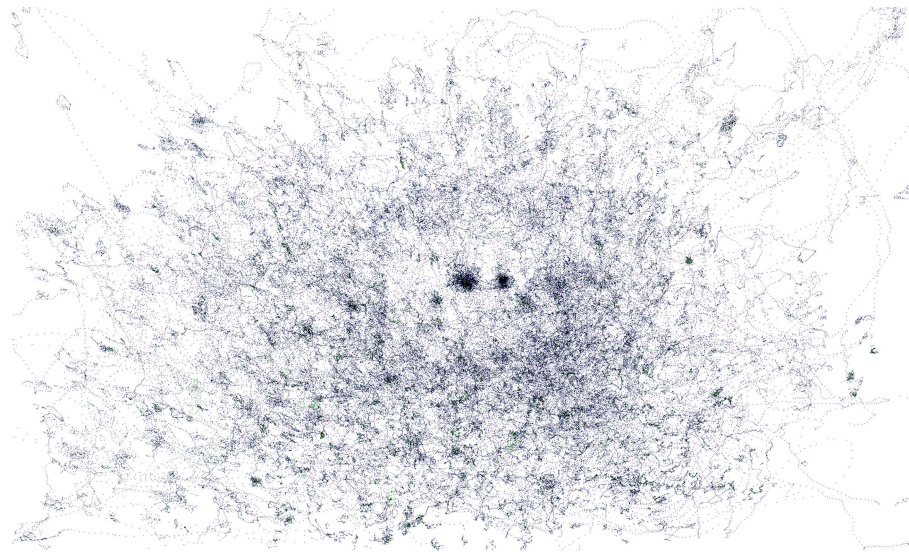


FIGURE 8 | Merged heatmap with pointer movements of all users.

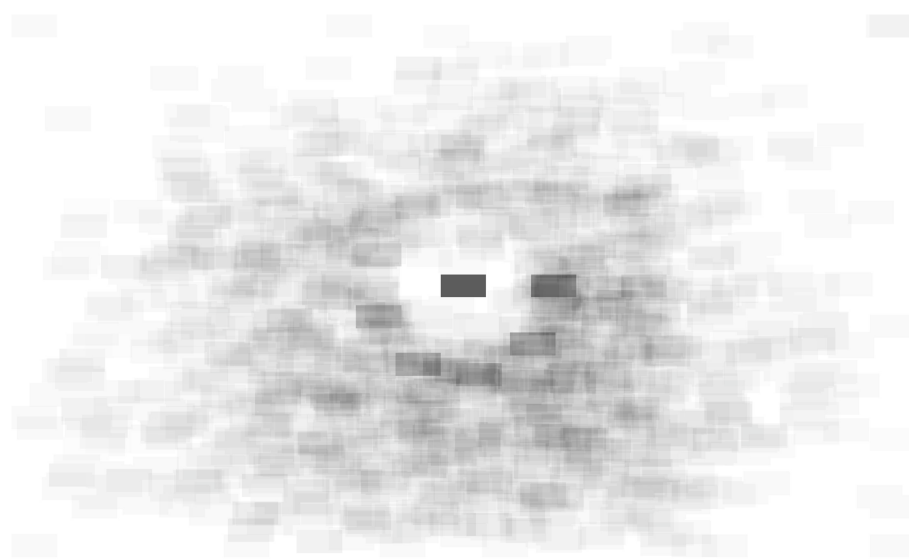


FIGURE 9 | Merged heatmap highlighting positions of selected nodes.

number of nodes at the end of the exercise and total time is shown in **Figure 7**.

This could be justified in several ways. First, the users with higher node count might have been simply more creative than the rest, and so it was easier for them to come up with new ideas. Moreover, as each step of the study was not limited by time but rather by a rough minimum of the number of nodes, participants had no problems creating more than enough nodes to continue. The flow of the session was also not interrupted so much by the time spent on thinking about possible ideas. The effect can also be caused by differences in the communication methods between participants. In any case, this confirms that the

speed of the brainstorming does not depend only on the system capabilities. Another results from the logs are shown in **Figures 8, 9**. **Figure 8** is created by merging heatmaps of all tested users. The points in the image represent positions in the mind map canvas, which were “hit” by a laser pointer while aiming at a canvas and selecting. The RGB colors determine the relative amount of hits at a given pixel with red being the most “hit” pixels while blue being the least “hit” pixels, whereas green pixels are somewhere in between. **Figure 9** shows an averaged heatmap of selected nodes of all users. This determines positions where nodes were selected for the longest time - in this case holds that the bigger the opacity is, the longer this position was covered by a selected node.

An observation regarding both heatmaps is the fact that the space near the corners of the mind map is unused. This suggests a tendency of users to place most of the nodes near the central root node. Another interesting point is the significant difference in the density of heatmap in the bottom and the upper half of

the canvas. This confirms that there might be reduced readability in the upper half of the canvas and users are therefore preferring nodes which are closer to them, i.e., at the bottom of the canvas. **Figure 9** also reveals that users generally like to move the nodes around as they wish, and they do not just stick to the default

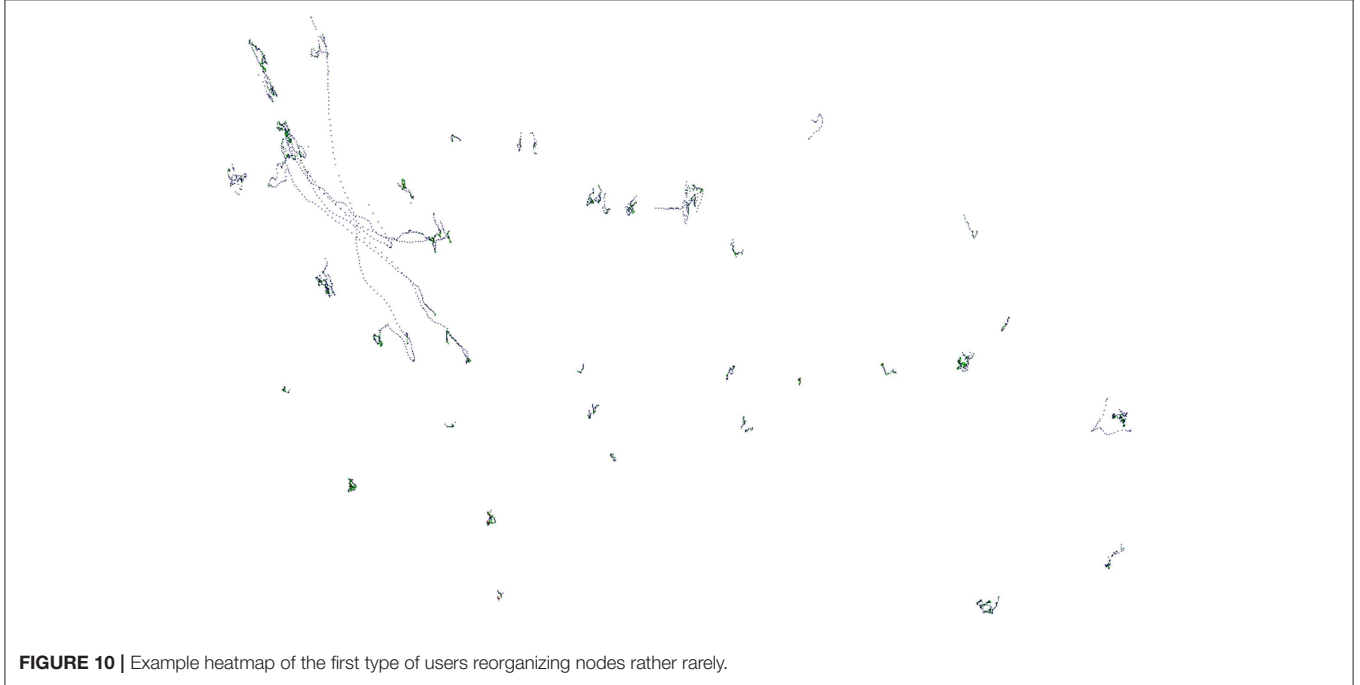


FIGURE 10 | Example heatmap of the first type of users reorganizing nodes rather rarely.

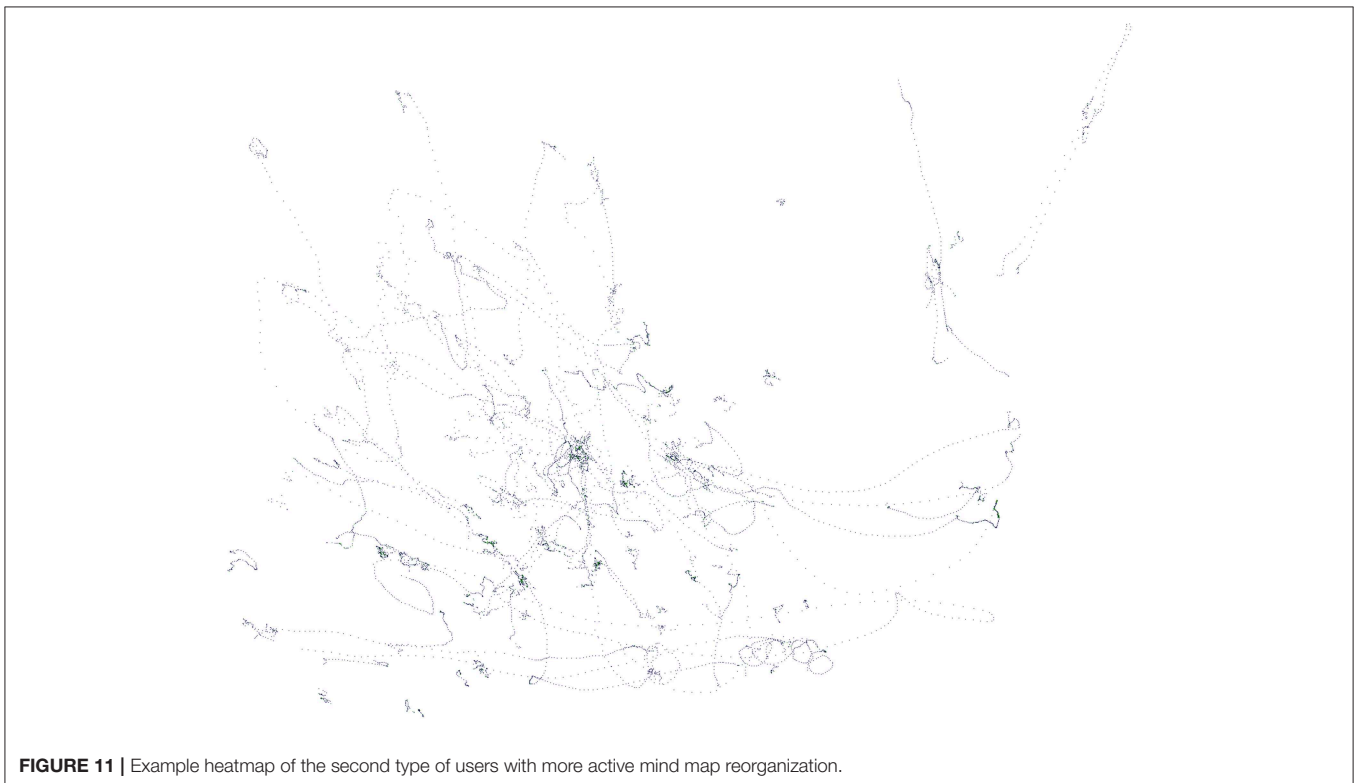


FIGURE 11 | Example heatmap of the second type of users with more active mind map reorganization.

automatic circular placement. This means that it is necessary to have a good interface for node movement. Regarding the movement, in order to be more precise, **Figures 10, 11** show two heatmaps which clusters the users into two categories. The first type is less common and prefers to stick to the default node placement and does only minor changes while the second category of users is more active in this regard. This is also related to another observed user behavior—some people use the laser pointer nearly all the time while others use it only when necessary.

6. CONCLUSIONS AND FUTURE WORK

This paper presented a collaborative multimodal VR mind map application allowing several participants to fully experience the process of brainstorming. It covered both aspects (a) idea-generation phase as well as (b) voting procedure. Multimodality was achieved based on the combination of speech and VR controller. To verify the usability of the system, experiment with 32 participants (19 males, 13 females) was conducted. Users were tested in pairs and filled several questionnaires summarizing their experience. The results indicate that the system performs according to its specifications and does not show critical problems. In terms of user feedback, comments include mainly minor issues of the usability of the environment and can be clustered as design issues.

Furthermore, there are many possibilities on how to improve and extend the application. Besides general improvements to the interface, avatars will be exchanged for some more realistic ones. Also, name tags will be added to identify individual participants. Thanks to the integrated voice solution, some speech-related features will be added, for example, automatic muting of users when they are saying a label for a node. Moreover, there is also going to be visual feedback, like icon or mouth animation, to

make it clear which user is speaking. Possibilities of hand gesture controls will be examined as well. Finally, a comparative user study will be made between traditional platforms for remote collaboration and the VR mind map to assess the advantages and disadvantages of each approach.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Research Ethics Committee of the Masaryk University with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Research Ethics Committee of the Masaryk University with approval number EKV-2018-093.

AUTHOR CONTRIBUTIONS

FL, JV, ZE, and DK conceived the research and defined the evaluation method. DK was in charge of the testing with participants with help from BK and MD. DK analyzed the user study data. MD developed the underlying networking framework. DK implemented the system presented in this paper. JV, ZE, and FL provided feedback during the development of the system. DK, BK, and FL wrote the paper.

FUNDING

This research was funded by Konica Minolta Laboratory Europe.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fict.2019.00014/full#supplementary-material>

REFERENCES

- Benford, S., Greenhalgh, C., Rodden, T., and Pycoc, J. (2001). Collaborative virtual environments. *Commun. ACM* 44, 79–85. doi: 10.1145/379300.379322
- Cruz-Neira, C., Sandin, D. J., and DeFanti, T. A. (1993). "Surround-screen projection-based virtual reality: the design and implementation of the cave," in *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '93* (New York, NY: ACM), 135–142.
- Dale, R., Kirkham, N. Z., and Richardson, D. C. (2011). The dynamics of reference and shared visual attention. *Front. Psychol.* 2:355. doi: 10.3389/fpsyg.2011.00355
- Dolezal, M., Chmelik, J., and Liarokapis, F. (2017). "An immersive virtual environment for collaborative geovisualization," in *2017 9th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)* (Athens).
- Faiola, A., Newlon, C., Pfaff, M., and Smyslova, O. (2013). Correlating the effects of flow and telepresence in virtual worlds: enhancing our understanding of user behavior in game based learning. *Comput. Hum. Behav.* 29, 1113–1121. doi: 10.1016/j.chb.2012.10.003
- Giraudeau, P., and Hachet, M. (2017). "Towards a mixed-reality interface for mind-mapping," in *Proceedings of the Interactive Surfaces and Spaces on ZZZ—ISS 17* (New York, NY), 344–389.
- Greenwald, S., Corning, W., and Maes, P. (2017). "Multi-user framework for collaboration and co-creation in virtual reality," in *Proceedings of the 12th International Conference on Computer Supported Collaborative Learning* (Philadelphia, PA).
- Hart, S. G. (2006). Nasa-task load index (nasa-tlx); 20 years later. *Proc. Hum. Fact. Ergon. Soc. Annu. Meet.* 50, 904–908. doi: 10.1177/154193120605000909
- Krahnstoeber, N., Kettebekov, S., Yeasin, M., and Sharma, R. (2002). "A real-time framework for natural multimodal interaction with large screen displays," in *Proceedings Fourth IEEE International Conference on Multimodal Interfaces* (Pittsburgh, PA).
- Liarokapis, F., and Newman, R. M. (2007). "Design experiences of multimodal mixed reality interfaces," in *Proceedings of the 25th Annual ACM International Conference on Design of Communication—SIGDOC 07* (El Paso, TX), 34–41.
- Liarokapis, F., Petridis, P., Andrews, D., and Freitas, S. D. (2017). "Multimodal serious games technologies for cultural heritage," in *Mixed Reality and Gamification for Cultural Heritage*, eds M. Ioannides, N. Magnenat-Thalmann, and G. Papagiannakis (Cham: Springer International Publishing), 371–392.
- Loomis, J. M., Blascovich, J. J., and Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behav. Res. Methods Instr. Comput.* 31, 557–564. doi: 10.3758/bf03200735
- Louwerse, M. M., Dale, R., Bard, E. G., and Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognit. Sci.* 36, 1404–1426. doi: 10.1111/j.1551-6709.2012.01269.x

- McCloy, R., and Stone, R. (2001). Virtual reality in surgery. *BMJ* 323, 912–915. doi: 10.1136/bmj.323.7318.912
- Miyasugi, M., Akaike, H., Nakayama, Y., and Kakuda, H. (2017). "Implementation and evaluation of multi-user mind map authoring system using virtual reality and hand gestures," in *IEEE Xplore Digital Library* (Nagoya).
- Oviatt, S. (2003). "The human-computer interaction handbook," in *Chapter Multimodal Interfaces*, eds J. A. Jacko, and A. Sears (Hillsdale, NJ: L. Erlbaum Associates Inc.), 286–304.
- Pan, Z., Cheok, A. D., Yang, H., Zhu, J., and Shi, J. (2006). Virtual reality and mixed reality for virtual learning environments. *Comput. Graph.* 30, 20–28. doi: 10.1016/j.cag.2005.10.004
- Papachristos, N. M., Vrellis, I., Natsis, A., and Mikropoulos, T. A. (2013). The role of environment design in an educational multi-user virtual environment. *Br. J. Educ. Technol.* 45, 636–646. doi: 10.1111/bjet.12056
- Redfern, S., and Galway, N. (2002). Collaborative virtual environments to support communication and community in internet-based distance education. *J. Inform. Technol. Educ. Res.* 1, 201–211. doi: 10.28945/356
- Rizzo, A., Parsons, T. D., Lange, B., Kenny, P., Buckwalter, J. G., Rothbaum, B., et al. (2011). Virtual reality goes to war: a brief review of the future of military behavioral healthcare. *J. Clin. Psychol. Med. Settings* 18, 176–187. doi: 10.1007/s10880-011-9247-2
- Sadler, P., Ethier, N., Gunn, G. R., Duong, D., and Woody, E. (2009). Are we on the same wavelength? Interpersonal complementarity as shared cyclical patterns during interactions. *J. Pers. Soc. Psychol.* 97, 1005–1020. doi: 10.1037/a0016232
- Schaffer, S., and Reithinger, N. (2016). "Benefit, design and evaluation of multimodal interaction," in *Proceedings of the 2016 DSLI Workshop. ACM International Conference on Human Factors in Computing Systems (CHI)* (San Jose, CA).
- Sharma, R., Pavlovic, V., and Huang, T. (1998). Toward multimodal human-computer interface. *Proc. IEEE* 86, 853–869. doi: 10.1109/5.664275
- Sharples, M. (2000). The design of personal mobile technologies for lifelong learning. *Comput. Educ.* 34, 177–193. doi: 10.1016/s0360-1315(99)00044-5
- Slater, M. (2003). "A note on presence terminology," in *Presence-Connect* (London).
- Tarr, M. J., and Warren, W. H. (2002). Virtual reality in behavioral neuroscience and beyond. *Nat. Neurosci.* 5, 1089–1092. doi: 10.1038/nn948
- White, M., Petridis, P., Liarokapis, F., and Plecinckx, D. (2007). Multimodal mixed reality interfaces for visualizing digital heritage. *Int. J. Archit. Comput.* 5, 321–337. doi: 10.1260/1478-0771.5.2.322
- Witmer, B. G., Jerome, C. J., and Singer, M. J. (2005). The factor structure of the presence questionnaire. *Presence Teleoper. Virt. Environ.* 14, 298–312. doi: 10.1162/105474605323384654
- Witmer, B. G., and Singer, M. J. (1998). Measuring presence in virtual environments: a presence questionnaire. *Presence Teleoper. Virt. Environ.* 7, 225–240. doi: 10.1162/105474698565686

Conflict of Interest Statement: ZE and JV were employed by company Konica Minolta. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors declare that this study received funding from Konica Minolta. The funder had the following involvement with the study: helped to define testing scenarios and overall design of user study, provided feedback during the development of the application. The funder was not involved in the collection, analysis, interpretation of data, the writing of this article and the decision to submit it for publication.

Copyright © 2019 Kut'ák, Doležal, Kerous, Eichler, Vašek and Liarokapis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.