

RESEARCH

Open Access



# Mechanism to capture learner's interaction in VR-based learning environment: design and application

Rumana Pathan , Ramkumar Rajendran and Sahana Murthy

\* Correspondence:  
rumana\_pathan@iitb.ac.in  
Indian Institute of Technology,  
Bombay, India

## Abstract

Virtual Reality (VR) is a multi-sensory technology that stimulates learning and has the potential for pedagogical applications. While researchers in VR have demonstrated several applications to support understanding and learning in STEM education, the research regarding which features of VR leverage learning is in its infancy. The existing studies exploring how learners interact with VR are based on human observations or learners' perceptions. This paper describes a novel mechanism to capture learner's interaction behavior, in the context of a mobile-based static VR to learn the human circulatory system. The data capturing mechanism is based on screen recordings of VR interaction, which is further annotated manually to form a time-sequenced action series. In a preliminary test conducted with three learners, the interaction data was analyzed based on the time spent in each action in the VR environment, frequently co-occurring actions, and sequence of actions. The test results are described and the implications of using such a mechanism to capture learners' interaction behavior is discussed. We conclude that capturing data in this manner gives a rich and detailed profile of learners and enables use of various analytics methods to provide personalized and adaptive support to learners.

**Keywords:** Virtual reality, Learner's behavior, Log-file, Sequential pattern mining

## Introduction

Virtual Reality (VR) is an emerging technology that has demonstrated the potential to change and improve the way learners are educated in many fields such as medical, engineering, social sciences, etc. (Jensen & Konradsen, 2018). Several applications have been developed for STEM domains in early years of VR in education (Dede, Salzman, & Loftin, 1996a, b) as well as more recent ones (Doak, Denyer, Gerrard, Mackay, & Allison, 2020; Tamaddon & Stiefs, 2017). Interaction with such VR environments has shown gains in constructs related to learning such as conceptual understanding, problem-solving, spatial ability, etc. (Jensen & Konradsen, 2018). By interaction in VR environment, we mean learners interacting with features such as VR menu, 3D image, navigation cues etc.

Researchers in VR have demonstrated several ideas concerning how VR might facilitate learning, but there is comparatively less focus on which of the VR features provides leverage to enhance learning. The existing research studies that explore how various features of VR affect learning has been mostly based on direct observations, learner perceptions, and learning outcomes (see e.g. Roussou, Oliver, & Slater, 2006; Salzman, Dede, Loftin, & Chen, 1999; Tudor, Minocha, Collins, & Tilling, 2018). Studies in VR looking at learner's finer level of interaction data (e.g. eye-gaze) is used mostly to enhance the VR experiences. But such data has not been used to understand learners' interaction in VR, or to understand the relationship between features of VR and learning.

In a few computer-based (non-VR) learning environments, log-files generated using learner's interaction with the environment are used to analyze learner's behavior to facilitate adaptation and personalization of the learning environment (Basu, Biswas, & Kinnebrew, 2017; Pathan, Shaikh, & Rajendran, 2019). A similar methodology to understand how VR features facilitate learning can be implemented by using log-files. However, existing VR applications does not generate log data similar to computer-based learning environment.

This paper describes a novel mechanism to capture learner interaction with VR using screen recording to reveal different aspects of learner interactions in VR environment. The paper also discusses the feasibility of such a mechanism through a preliminary test example with three learners to demonstrate how it can be used to inform researchers as well practitioners for using VR for learning and instruction. Based on the data collected from the three learners, the test identified the frequent actions, the time distribution of each action in the VR environment, and frequent sequence of actions to reveal learner interactions in VR environment in an objective and accurate way.

In the sections that follow, Section 2 reports a brief literature review of the existing VR studies and the data collection methods involved in several VR and non-VR (computer-based) studies. Section 3 describes the data capturing mechanism. The results of analysis performed from three learner's data is described in the section 4. In section 5, the results of implementing and testing the mechanism to collect data in a VR environment is described and this paper is concluded in Section 6.

### **Literature review**

Several VR projects have been developed and employed in educational setups to evaluate the potentials of VR. For example, VR is used in teaching and learning of several STEM concepts in projects such as Newton's World and Maxwell World (Dede et al., 1996a, b), Construct 3D (Kaufmann, Schmalstieg, & Wagner, 2000), NICE environment (Roussou et al., 2006), the Round Earth project (Johnson, Moher, Ohlsson, & Gillingham, 1999), Peppy (Doak et al., 2020), and the ISS virtual tour (Tamaddon & Stiefs, 2017). Newton's World and Maxwell World (Dede et al., 1996a, b), is used to explore physics concepts such as kinematics. In Construct3D, learners worked in a 3D space to solve complex spatial problems in Geometry (Kaufmann et al., 2000). Similarly, NICE project is an interactive VR environment that provides young children to collaborate in a fantasy world to cultivate a virtual garden (Roussou et al., 2006). Peppy, however, is a virtual environment to explore principles of polypeptide structure for undergraduate biochemistry class (Doak et al. 2020). ISS virtual tour, on the other hand, provides an

interactive virtual journey near ISS, where learners can also experience microgravity (Tamaddon & Stiefs, 2017). Research studies using these VR learning environments report that features of VR such as multiple representations, the 3-dimensional space, and multisensory cues (such as visual, auditory, and haptic) might have enhanced skills like problem-solving, understanding the concepts, spatial ability, and other psychomotor skills. However, the widespread use of such VR technologies at real educational settings have been a bottleneck because of the high cost associated with the required hardware and software (Olmos, Cavalcanti, Soler, Contero, & Alcañiz, 2018). Also, it limits the use of such technology in a lab set up.

With the advent of mobile VR (mVR), configuring a VR environment using a headset based on smartphones became more accessible and affordable. There have been several research studies conducted in large classrooms using mVR (Tudor et al., 2018; Vishwanath, Kam, & Kumar, 2017). A lot of educational content has been created (e.g. Google Expedition) to be used in mVR. Chittaro, Corbett, McLean, and Zangrando (2018) have developed an aviation safety mVR to educate passengers about flight safety engagingly and comprehensively. Similarly, Tudor et al. (2018) and Craddock (2018) have used Google expedition application to teach students concepts on environmental impact and exploring organelles in cell biology, respectively. Studies in these projects have reported engagement and learning with the help of mVR features such as screen interaction, the three-dimensional view, and navigation cues.

Although existing research studies have discussed how VR might facilitate learning, the field has little information concerning which VR features provide leverage for enhancing learning. Hence more recently, a few research studies have started exploring the relationship between VR features and learning (see Roussou et al., 2006; Salzman et al., 1999; Tudor et al., 2018). For example, in the ScienceSpace project, Salzman et al. (1999) studied the relationship between VR features such as 3-D immersion, the frame of reference, and multisensory cues that influence learning. The data from human observation, learner's usability questionnaires, and interview feedback suggests that 3-D immersive representations motivated learners to perform better than 2D representations. In a similar study Tudor et al. (2018) explained how features such as 3-D immersion, navigation, and emphasis to highlight aspects of a scene fostered learners to become aware of environmental challenges. The inferences made in this research relied on learners' post-intervention reflections in written format and group-interviews conducted by the educators. To summarize, the studies mentioned above have explored the effect of various VR features on learning based on either human observations or learner perceptions. However, finer level of learner's interaction data such as eye-gaze, head orientation, or actions performed in the environments are not captured or utilized in the above research studies.

In a few research studies, learner's interaction data have been considered to improve aspects such as compression algorithms required to render images in high-end VR (powered by external computers or game consoles), create models of visual saliency, etc. For example, Sitzmann et al. (2018) used head and gaze trajectories to find similarities in viewing behavior of different learners in VR. Similarly, other studies have utilized such data to enhance visual fidelity (Marmitt & Duchowski, 2002) and human-computer interaction (Ruhland et al., 2015). In another instance, Pillai, Ismail, and Charles (2017) analyzed screen recordings of VR interaction (specifically gaze-pointer

interactions) to understand and design guidelines for the development of visual cues in an mVR. But, the exploration of such interaction data has not been used to discuss how various features of VR leverage learning. Moreover, the gaze and head trajectory data are an expensive way of capturing the interaction and involves the use of additional hardware to capture the learner data. Based on our knowledge, there exists no application to collect learner's interaction in mVR. Hence our mechanism to capture VR interaction using screen-based recordings is both novel and innovative.

In a few computer-based (non-VR) learning environments, a comparatively inexpensive way to record learner actions along with relevant contextual information has been implemented. This interaction can be viewed via log-files generated using the learner's clickstream or screen-recordings. For example, systems such as Betty's brain (Leela-wong & Biswas, 2008), Metatutor (Azevedo, Johnson, Chauncey, & Burkett, 2010), MEttLE (Pathan et al., 2019) record all actions that learners perform in the learning environment. Action logs of learners such as reading resources, interacting with an agent, answering questions are captured along with contextual information such as time-stamp. This list of actions, along with their labels, facilitates the application of various analytic and mining methods like pattern mining, process mining and clustering. Implementation of such algorithms allows one to gain more insights regarding learner behavior and learning with the features of the system and facilitates adaptation and personalization by adapting to learners' needs as they interact with the system (Basu et al., 2017; Munshi et al., 2018). A similar methodology can be explored to understand how VR features facilitate learning.

The need for a mechanism to capture detailed learner interaction in VR is evident from the existing research studies. Moreover, the existing data capturing mechanisms are expensive, hardware intensive, and cannot be used on lighter platforms such as mVR. Inspired by a few computer-based environments, a screen-recording based data capturing mechanism for mVR can be implemented to capture primary mode of interactions include eye and head movement along with some click-based screen interaction. The screen-recording that captures the interactions in mVR can be further annotated to signify various activities of the learner. Also, using the action and context sequence generated log files, various features of VR, learner behavior and other interesting details can be explored. The paper describes one such mechanism to capture learner interaction in the context of a mVR application.

### **Mechanism to capture learner interaction in VR: implementation**

An exploratory test with three learners is conducted to test the mechanism to capture data in mVR. To describe use of the data capturing mechanism, a preliminary test (pilot study) with three learners is conducted as a proof of concept. Hence, the specific goal of this preliminary test is to explore how learners interact with different features of VR environment and how it can be captured. Through an analysis of learner interaction with VR features, it is aimed to gather insights and further investigate several questions about the behavior of learners exploring the virtual environment. This paper analyzes learners' characteristics at three levels, time spent in each action, frequently co-occurring actions, and when an action occurred.

The learners in this test are 8th grade students (2 females, one male) from a school located in the suburban area of Mumbai with English as the primary medium of

instruction. All the three students were purposively sampled because a) they were formally introduced to the topic of the human circulatory system in their schools, b) belonged to a sub-urban school, and c) had interacted with high-end VR before (for gaming purpose), but were new to mobile VR. The students and their parents signed an informed consent before the participation in test.

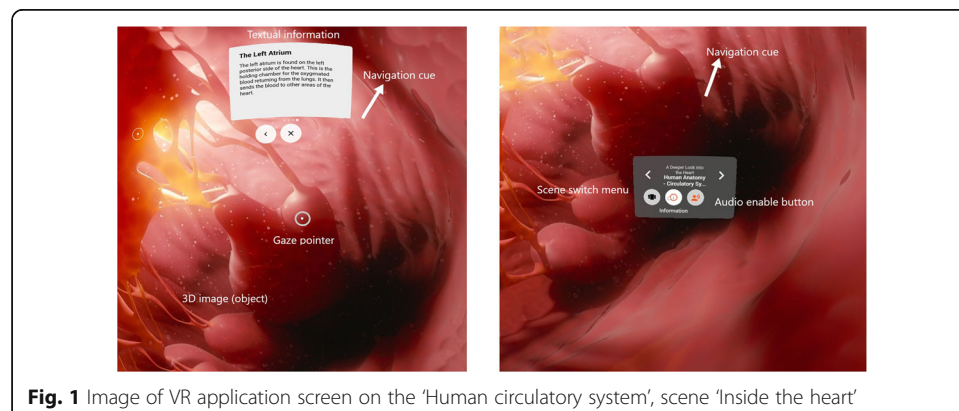
### Mobile-VR learning environment: human circulatory system

The mVR application on ‘human circulatory system’ used in this project is a part of Google expeditions application available on the google play store. This application is developed by VidaSystems and is available for the public. It has eight scenes in total arranged hierarchically. The first two scenes are introductory and define the circulatory system. The remaining six scenes dwell deeper into details such as ‘structure of heart’, ‘blood vessels’, etc. and explain how blood circulation takes place in a human body using 3D images and relevant text as shown in Fig. 1. A gaze pointer (a small white dot) that represents where the learner is currently looking can be seen in the Fig. 1. There are navigation cues in the form of arrows which leads the learner from one content to another in a hierarchical manner. The learner may choose not to follow the navigation cue and instead navigate or move on his own. The learner can switch between scenes by clicking on a dialogue box (VR-instruction) and choose the scene of his choice. The same dialogue box also enables audio, but the functionality was switched off for the test mentioned in this paper.

The virtual trip facilitated by Google Expeditions can be experienced by placing a smartphone in a phone holder (e.g. Google Cardboard), which can be mounted on eyes using straps. Learners can then look at their device through eye holes that give a VR experience.

### Data collection design and procedure

We collected the learners’ interaction data in a lab set-up. The set up consisted of an android mobile with Google expedition application installed in it, which was used to run the virtual reality program on the human circulatory system. A VR headset was used to convert the mobile into a VR gateway, and the learners were provided with writing material (A4 sheets, pencil, eraser). There were 2 video recordings, 1) mobile screen where VR content was displayed to the learners, and 2) learners overall



interaction video. A screen recording application was used to record the screen of the mobile, while the learner was interacting with the VR program.

Researchers helped the learners to resume the application if terminated accidentally, adjust VR headset, and to remind them to take breaks in between. None of the learners sought help from researchers regarding learning content/exercise. The learners could move freely and were provided with a 360 degrees swivel chair. They were also strictly instructed to remove the headset even if slight discomfort was felt. After every 10 min of continuous VR interaction, learners were reminded to take a break and then resume.

A set of questions (pretest) were asked at the beginning to determine the learners' prior knowledge of the human circulatory system. The learners were then introduced to the VR application (i.e. 'the ocean' in VR) to familiarize them with the different controls. Once the learners were comfortable with the application, the VR program on "human circulatory system" was introduced to them. Learners constructed a concept map (of the structure of the heart) during VR interaction. Altogether eight scenes were studied by the learners, which provided the view of the human circulatory system at different levels. Learners answered a set of questions (post-test) at the end, which is similar to the pretest questions.

#### **Proposed mechanism to capture learner's interaction in mVR**

To collect learner's interaction with the VR environment we collected data from the following resources:

- Screen recording: The interaction with the VR application was recorded using A-Z<sup>1</sup> screen recording software. The video was later analyzed and coded manually according to the actions performed by the learners.
- Video and audio recordings: The test was video recorded to capture the interaction of the learner with the VR headset and the writing material used. The video recordings and the screen capture were time synchronized to obtain an overall view of what the learner was doing.
- The concept map drawn by learners during interaction with VR
- Response to pre-test, post-test and interview questionnaire

Since our goal is to understand how learners interact with the various features of VR, the pre and post test results were not analyzed in this paper.

To interpret how learners are interacting with mVR, we listed all the actions a learner can do in the VR environment. With the help of screen recordings of learners interacting with VR and position of gaze pointer, we found that learners performed different actions such as looking at the 3D images, reading the text relevant to the 3D images, navigating in the 3D space either with the help of navigation cues or on their own, changing from one scene to another, and controlling the application such as taking pauses. With reference to the actions performed by learners, an initial table to record the action and its contextual information was developed by the researchers (Table 1).

Table 1 summarizes all the actions along with details such as the action duration, and its description. For example, action READ means that the learner is reading the

<sup>1</sup><https://play.google.com/store/apps/details?id=com.hecorat.screenrecorder.free&hl=en%20IN>

**Table 1** Learner actions and contextual details

Action	Time in sec	Description	View
READ-long	READ > 10	The learner is reading textual information in a scene	global_scene = "
READ	3 < READ <= 10		local_scene = "
READ-short	READ <= 3		scroll = Y/N
LOOK-long	LOOK > 10	The learner is looking at the 3D images in a scene.	global_scene = "
LOOK	3 < LOOK <= 10		local_scene = "
LOOK-short	LOOK <= 3		3D object = "
FN-long	FN > 10	The learner is following a system-generated navigation cue, which is in the form of an arrow.	from = "
FN	3 < FN <= 10		to = "
FN-short	FN <= 3		-
MOVE-long	MOVE > 10	The learner is navigating inside the VR scene without the help of navigation cue.	from = "
MOVE	3 < MOVE <=10		to = "
MOVE-short	MOVE <= 3		-
SC-seq	-	The learner is changing the scene in VR. The next scene is selected in a sequential manner.	from = " to = "
SC-ran	-	The learner is changing the scene in VR. The next scene is selected in a sequential manner.	from = " to = "
VR-i	-	The learner is interacting with a dialogue box, which allows one to either change a scene or click on the information.	select_scene = Y/N select_info = Y/N
CON-app	-	The learner is controlling the application by either starting it, pausing it or exiting it.	start_app = Y/N pause_app = Y/N exit_app = Y/N

textual information present in the VR application. The position of gaze pointer on the textual information determines the time spent by the learner reading. The action READ if less than 3 s is termed as READ-short, and if it is more than 10 s it is termed as READ-long. The values used to classify the actions as READ-short or READ-long is based on the time required to read the content provided in the VR environment that we selected. Based on several attempts and analysis, we found that, learner can decide whether to continue to read the page or skip within 10 s. It is to be noted that, the threshold chosen to classify the actions is not generalizable. A similar classifying technique has been used by Rajendran et al. (2018) in the context of a computer-based learning environment. The 'view' section of the table gives us more information about the context, for e.g. 'global scene' highlights the current scene (e.g. Structure of the heart), and 'local scene' the 3D image (e.g. Myocardium). Whether the learner is scrolling the text is also captured in this case. Similarly, there are other actions such as LOOK, FN, MOVE, SC-seq, SC-ran, VR-i, and CON-app described in Table 1 in detail.

Using Table 1, two researchers independently coded a common 5-min video. The unit of analysis was decided as 'change in action'. Inter-rater reliability was established with Cohen's Kappa = 0.75. Following which all the videos were coded using the same coding scheme (Table 1) and a time-sequenced action series (log data files) containing learner id, action, its context, and the time for all the learners were generated.

**Excerpt from a sample log data file generated using the data capturing mechanism**

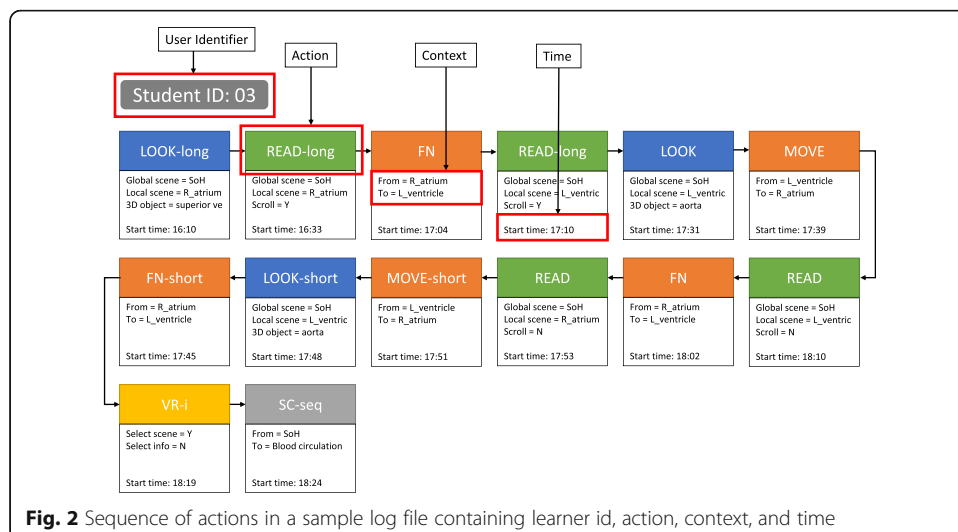
Figure 2 is an excerpt from log data of learner\_03, exploring the scene ‘Structure of Heart (SoH)’ in mVR. The log file contains the learner identifier, the action, its context, and the time. In the figure, the learner at 16:10 can be seen exploring the right atrium and is looking (LOOK-long) at superior vena cava for 23 s (i.e. more than 10 s). The time duration is calculated by subtracting start time of current action from the start time of the next action. At 16:33, the learner reads (READ-long) about the textual content associated with it for 31 s and can also be seen scrolling the content. Further, at 17:04, the learner moves (FN) from the right atrium to left ventricle using a navigation cue (FN). The learner can be seen reading (READ-long) textual content associated with the left ventricle, followed by looking (LOOK) at the aorta in the left ventricle for 8 s (i.e. more than 3 s, but less than 10 s). The learner then moves (MOVE) from the left ventricle to the right atrium on his own, i.e. without the help of a navigation cue. In the next few action logs, the learner can be seen navigating and reading/looking at the content related to the aorta and superior vena cava. At the end of the scene’s exploration, the learner chooses to select a scene (VR-i) and selects the next scene in sequence (SC-seq), i.e. Blood circulation.

**Analysis of learner’s interaction data**

The action sequence series generated was analyzed at three levels, i.e. 1) the time spent by the three learners in each action was calculated, 2) frequent patterns of actions among the learners were mined using sequential pattern mining algorithm, and 3) the contexts in which actions occurred are discussed in detail.

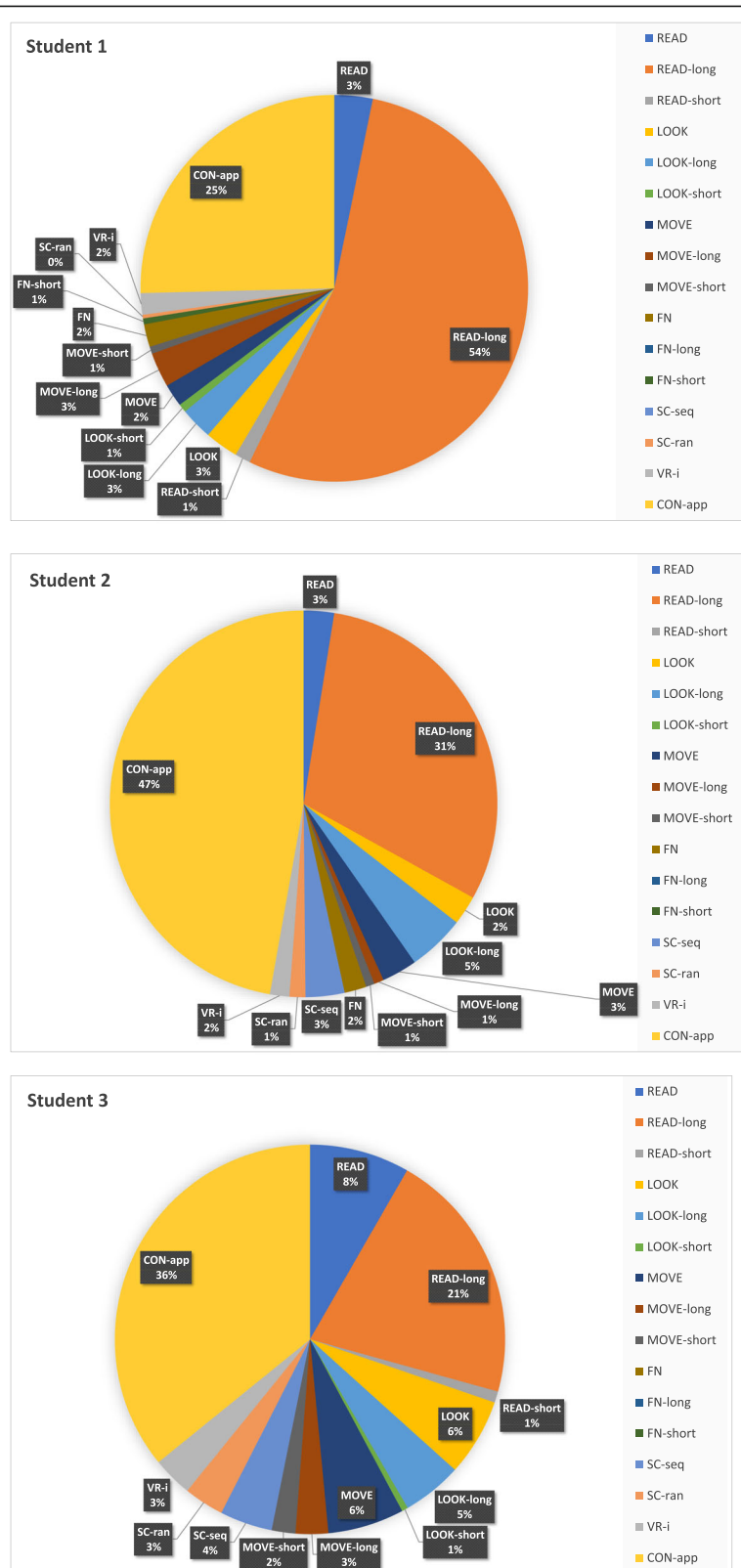
**Time distribution of each action**

The time spent by all the three learners in each action (e.g. READ, LOOK, etc.) is depicted in the pie charts shown below. Figure 3a, b, and c depict the difference in behavior between the three learners. For example, learner 1 is spending most of her time doing read action as compared to learner 2 and 3. Learner 2 (refer Fig. 3b), on the



**Fig. 2** Sequence of actions in a sample log file containing learner id, action, context, and time





**Fig. 3 a** Time spent in various actions by learner 1 in VR. **b** Time spent in various actions by learner 2 in VR. **c** Time spent in various actions by learner 3 in VR

other hand, is seen pausing the application, i.e. CON-app for a longer duration as compared to learner 1 and 3.

Figure 4 summarizes the total time (in mins) spent by each learner in each scene. All the three learners can be seen spending maximum time in exploring scene 3, which was regarding the structure of the heart and included concepts such as the heart's exterior, cardiac muscle, within the heart's chambers, and the network hub.

### Patterns found in actions

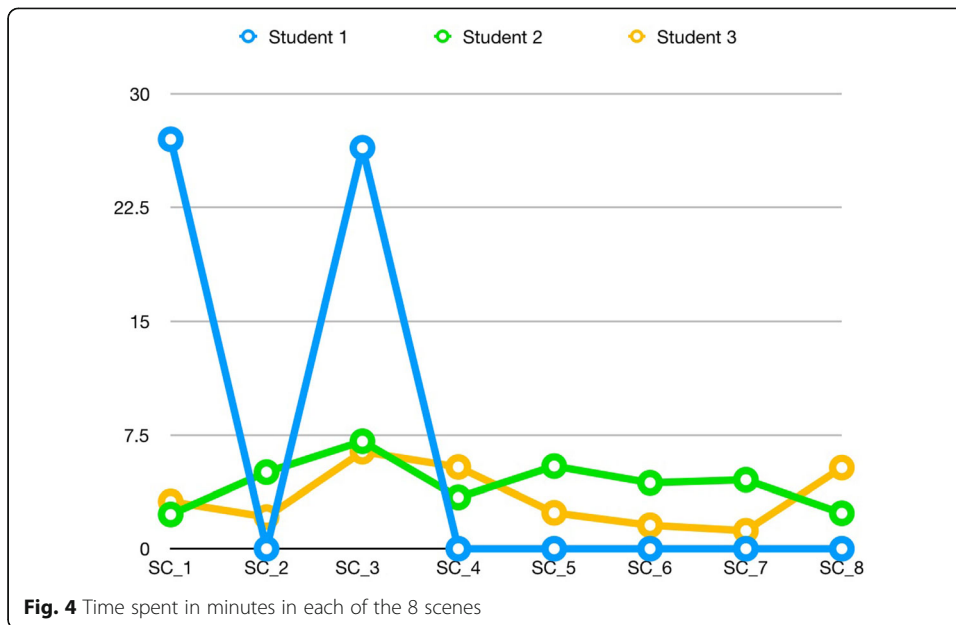
To understand the frequent patterns of actions, the time-sequenced action series was processed using a sequential pattern mining (SPM) algorithm. SPM is the mining of frequently occurring ordered events as patterns (Agrawal & Srikant, 1995; Kinnebrew, Loretz, & Biswas, 2013). The algorithm was implemented with minimum threshold of 0.5, i.e. minimum 50% of the total learners must have the same pattern in common, also called as minimum support. When mining frequent patterns in learning interactions (VR interaction for e.g. READ -> LOOK -> MOVE), students may perform additional actions that are also interspersed with actions that constitute the pattern. Therefore, a maximum gap constraint was applied, i.e. between each consecutive pair of actions in a given pattern, the algorithm allowed up to 'gap number' additional actions. The SPM algorithm was implemented using LASAT (Learning Activity Sequence Analysis Tool) with support 0.5, and maximum gap 1 (Mishra, Munshi, Rushdy, & Biswas, 2019). Altogether 965 patterns were generated, which were then filtered. The filtering criterion included removing patterns which were one action long, and patterns whose average frequency was equal to or less than one. The filtering criteria reduced the patterns to 173.

The pattern most commonly found was FN -> READ-long, which indicated that a learner followed a navigation cue (for 4–10 s) and read content (for more than 10 s). This pattern was observed 20 times in total in all three learners. The next frequent pattern was LOOK -> READ-long, which was found 19 times in all the participants. This pattern signifies that learners were looking at an image (for 4–10 s) and were then reading content associated with the image for more than 10 s. Actions such as SC-seq/ran (change scene in VR either sequentially or randomly), VR-i (VR-instruction that enables scene change or information) were used less frequently as compared to other actions by the learners.

### When an action happened - the context

In this subsection, we analyze when the actions occurred during the learner's interaction with the environment. With the help of the concept map developed by the learners, it was observed that, learners who spent most time reading were able to recollect the structure of the circulatory system with proper biological terminology. The learners who spent most time at looking at the images were able to describe the structure of human circulatory system while capturing more visual details, such as blockages found inside a blood vessel, capillaries in their arteries, etc.

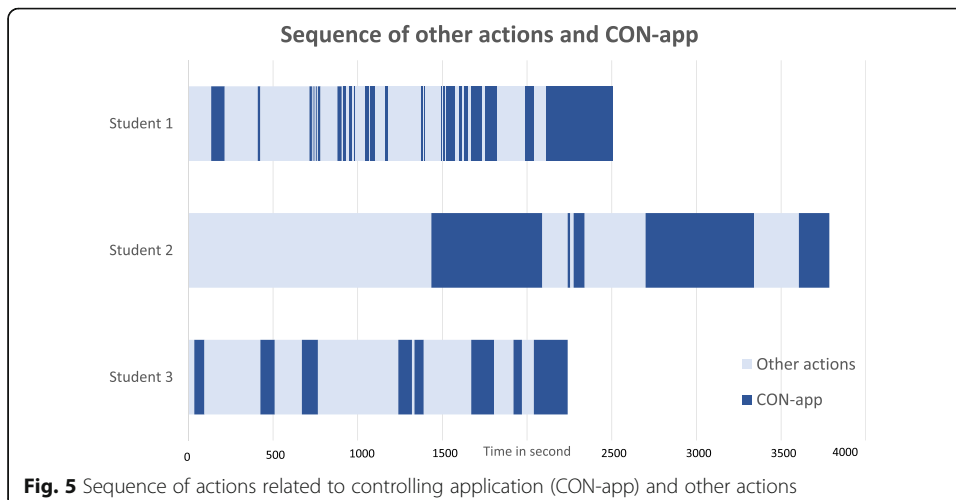
To analyze learner's CON-app behavior, sequence of 'CON-app' action and 'other actions' were plotted. We observed that (refer Fig. 5) learner 1 paused 36 number of

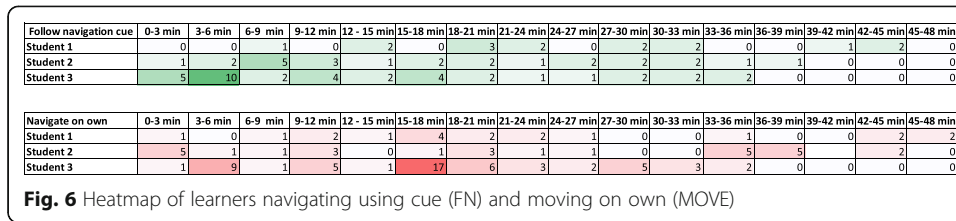


times, whereas learner 2 and 3 can be seen pausing the application 5 and 8 number of times, respectively.

Similarly, we looked at the navigation actions, FN and MOVE. Figure 6 is a heat map generated based on the use of FN and MOVE by all the three learners. The frequency of 'FN' and 'MOVE' for every 3 min for all learners is recorded and color-coded. More number of FN and MOVE actions has been highlighted in a darker shade of green and red color respectively, and lesser use in their lighter shades. Learner 2 and 3 can be seen using FN and MOVE more as compared to learner 1.

In the patterns containing MOVE and FN, learners were found using both the ways of navigation almost equal number of times. It was also observed that in a new scene, most of the learners always used a navigation cue to look at the VR space before moving inside VR on his own. This result of the analysis is in line with the results of a prior study which





had stated that navigation in VR is difficult and introduction of specific navigation cues is helpful for learners to move around inside a VR space (Vinson, 1999).

### Discussion

The data capturing mechanism mentioned in this paper is currently implemented in the context of mVR. This mechanism can be applied to any existing VR application. However, In order to implement the mechanism in another VR environment, the following steps will have to be ensured, 1) identify various VR features/affordances in the new VR environment (check for additional feature e.g. ‘VR audio’), 2) for each individual learner, capture the time of action along with, 3) contextual information such as the global and local scenes, 4) capture other details of the interaction such as scrolling, clicking etc.

The data resulting from the mechanism mentioned in the paper will give deeper insights on how learners interact with VR features. The mechanism is equipped to capture the learner identifier, the action (learner use of VR features such as ‘LOOK’ at 3D images), its context, the time at which the action took place and its duration. The time sequenced action series generated using such data enables us to capture and observe how different learners interact with various VR features. For example, in the preliminary test mentioned above, the patterns containing FN and MOVE actions inform us about the different navigation behaviors found among the learners. Similarly, in the study it was observed that learners spent more time reading the textual content as compared to looking at the 3D images. This work can be extended with larger N to further understand the correlation between different navigation behaviors and its corresponding effect on learning in VR. Similarly, image viewing, and reading behaviors can be analyzed. Moreover, with larger N, effective interaction paths can be obtained for different kinds of learners, and an informed decision can be made regarding re-design of the existing VR application that would aid students to achieve their learning goal.

As described earlier, the log files generated in non-VR (such as computer-based learning environments) systems are used to provide personalized and adaptive feedback to improve learner learning (Basu et al., 2017). Our work forms the basis for a similar automatic mechanism in VR learning environments. Capturing data in this manner gives a very rich and detailed profile of the learners’ interaction with the system, which is helpful in many ways. For example, the data captured in VR environment similar to log data allows usage and implementation of a number of learning analytics methods (such as prediction, pattern mining, etc.). For instance, the study described in this paper implements a SPM algorithm to mine common patterns found in learners while interacting with VR. Extending a similar analysis with larger N will enable us to identify various desired patterns to effectively learn in VR, which can be used to provide scaffolding to novice learners.

Recently data such as head and eye gaze movement has been captured in VR to improve the VR viewing experience. However, such data has not been used to study the impact of different VR features. Moreover, capturing data such as head and eye gaze employs use of high cost equipment such as eye-trackers. Coding mechanisms such as the one mentioned in this paper uses screen-recording of VR to annotate several learner actions and thus is a cost-effective way to capture the learner interaction. Although comparatively cheaper, the mechanism has a few limitations; for example, a) the methodology involved in coding is manual and time-intensive, and b) during implementation, the mechanism is prone to human errors such as interpreting change in actions. Because of which, analysis with large N becomes difficult. However, an algorithm to automatically identify learner interaction can also be developed using computer vision or video processing software to conduct studies with larger N. Similarly, with the inclusion of data from sensors such as gyroscope, accelerometer etc. information such as device's acceleration, vibration, tilt, orientation details can be captured precisely.

## Conclusion

This paper describes a novel data capturing mechanism in mVR applications. To describe use of the data capturing mechanism, a preliminary test with three learners was conducted. The aim of the test was broadly to understand how learners interact with VR. To do so, learner actions in VR such as reading, looking, following a navigation cue, etc. were captured by manually coding the screen recordings of learners' interaction. With the help of the time-sequenced action series, learners' characteristics such as time spent on each action, frequently co-occurring actions and when an action occurred were analyzed.

We have described the data logging mechanism for mobile-based VR application. Since the mechanism involves manually coding screen recordings of VR interaction, this strategy can be also applied to any existing VR. Similarly, a computer-generated logging mechanism will find its usage in VR applications to analyze learner behavior and to facilitate adaptive and personalized learning environment in VR at a large scale.

The preliminary test described in this paper is to show the data analysis in a VR environment, hence the smaller sample size. A similar study can be extended with a larger sample size to obtain process models which will provide more precise details regarding how learners are learning in a system. With an increased N, other analytics can be applied to the data to gain more insights. However, the process of coding the VR screen recording to obtain the finer level of interaction data was manual and time-intensive and is not recommended for a larger sample. Also, the VR application used in this test is mobile based. In contrast, there are many high-end VR applications available where learners might choose to navigate or learn differently due to additional features available in them (e.g. haptic feedback, joystick/hand gears to control VR environment etc.). In the future, we propose to conduct similar research studies with larger N and high-end VR applications. Based on the research studies, we also aim to explore how various VR features impact learning in future.

## Abbreviations

VR: Virtual Reality; mVR: Mobile Virtual Reality; 3D: Three dimensional; SPM: Sequential pattern mining; FN: Follow navigation-cues; SC-seq: Scene change sequential; SC-ran: Scene change random; VR-i: Virtual Reality instruction; CON-app: Control application; SoH: Structure of heart

### Acknowledgements

The authors would like to thank the participants and their parents for their participation in the study. The authors acknowledge Research Scholars of Indian Institute of Technology (IIT) Bombay and Next Education Research Lab for the support in conducting this research.

### Authors' contributions

RP contributed in design of the work, data collection, data analysis, drafting the work, and substantively revising it. RR contributed in design of the work, data analysis, drafting and substantively revising the manuscript. SM contributed in design of the work, data analysis, and substantively revising the manuscript. The author(s) read and approved the final manuscript.

### Authors' information

Interdisciplinary Programme in Educational Technology (IDP-ET) at Indian Institute of Technology (IIT) Bombay, Mumbai, India Rumana Pathan, Ramkumar Rajendran & Sahana Murthy.

### Funding

Next Education Research Lab at IDP-Educational Technology, IIT Bombay

### Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available because of minor aged (under 14 years) participants involved in the study.

### Competing interests

The authors declare that they have no competing interests.

Received: 31 July 2020 Accepted: 6 October 2020

Published online: 26 November 2020

### References

- Agrawal, R., & Srikant, R. (1995). Mining sequential patterns. In *Proceedings of the eleventh international conference on data engineering*, (pp. 3–14). Piscataway: IEEE.
- Azevedo, R., Johnson, A., Chauncey, A., & Burkett, C. (2010). Self-regulated learning with MetaTutor: advancing the science of learning with MetaCognitive tools. In *New science of learning*, (pp. 225–247). New York: Springer.
- Basu, S., Biswas, G., & Kinnebrew, J. S. (2017). Learner modeling for adaptive scaffolding in a computational thinking-based science learning environment. *User Modeling and User-Adapted Interaction*, 27(1), 5–53.
- Chittaro, L., Corbett, C. L., McLean, G. A., & Zangrando, N. (2018). Safety knowledge transfer through mobile virtual reality: a study of aviation life preserver donning. *Safety Science*, 102, 159–168.
- Craddock, I. M. (2018). Immersive virtual reality, Google expeditions, and English language learning. *Library Technology Reports*, 54(4), 7–9.
- Dede, C., Salzman, M. C., & Loftin, R. B. (1996a). ScienceSpace: virtual realities for learning complex and abstract scientific concepts. In *Proceedings of the IEEE 1996 virtual reality annual international symposium*, (pp. 246–252). Piscataway: IEEE.
- Dede, C., Salzman, M. C., & Loftin, R. B. (1996b). MaxwellWorld: learning complex scientific concepts via immersion in virtual reality. In *Proceedings of the 1996 international conference on learning sciences*, (pp. 22–29).
- Doak, D. G., Denyer, G. S., Gerrard, J. A., Mackay, J. P., & Allison, J. R. (2020). Peppy: a virtual reality environment for exploring the principles of polypeptide structure. *Protein Science*, 29(1), 157–168.
- Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, 23(4), 1515–1529.
- Johnson, A., Moher, T., Ohlsson, S., & Gillingham, M. (1999). The round earth project-collaborative VR for conceptual learning. *IEEE Computer Graphics and Applications*, 19(6), 60–69.
- Kaufmann, H., Schmalstieg, D., & Wagner, M. (2000). Construct3D: a virtual reality application for mathematics and geometry education. *Education and Information Technologies*, 5(4), 263–276.
- Kinnebrew, J. S., Loretz, K. M., & Biswas, G. (2013). A contextualized, differential sequence mining method to derive students' learning behavior patterns. *Journal of Educational Data Mining*, 5(1), 190–219.
- Leelawong, K., & Biswas, G. (2008). Designing learning by teaching agents: The Betty's brain system. *International Journal of Artificial Intelligence in Education*, 18(3), 181–208.
- Marmitt, G., & Duchowski, A. T. (2002). *Modeling visual attention in VR: measuring the accuracy of predicted scanpaths* (Doctoral dissertation, Clemson University).
- Mishra, S., Munshi, A., Rushdy, M., & Biswas, G. (2019). LASAT: learning activity sequence analysis tool. In *Technology-enhanced & evidence-based education & learning (TEEL) workshop at the 9th international learning analytics and knowledge (LAK) conference, Tempe, Arizona, USA*.
- Munshi, A., Rajendran, R., Ocumpaugh, J., Biswas, G., Baker, R. S., & Paquette, L. (2018). Modeling learners' cognitive and affective states to scaffold SRL in open-ended learning environments. In *Proceedings of the 26th conference on user modeling, adaptation and personalization*, (pp. 131–138).
- Olmos, E., Cavalcanti, J. F., Soler, J. L., Contero, M., & Alcañiz, M. (2018). Mobile virtual reality: a promising technology to change the way we learn and teach. In *Mobile and ubiquitous learning*, (pp. 95–106). Singapore: Springer.
- Pathan, R., Shaikh, U., & Rajendran, R. (2019). Capturing learner interaction in computer-based learning environment: design and application. In *2019 IEEE Tenth International Conference on Technology for Education (T4E)*, (pp. 146–153). Piscataway: IEEE.
- Pillai, J. S., Ismail, A., & Charles, H. P. (2017). Grammar of VR storytelling: visual cues. In *Proceedings of the virtual reality international conference-laval virtual 2017*, (pp. 1–4).

- Rajendran, R., Munshi, A., Emara, M., & Biswas, G. (2018). A temporal model of learner behaviors in OELEs using process mining. In *Proceedings of ICCE* (pp. 276–285).
- Roussou, M., Oliver, M., & Slater, M. (2006). The virtual playground: an educational virtual reality environment for evaluating interactivity and conceptual learning. *Virtual Reality*, *10*(3-4), 227–240.
- Ruhland, K., Peters, C. E., Andrist, S., Badler, J. B., Badler, N. I., Gleicher, M., ... McDonnell, R. (2015). A review of eye gaze in virtual agents, social robotics and hci: Behaviour generation, user interaction and perception. *Computer Graphics Forum*, *34*(6), 299–326.
- Salzman, M. C., Dede, C., Loftin, R. B., & Chen, J. (1999). A model for understanding how virtual reality aids complex conceptual learning. *Presence Teleoperators and Virtual Environments*, *8*(3), 293–316.
- Sitzmann, V., Serrano, A., Pavel, A., Agrawala, M., Gutierrez, D., Masia, B., & Wetzstein, G. (2018). Saliency in VR: how do people explore virtual environments? *IEEE Transactions on Visualization and Computer Graphics*, *24*(4), 1633–1642.
- Tamaddon, K., & Stiefs, D. (2017). Embodied experiment of levitation in microgravity in a simulated virtual reality environment for science learning. In *2017 IEEE virtual reality workshop on K-12 embodied learning through virtual & augmented reality (KELVAR)*, (pp. 1–5). Piscataway: IEEE.
- Tudor, A. D., Minocha, S., Collins, M., & Tilling, S. (2018). Mobile virtual reality for environmental education. *Journal of Virtual Studies*, *9*(2), 25–36.
- Vinson, N. G. (1999). Design guidelines for landmarks to support navigation in virtual environments. In *Proceedings of the SIGC HI conference on human factors in computing systems*, (pp. 278–285).
- Vishwanath, A., Kam, M., & Kumar, N. (2017). Examining low-cost virtual reality for learning in low-resource environments. In *Proceedings of the 2017 conference on designing interactive systems*, (pp. 1277–1281).

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)

---