Annotation vs. Virtual Tutor: Comparative Analysis on the Effectiveness of Visual Instructions in Immersive Virtual Reality

Hyeopwoo Lee, Hyejin Kim, Diego Vilela Monteiro, Youngnoh Goh,
Daseong Han, Hai-Ning Liang, Hyun Seung Yang, and Jinki Jung

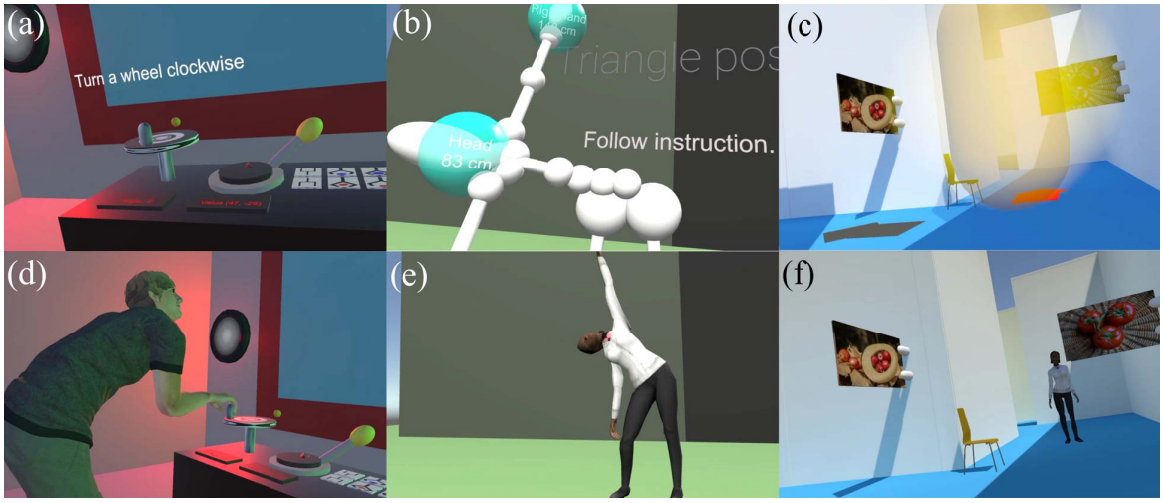

Fig. 1: Screenshots of the use of visual instructions, i.e., annotation (ANN) and tutor (TUT), in crane manipulation (CM), stretching exercises (SE), and maze escape (ME). (a) ANN-CM, (b) ANN-SE, (c) ANN-ME, (d) TUT-CM, (e) TUT-SE, (f) TUT-ME

Abstract— In this paper we present a comparative study of visual instructions in Immersive Virtual Reality (IVR), i.e., annotation (ANN) that employs 3D texts and objects for instructions and virtual tutor (TUT) that demonstrates a task with a 3D character. The comparison is based on three tasks, maze escape (ME), stretching exercise (SE), and crane manipulation (CM), defined by the types of a unit instruction. We conducted an automated evaluation of user's memory recall performances (recall time, accuracy, and error) by mapping a sequence of user's behaviors and events as a string. Results revealed that ANN group showed significantly more accurate performance (1.3 times) in ME and time performance (1.64 times) in SE than TUT group, while no statistical main difference was found in CM. Interestingly, although ANN showed statistically shorter execution time, the recalling time pattern of TUT group showed a steep convergence after initial trial. The results can be used in the field in terms of informing designers of IVR on what types of visual instruction are best for different task purpose.

Index Terms—Virtual reality, Evaluation, Visual guidance, Computer-aided instruction, Human-computer interaction.

1 Introduction

The reality-virtuality spectrum indicates that virtual reality (VR) and augmented reality (AR) transfer information to the user by synthesizing the virtual contents in the virtual world and the real world, respectively [29]. The reason why VR and AR are useful for educational and instructional purposes is because the immersion driven by the high fidelity (VR) or the use of real environment (AR) improves the user's learning effect based on context-dependent memory [4, 40]. For instance, immersive VR (IVR) is a highly interactive, fully immersive, multi-sensory VR that even further amplifies the immersive experience [37]. Excluding the aspects of immersion, virtually generated contents deliver information through visual [27, 35], auditory [30], and haptic [12] sensations depending on the application. Most efforts have been focused on visual aspects (from which humans obtain most of the information). The advantages of IVR are amplified when it is applied to procedural learning/training requiring gesture and providing corresponding feedback [8, 14, 22]. There are, however, very few studies on the format of the interface for making IVR more informative in such procedural learning. Here, the questions can be addressed: **What kind of visual instruction efficiently conveys information? Does the efficiency of this information transfer have the same effect on the diversity of the application domain?** Our study investigates the answers in part by performing an empirical experiment of examining the influence of visual instructions on procedural skill learning in IVR.

The aim of our study was to assess the effectiveness of visual instructions in IVR. Our study deals with two types of visual instructions: Annotation (ANN) and Tutor (TUT). ANN is a 3D representation of

- *Hyeopwoo Lee and Hyun S. Yang are with the School of Computing at KAIST. Emails: leehyeopwoo@kaist.ac.kr; hsyang@kaist.ac.kr*
- *Hyejin Kim, Youngnoh Goh, and Daseong Han are with Handong Global University. Emails: gobetty20@gmail.com; youngnoh.goh@gmail.com; dshan@handong.edu*
- *Diego Vilela Monteiro and Hai-Ning Liang are with the Department of Computer Science and Software Engineering at Xi'an Jiaotong-Liverpool University. Emails: d.monteiro@xjtlu.edu.cn; haining.liang@xjtlu.edu.cn*
- *Jinki Jung (corresponding author) is with Digital Maritime Consultancy. Email: your.jinki.jung@gmail.com*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

978-1-7281-4765-9/19/$31.00 ©2019 IEEE
DOI 10.1109/ISMAR.2019.00030

spatial anchor and/or abstraction of gestures that a user should take in the form of virtual objects, e.g., a 3D text, an arrow, a bounding sphere. ANN also uses a combination of virtual objects for clarity. TUT is a prior demonstration of a task by a virtual character such as a tutor. In TUT, the user is asked to imitate the tutor's demonstration. ANN is the one of the most commonly used interfaces for delivering information in VR and AR, and TUT can be considered to simulate the traditionally performed mentor-mentee (or apprenticeship) training model.

In order to delicate and precise design of tasks, we define a visual instruction, which conveys information on requirements of task completion visually. A visual instruction is defined as virtual objects that exist separately from the virtual environment (VE). We define the two elementary features of a visual instruction, i.e., **terminus** which is a description of a 3D position with a range and **action** that the user should take in the task. Terminus-action (TA) for a visual instruction will be used to evaluate the integrity of the information transfer as well as the implementation of it. The use of TA as a unit of procedural task is also to convey the same amount of information in two different instructions. Our experiment is limited to IVR which has enormous advantages over controlled experiments involving the collection of participant experience, data collection scope, and replicable scenario and environments [9]. Moreover we propose the mapping of user's behavior and a string based on the terminus and the action which enables an automatic evaluation of the user's performances.

Our experiment has been designed to test the effect of visual instructions on user behavior with tasks that focus on terminus (maze escape, ME for short), action (stretching exercises, SE for short), and both terminus and action (crane manipulation, CM for short) to simulate the context of various applications (Figure 1). The effectiveness is analyzed into participants' recall performances, i.e., recall time, accuracy, and error. We also measured skill decay through the weekly results measurement as procedural skills are highly susceptible to skills decay [19, 23]. Our contributions are as follows: 1) Presenting a novel study on the comparison of visual instructions for procedural learning in immersive VR. 2) Proposing a procedural task design scheme based on terminus-action unit. 3) Conducting implementation and evaluation of different types of procedural tasks (ME, SE, and CM). 4) Proposing a string distance based error measurement for a procedural task evaluation. As IVR has been studied as an instructional medium through experience [36], our experimental results are expected to be used for effective task design in IVR, taking into account the relationship between terminus and action on a particular topic. Based on the second and third contributions, our automated log-based user evaluation can also be used extensively in quantitative analysis of user studies in IVR. In Section 2, we review related literature on visual instruction and procedural tasks, while we define several terminologies with regards to our experiment design in Section 3. Section 4 and 5 describe the experimental results and the following discussion. Conclusion is drawn in Section 6.

2 Related work

We will first review the previous studies on the visual instructions, in terms of annotation and tutor and then describe related literature (in terms of procedural tasks in IVR) and the three tasks defined in our experiment.

2.1 Visual instructions

Although much work has been done on the effect of visual attributes on VR and AR (e.g., field of view [35], visual complexity [27, 35], and point of view [17]). there are very few studies on visual instructions in IVR. Ritz and Buss have noted the potential of IVR as effective, practical learning environments and as instructional design of CAVE systems for pedagogical purpose [37]. They emphasize the compactness of instructions, information, and connectivity to physical sources of information in VEs, e.g., placing text or visual cues directly next to relevant parts of a diagram to prevent learners from using memory resources to mentally integrate them. The advantages of the less abstract three-dimensional perceptual representation of the task was also pointed out in Wearable AR [6]. They analyzed the effects of the three instruction types (paper, static 3D, and dynamic 3D) for a LEGO assembly task and concluded that the dynamic 3D representation was superior to others in time performance, but with no statistical difference in subjective reports. Their results aligned with the fact that the integration of motion and stereo cues is advantageous in perceiving 3D content.

Annotations in VR place semantic and spatial markers to convey the high-dimensionality of data efficiently [11] as well as intuitively handle such data through interaction in three-dimensional space [32]. Pick et al. proposed a guided workflow by a series of interaction techniques such as annotations, spatial markers, and mid-air display for solving individual subtasks that showed better task performance and user experience than other subject groups [32]. Annotations are also used to induce a user's specific behavior through an intuitive three-dimensional indicator. Covaci et al. investigated the effect of the guidance type that expresses the anticipated path of the perspective and the ball on the training effect in free-throw training in basket ball [10]. The result supported that the visual instructions assist users in successfully performing a precision-aiming task, which could achieve optimal release parameters, as experts do. Learning from a tutor has been a traditional way of training, based on strength in many prior experiences from the learner's perspective [20]. A virtual tutor, also referred to as a pedagogical agent, is a human-like virtual agent that helps a student understand and learn contents [21]. Although the participant was aware that the agent was purely virtual, social interactions with the agent have been found to further enhance the immersive feeling and amplify the learning experience [15]. Hassani et al. has shown a learning-evaluation system with embedded pedagogical agents by learner's interaction and their computational model to measure improvements in language skill proficiency [18].

2.2 Procedural tasks

Procedural tasks are defined as tasks that have a number of coherent steps, which include the application of both cognitive and motor skills [22, 25]. It has been noted that VR can transfer a high degree of detail of the procedural tasks depending on the learning objectives and required levels of physical and psychological fidelity [25]. The ability to navigate is one of the most commonly used skills in both real and virtual environments [16, 41] and a maze has been used from very early on as an environment for evaluating spatial knowledge transfer [5,43]. Waller et al. measured the spatial knowledge performance for the groups with exposure to the actual maze, a map of the environment, desktop VR, or immersive VR with a head-tracked HMD [43]. The result showed that the real-world training was the most effective overall, and immersive VR showed better performances than the other non-real conditions for longer periods of training. Volmer et al. measured effectiveness of spatial augmented reality cues for procedural tasks. They demonstrated that the spatial cues provided by the projector improved users' task time and mental effort in the procedural tasks. [42] Motor skills have also been studied extensively as a type of procedural task [33, 38]. Most of the studies on motor skills were designed to evaluate a user's behavioral performance using sensors like Kinect [1] and Wii-mote [33]. NrityaGuru [2], the study most similar to our study, assessed whether the learner took a similar pose as the expert (through the position and angle of the 20 body parts) for eight consecutive poses. Complex machines and equipment manipulation has been the most popular and important procedural task in the use of VR for learning transfer due to its industrial importance and practicality [7, 14, 19]. Ganier et al. mentioned the efficiency of VE in sensory-motor skill learning and conducted experiments on a task to set up an element of the tank suspension system [14]. The results suggest that a procedure can be successfully transferred from the virtual to the real world. Studies indicate the immersive type of interface induced better performances in the assessment of technical skills [22]. Although many studies have been conducted on procedural tasks, few have focused on the effect of the visual instructions in IVR.

2.3 Hypotheses

Even though there are no perfectly matched studies on the comparison of ANN and TUT, we set hypotheses on the effect of ANN and TUT for

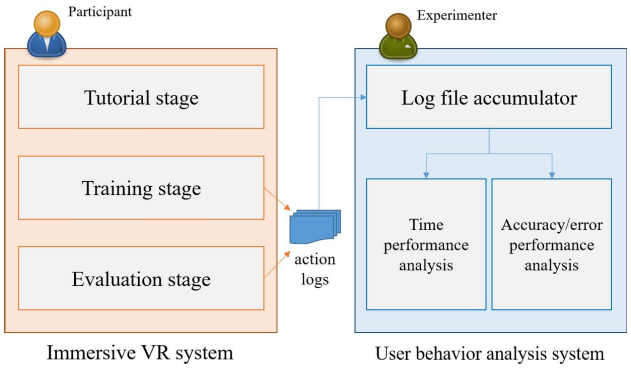

Fig. 2: User-centric evaluation system architecture

clarity. With regard to providing a demonstrating avatar in TUT, ANN and TUT can be approximated with the learning transfer comparison of the first and third person perspectives. From the founding that having the first person view contributes wider scope of visual perception in 3D navigation [24], we hypothesize that ANN is more suitable than TUT in ME. We also set two hypotheses that TUT is a better visual instruction for SE based on the previous work which demonstrated that the existence of virtual tutor influenced positive effect on self-explanation related to motor skills [3]. Based on the results that participants with the lower levels of complexity resulted much higher score and better strategy [35], we assume that ANN that has lower complexity than TUT induces a more accurate and faster transfer of learning in complex tasks such as CM. There are also related works on the use of ANN in assembly tasks [8, 34]. In summary:

Hypothesis 1. *ANN induces more accurate transfer than TUT in ME.*

Hypothesis 2. *ANN induces faster transfer than TUT in ME.*

Hypothesis 3. *TUT induces more accurate transfer than ANN in SE.*

Hypothesis 4. *TUT induces more faster transfer than ANN in SE.*

Hypothesis 5. *ANN induces more accurate transfer than TUT in CM.*

Hypothesis 6. *ANN induces faster transfer than TUT in CM.*

3 Material and Methods

It is important to maintain consistency in training transfer during the experiment, even if the two visual instruction methods provide different user experiences. We applied the part-task training design [25] which provides a divide-and-conquer approach to transfer one or a limited number of constituent skills in order for the learner to reach a high level of automaticity for the skills. We designed a pair of terminus and action as a part-task training scenario unit, and implemented a scenario by weaving the pairs sequentially. By following the part-task design, ANN provides visual description of terminus and action using static spatial UI elements in each step. TUT on the other hand provides an experience of observing demonstrations of a virtual tutor. Our empirical study investigates three types of procedural tasks to testify the effectiveness of visual instructions in various contexts. Our experiment covers the implementation of ANN and TUT according to three kinds of tasks, evaluation of performance, and usability test. Figure 2 shows the overall diagram of our system. User's behavioral data is captured from the immersive VR system in the training and evaluation stage, and then used to evaluate user's performance in the user behavior analysis system. We applied the string distance metric, which has been used in DNA sequence comparison [44], to the recall error measurement of the behavioral data by using character mapping for the TA pair. In the experiment, other information except the visual instructions, such as virtual environments and audio effects, were controlled.

3.1 Terminus-action-based training design

The idea of applying the divide-and-conquer concept to training is based on the segmentation of part-task in which a task can be considered as a series of subtasks that has identifiable endpoints [45]. We define a set of *condition, terminus, and action* (referred to as CTA) as an elementary segmented unit of a training. Condition in the CTA is defined as a contextual trigger that causes execution of terminus and conduct. For example, systematic failures or human faults in an accident model can be applied to the condition. Terminus in the CTA is defined as range-based spatial knowledge that must be memorized for problem solving. The terminus can be represented by a three-dimensional position with a range, which can be either larger than a person, or very small, depending on the context. In the systematic level, the terminus is used as an explicit boundary of whether a behavior happened within the spatial ranges. Action in the CTA is defined as a motor skill of person to be performed at the terminus. The action can also be a series of CTAs to express a hierarchical action or skill. CTA has been expressed in the form of texts, voice, materials, and a demonstration of assistants. Our comprehensive implementation of CTA is available online[1].

For simplicity, our experiment limits the condition of a part-task to the success of previous part-tasks so that the procedural task has no conditional branch. We therefore use a terminus-action (TA) instead of a CTA. A part-task, which is a subset of constituents of a procedural task, is represented as a TA.

In the implementation of TA with ANN, we used a position of 3D text or a geometric model to indicate the terminus and its content to describe an action a user should take (in either a descriptive or an abstract level). ANN then provides a visual guide of the minimum requirements to move on to the next step. TUT transfers a terminus and action through the consecutive and natural behavior of the tutor. The learner also has the effect of simulating TAs through the meta-cognition of the tutor [3]. The difference between TUT and ANN can also be found from the point of observation; ANN provides TAs from the first-person perspective while TUT demonstrates TAs in the third-person perspective.

3.2 Task description

In our experiment, we used three tasks based on the TA structure to accommodate the various contexts of virtual training. Maze escape (ME) is a task to escape a maze by successfully passing five intersections from the starting point to the escaping point. The terminus in ME is the direction at each intersection and walking toward terminus is the only one action in ME. Stretching exercises (SE) requires recalling five consecutive poses (taking into account the action as the target pose). The terminus in the SE is the location of each body part, and there is no change in the terminus until the end of the task. Crane manipulation (CM) asks the user to conduct five consecutive machine operations that make changes in both terminus and action. The terminus in CM is the location of each device and the action is defined differently depending on how the device is operated. Therefore, CM requires higher cognitive loads due to these contextual changes. In a later section, we describe the three tasks in detail. We also describe the pilot studies for each task in order to enhance better usability of visual instructions and plausibility of a circumstances in a virtual scenario from given prior knowledge [39].

3.2.1 Task1: Maze escape (ME)

The ME task design is inspired by Waller et al's work [43] which measures the procedure learning transfer in spatial knowledge. The goal of ME is to reach the exit of a virtual maze with five intersections that have two choices in each. We set up cognitive anchors in the maze inspired by previous works [26, 43] to offset the variances in the spatioperceptual ability of an individual participant. As the cognitive anchors, we placed reference pictures on each intersection to minimize this effect as shown in Figure 1 (c), (f). The referenced pictures presented apple and tomato at the first intersection, flamingo and blue bird at the second, car and bus at the third, computer and Bit-coin at the fourth, water

[1] https://github.com/VirtualityForSafety/ACTA

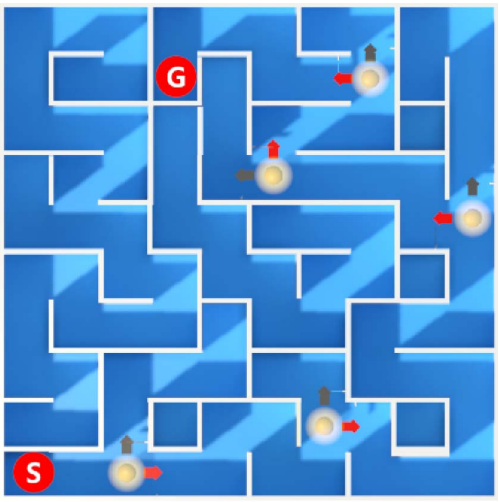

Fig. 3: Top view of the map of ANN in ME. The red circle with 'S' and 'G' indicate start goal positions respectively. From S to G, there exist 5 intersections that are represented as white sphere with two choices.

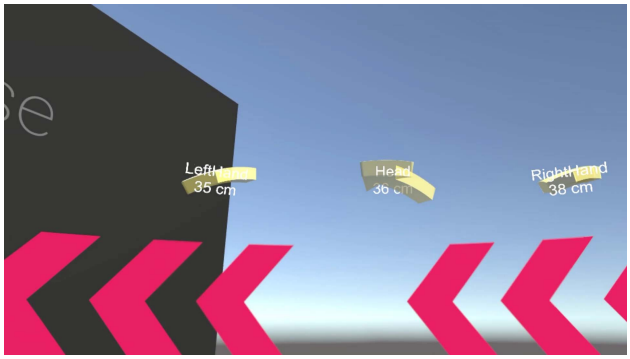

Fig. 4: Out-of-view visual guidance indicator of ANN in ME; three yellow 3D arrows for the body parts (middle) and red 3D arrows to make user face the front side (low).

and juice at the last. The correct choices of referenced pictures were tomato, blue bird, car, Bit-coin, and water in order. We designed the maze to have only one correct path. The total number of possible binary actions through five intersections was 2^5. For the walking inside the maze, we employed a swinging back-and-forth gesture while the user was pressing both trigger buttons [13]. More specifically, the user was able to walk forward along the direction they were looking at, press the trigger button and make a swing gesture. Also, the user could move backwards while pressing touch button and making a swing gesture. The tutorial stage in ME helped familiarize the user with the controller-driven walking method. The user learned how to arrive at the exit of the maze with the given visual instructions during the training stage. In the evaluation stage, the user had to escape the maze themselves as soon as possible in the same environment. When the user took a wrong path, they had to return to the intersection again and move down the correct one. Figure 3 represents the maze we designed for our experiment. The distance between the walls in VE was 4 meters.

The quantitative evaluation criterion of the ME was whether the user had entered the wrong path, which was recorded through invisible collision objects placed at the entry point of a wrong path. In ANN of the ME, the intersection was visualized as a glowing and transparent capsule in the air, while the path guidance was two three-dimensional arrows on floor; red one for the correct path and the gray for the wrong path Figure 1 (c). In TUT of the ME, a virtual tutor was shown to the user to guide them to the correct path. We implemented the tutor guidance using Unity's A* algorithm-based navigation function. A virtual character always moved along the correct path and approximately 5 m ahead of the user so that the user wouldn't be able to miss the character. The user followed the path that the character walked. Unlike ANN of the ME, the virtual character was visible to the user the entire time. The MIT license version of the ME implementation is available online [2].

3.2.2 Task2: Stretching exercises (SE)

Stretching exercises (SE) was designed to measure the performance of procedural motor skills. SE was designed to take five upper-body poses in order through a simple stretching process. An evaluation in SE was performed to determine how accurately the user took the five poses in the correct order. We also measure the accuracy and elapsed time for the sequence of poses. Figure 5 represents the sequence with a dummy skeleton: 1) Forward bend pose: the tutor bends forward and stretches both arms downwards; 2) Triangle pose: the tutor bends left while laterally stretching out both arms so that the left hand moves closer to the floor; 3) Side-bend stretch pose: The tutor puts his right hand on the right of her waist and bends right while lifting the left hand; 4) Mountain pose: the tutor lifts both arms above their heads while stretching the arms straight; 5) Neck relaxing pose: the tutor puts both hands on their waists while standing straight and tilts their head backwards to look up at the sky. For each pose, the voice explanation of taking the pose in sequence was given. The tutorial stage in SE gave a brief explanation on the task objective and capturing their pose with the controllers. The tutorial stage also carried out calibration for accurate visualization and evaluation so that the reference pose of the dummy skeleton (Figure 5) was adjusted to match the length of the arm and neck of the user. During the calibration, the head and hands were forced to take a 'T' pose and straight upward pose to measure arm and shoulder length. In the training stage in SE, the user had to take pose and push the trigger button of both hands to move to the next step. In the evaluation stage, the user was asked to take the pose in order.

The evaluation of SE was based on the three-dimensional position of the head and hands when the user was pressing the trigger button of the controller. Here the position of three body parts were given by the tracking of VR hardware. However, as there was no restriction on the action that the user can take as well as a reference for the perfect pose, operation and direction were available. The experimenter evaluated the accuracy of the user's poses behind the user. For the sake of simplicity, the accuracy criterion for the individual static pose was whether each body part was on the required position (e.g, was the right hand near the side of body) and did not take into account other factors such as head direction and hand posture. If the user took the learned pose, the label of the action was recorded like 'pose:mountain'. The poses that were not in the learned poses were recorded as 'wrong'. SE had 6 choices during the 5 procedures, so the number of all possible actions during the 5 procedures is 6^5. Although the actions were categorized as one of six, the actual number of degrees of freedom was the highest as it did not impose any limitations on the actions that could be taken.

In the case of ANN in SE, target position in SE was defined as the three-dimensional coordinates of the hand and the head that should be located at the end when posing. The portions where the head and the hand position had to be entered were represented as a three-dimensional bounding sphere. In order to provide out-of-view visualization of the body parts, we employed 3D arrows and texts to indicate the direction and distance from each body part to the target bounding sphere (Figure 4). In ANN, the user had to press the trigger button while all the body parts were in the bounding sphere to move on to the next step. Here the criteria of inclusion of the bounding sphere was 30 cm. In representing a pose, ANN initially utilized only three spheres, but it was modified to express its relationship using a dummy skeleton with the opinion that it was too difficult to find the relationship among spheres in the pilot study. TUT for SE implemented the tutor's guidance by voice and demonstration. As a real trainer demonstrates exemplary stretching

[2]https://github.com/VirtualityForSafety/MazeEscape3D

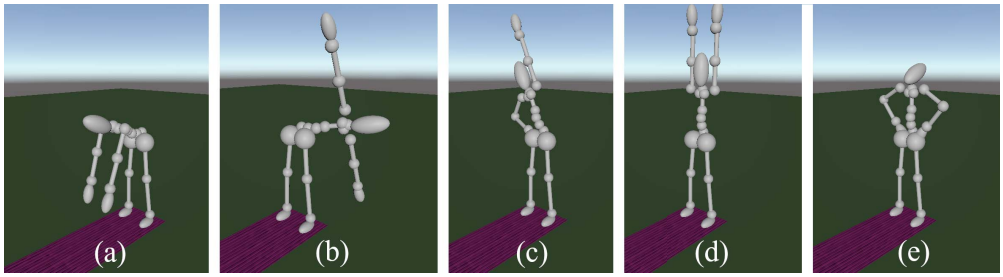

Fig. 5: Five poses in SE: (a) forward bend; (b) triangle; (c) side bend stretch; (d) mountain; (e) neck relaxing.

postures while the learner watches and typically follows these in the opposite direction, we set the virtual tutor to stand face to face with the learner at the fixed position in the virtual training environment. In the tutorial stage, the learner was instructed to use his/her left hand to imitate the motion of the tutor's right hand. During the training stage, the user was trained on five stretching postures by imitating the tutor's demonstration. In TUT there was no restriction on the pose that the user would take. The MIT license version of the SE implementation is available online [3].

3.2.3 Task3: Crane Manipulation (CM)

The goal of the CM is to train a user to manipulate a crane to lift a green box. We designed the crane manipulation because the VR-based operating training for complex technical system is considered to a worthwhile method in the industry [14, 22]. A crane that is impossible to perform by a person, e.g. lifting a huge container, accomplishes the task by manipulating a series of operations with a controller. A crane controller (low level) to be operated directly and a crane to move through the operation (high level). The crane is shown to the user only through the large screen on the front. The user performs a fine-level operation based on the motivation for the success of the high level, and the feedback of the operation is provided through the display of the crane. This hierarchical design was motivated by our CM pilot study which provides only a controller that does not provide either motivation or feedback. The constraints of the fine-level equipment are inherited from the high level, e.g., when the crane hook holds an object, the joystick was locked and inoperable. Figure 7 shows four types of controllers in the fine-level: joystick, wheel, buttons, and lever. The joystick has x and y coordinates that represent a scaled-down position of the hook of the crane projected on the ground plane. When the user located the joystick in a certain coordinate, the crane in the screen moves to the corresponding position. The lowering or lifting of the hook is done by rotating the wheel controller clockwise or counterclockwise. There were six buttons used to grab or release an object with the crane hook. The lever is used to change the gear level for lowering or lifting the hook. The order of manipulation with the equipment is as follows: 1) The joystick should be handled to make its coordinates (x, y) = (43, -29) so as to locate the hook of the crane over the box; 2) The wheel should be rotated by 360 degrees clockwise to lower the hook until it touches the box; 3) The correct button (located in the middle of lower row in Figure 7) should be pressed to grab the box; 4) The lever should be pulled down to change the gear level from 2 to 1; 5) To lift up the box, the wheel should be rotated by 360 degrees counterclockwise. Finite numbers as the result of the manipulation was shown in the terminal that is located below each equipment (Figure 7). To interact with virtual equipment, the user had to hover a controller over the equipment. They were then able to manipulate it by holding the trigger button. The tutorial stage gave explanations of name and manipulation of each equipment and allowed the user to practice manipulation. In the training stage, the five subtasks were informed in order. In the evaluation stage, the cranes had to be manipulated by the user so as to pick up the object in the same environment as the training stage. For quantitative evaluation, the user's behavior was labeled by a combination of the name of the equipment and the manipulation. For example, when the user rotated the wheel continuously more than 30 degrees clockwise (CW) or counter-clockwise (CCW), the behavior for the wheel was logged as 'wheel:CW' or 'wheel:CCW' respectively.

According to the annotation definition introduced in this paper, the instruction of ANN for each equipment was given as a 3D text located on the corresponding equipment (Figure 1 (a)). The instruction contains the target number that the user must make through manipulation. In the case of the wheel, the direction of rotation (clockwise or counter-clockwise) was displayed as a 3D arrow. In the case of the button, the annotation indicated the correct button (labeled as fourth button) as a 3D bar. TUT employed the tutor to demonstrate the equipment, allowing the user to operate the equipment alongside the tutor. Participants in TUT have to follow the manipulation after watching tutor's demonstration.

After the demonstration in each step, we also make the tutor step aside and look at the controller that has been used just before so that the user may be aware of his/her turn to practice the current step with the controller. When the user accomplishes the practice, the tutor moves to the place near the controller to be used in the next step. This cycle between tutor's demonstration and user's practice is repeated until the last step. The MIT license version of the CM implementation is available online [4].

3.3 Measures

User performance measurements are made with execution time and accuracy. Accuracy indicates whether the user has performed the task without a mistake. Although measuring the accuracy of TA is defined by task, it is very important to consider the execution order in the procedural task evaluation. An example of CM evaluation is given in Table 1 to illustrate how it evaluate a user taking into account the order. From the human-readable instructions, it is possible to derive the TA of each level and automatically check the completeness of the TA according to the context of the task described above. We have mapped the TA pair as a behavior code, a unique character, assuming that the user's behavior with intention can be expressed as a TA pair. Under such assumptions, the correct behavior code can be defined step by step as shown in Table 1. Note that in the example, order 2 and 5 have the same terminus, but the correct behavior code is mapped to B and E, respectively, because of the difference in action. In the implementation, a behavior code was generated when the subtask terminating criterion is satisfied, e.g., the moment the user satisfies the target parameter of the equipment in the CM, the moment the user presses the trigger buttons in the SE, and the moment the user collides the invisible trigger of the correct/incorrect path in the ME, with the corresponding time and user's behavior remained as a log.

At the completion of a task, the user's behavior is represented by a behavior string (e.g., ABCDE), which is a sequence of behavior codes. Accuracy was simply measured by matching the string in the exact order, i.e., 1 for correct and 0 for incorrect. As a subsidiary measurement to quantify how the visual instruction negatively impacts on procedural recall, we propose the use of string comparison as an

[3] https://github.com/VirtualityForSafety/StretchingExercise3D

[4] https://github.com/VirtualityForSafety/CraneManipulation3D

Fig. 6: A participant taking the triangle pose in the SE evaluation

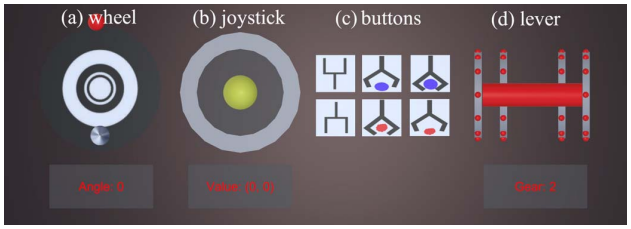

Fig. 7: Top view of four equipment in CM. (a) Wheel that turns clockwise or counterclockwise to pick up a crane hook, (b) Joystick that contains 2 dimensional degrees of freedom (x,y) to translate the position of crane hook, (c) Six buttons, including the grab button, (d) Lever which is pulled up or down to change gears of from 1 to 3

procedural task error metric. In particular we apply the Levenshtein distance (LD) [31] to comparison of behavior strings. LD calculates the least number of edit operations that are necessary to modify one string to obtain another string. We use the properties of LD to obtain the number of unit behaviors (number of edit) that the user has to be trained further to get the correct behavior string. For example the LD distance of the correct and example behavior codes ("ABCDE" and "EABCD" respectively) in Table 1 is 2. This means the user should do not perform "E" from the front and should perform "E" at the end, by maintaining the rest of the connectivity of behavior codes. Note that the LD increases the distance proportionally to the string length mismatch, which can indicate penalties for repeated misbehavior for a participant. In the same example with the participant's behavior string 'ABCDEISMAR', the LD outputs 5. In the behavior string generation we disregard redundant behavior codes until different code appeared. For the time of task completion we use the completion time of the fifth subtask in CM and SE, and reaches the goal point in ME.

3.4 Apparatus

In TUT, the tutor's motion was captured by a whole body tracker with a suit capturing the body using 37 markers attached and 12 OptiTrack Motion capture cameras (1664 $\times$ 1088 resolution and 120 FPS for each). One of the experimenters volunteered to demonstrate the tutor's motion for three tasks. The captured motions were post-processed via OptiTrack Motive 2.0.2 Final with an i7-6700 CPU and 16GB of RAM. At runtime, we applied the IK controller which is one of Unity's default functions to make the tutor look at the correct equipment direction.

We individually configured the space for evaluation of each task in the experiment as shown in Figure 6. Each space for a task was larger than 3 m $\times$ 3 m. An HMD and controller for the HTC Vive were used as immersive VR hardware. All groups were trained and measured on a desktop environment with Intel i7-7700, 8GB memory, NVIDIA GeForce GTX 970 on Windows 10 (64 bit). Audio effects were provided through Britz G2 headphones. The voice explanation was provided by converting the manuscript, originally written in text, to a female voice using Microsoft's text-to-speech library (System.Speech.Synthesis in Microsoft NET Framework). For CM, however, due to the specific implementation issue, we used a male model for the demonstration and a female model for the explanation. For the interaction, hovering of the controller over a virtual object and pressing the trigger button of the controller were chosen and 3D menu for manually selecting stage was given.

3.5 Participants

There were 24 participants aged between 19 and 33 (M = 22.96, SD = 3.52). There were 15 males and 9 females with undergraduate/graduate students. Each participant received approximately 30 USD compensation. Nineteen subjects had prior experience with VR, and none of them had experience similar to crane manipulation. The average score of people confident about operating the equipment for the first time was 3.46 (Likert scale = 5, sd = 0.82) There were 6 people who had practiced body exercises such as yoga or Pilates. The average score of a person who is confident about taking others poses was 3.83(Likert scale = 5, sd = 0.75) 11 people had experienced a real maze escape. The average score of a person who is confident about escaping the maze was 3.33 (Likert scale = 5, sd = 0.75). The average score of person who is confident about reaching the destination from the first trip was 4 (Likert scale = 5, sd = 0.82) 14 subjects had prior experience of visual instructions similar to ANN (8 from VR, 5 from games, and 1 from AR). 7 people had prior experience of a virtual tutor (1 from VR and 6 from games). One subject had color amblyopia.

3.6 Procedure

Participants were asked about demographic information and prior experience at the beginning of the experiment. Participants were randomly assigned to two groups. In the initial test, all participants performed the specified visual instructions condition for the three tasks assigned in random order. For all tasks, the participants played the tutorial stage which provided the description of the environment and training information so that they could adapt to the environment and interaction. At the tutorial stage, participants were able to test all interactions freely without time constraints. After completing the tutorial, they entered the training stage without time limitation through the 3D menu selection. Participants who completed the training stage carried out the evaluation stage three times in a row for assessing the performance. After finishing a task, participants moved to the next task space and repeated the aforementioned procedure. Experimenters placed in each task space helped the participants wear the VR equipment and headphones, and responded to queries throughout the experiment. After all tasks were completed, the participants filled out questionnaires on the workload, experience, and memory of procedural information for each task. In the retention test, all the participants performed the evaluation of the three tasks again (in a random rearranged order three times without any training or tutorial stages) and conducted a post-questionnaire survey.

4 Result

The results from the initial and retention tests are presented here on a task-wise basis. We measured three times per test to see the recall pattern [19]. Figure 8, Figure 9, and Table 3 shows time performance graphs, accuracy graphs, and a statistics of recall performance measurements for three tasks. We applied strict criteria ($p \leq 0.005$) for a significance.

Maze escape (ME) To comparatively evaluate ANN and TUT (n = 24), the recall time, procedure accuracy, and recall error were submitted to a 2 $\times$ 6 mixed-factorial analysis of variance (ANOVA), in which the visual instruction group served as the between-subjects variable, and the time of attempt served as the within-subjects variable. 2 $\times$ 6 ANOVA analysis for the recall time of ME found a significant main effect for the attempt, $F(1,22) = 59.52$, $p < 0.001$, but not the group, $F(1,22) = 0.08$, $p = 0.778$. Analytical pairwise comparisons of the time for group per attempt showed no statistical significance. In the ANOVA for the procedure accuracy of ME, a significant main effect was found for the group ($F(1,22) = 15.66$, $p < 0.001$), but not the

No	Instruction	Terminus/Action	Correct behavior code	Example behavior code
1	Control the joystick to set the X and Y coordinates to 43 and -29	joystick/set position	A	E
2	Turn the wheel clockwise to go down the crane hook.	wheel/turn clockwise	B	A
3	Press the grab button to let the crane hook hold the object.	grab button/press	C	B
4	Pull the lever to make the gear level to one.	lever/pull	D	C
5	Turn the wheel counterclockwise to lift the crane hook.	wheel/turn counterclockwise	E	D

Table 1: Step-by-step mapping of a behavior code from procedural instruction. The scenario of CM is used.

attempt ($F(1,22) = 7.57$, $p = 0.006$). There was no interaction effect, ($F(1,22) = 2.11$, $p = 0.148$). ANOVA for the recall error of ME revealed a significant main effects for the group, $F(1,22) = 13.65$, $p < 0.001$, but not the attempt, $F(1,22) = 7.58$, $p = 0.007$.

Stretching exercise (SE) From the 2×6 ANOVA analysis for the recall time of SE, a significant main effect was found for the attempt, $F(1,22) = 69.54$, $p < 0.001$, and also for the group, $F(1,22) = 15.78$, $p < 0.001$. There was a significant interaction for the time between the group and attempt, $F(1, 22) = 35.89$, , $p < 0.001$. Post hoc comparison using the Tukey HSD test indicated that the TUT were significantly faster than ANN. In the ANOVA for the procedure accuracy of SE, no effects were found for neither both groups, $F(1,22) = 1.56$, $p = 0.214$, nor attempt ,$F(1,22) = 0.5$, $p = 0.478$. ANOVA for the recall error of SE revealed no effects for groups, $F(1,22) = 1.03$, $p = 0.311$, and attempt, $F(1,22) = 1.27$, $p = 0.261$. Analytical pairwise comparisons of both accuracy and error for a group per attempt showed no statistical significance.

Crane manipulation (CM) 2×6 ANOVA detected a significant main effect of the recall time for the attempt, $F(1,22) = 30.06$, $p < 0.001$, but no effect for the group, $F(1,22) = 0.02$, $p = 0.879$. There was no interaction effect, $F(1,22) = 0.62$, $p = 0.678$. ANOVA for the procedure accuracy of CM showed no effects for neither groups ($F(1,22) = 0.03$, $p = 0.861$) nor attempts ($F(1,22) = 0.13$, $p = 0.719$). No interaction effect was found, $F(1,22) = 2.930$, $p = 0.0892$. ANOVA for the recall error of CM revealed no significant main effects for the attempt, $F(1,22) = 8.39$, $p = 0.005$, but not the group, $F(1,22) = 2.22$, $p = 0.142$. Analytical pairwise comparisons of both accuracy and error for a group per attempt showed no statistical significance.

5 Discussion

In our experiment, ANN group showed statistically better performances than TUT for the measurement in accuracy and error at ME, and time performance at SE. This results coincide with the previous works which demonstrate the lower complexity yields the better strategy and learning transfer [27, 28, 35]. For ME, our result confirmed Hypothesis 1 (Figure 9 (a)). The accuracy of the ANN was statistically 1.3 times higher than that of the TUT, although the accuracy of the ME was much higher than that of the other tasks. But there was no significant difference in ME between groups in terms of time. In ME, which mainly taking account into terminus, ANN provides static and more explicit three-dimensional arrows for terminus, while the guiding tutor that are always in the perspective of the participant seemed to disturb the perception of the surrounding environment and paradoxically became a factor that impeded the localization. This distraction can be also interpreted as the peripersonal space [24] which paradoxically became a factor that disturbed the perception of the surrounding environment. We conclude that the use of static spatial anchors is beneficial to transfer terminus more accurately than the use of a virtual guide. In SE, the comparison in the time performance showed the superiority of ANN which is the complete opposite of Hypothesis 4, but the learning curves showed distinctive patterns in the initial and retention tests (Figure 8 (b)). The TUT group had a very long recall time initially, but the subsequent recall time of them was very short and stable. Interestingly the comparison on the time performance graphs of the retention test suggests a higher learning curve for ANN is found than TUT, which is the opposite to the pattern of the initial test. In the training the participants of ANN were asked to position their head and hands in a three-dimensional spherical annotation. Participants were able to see the overall pose but showed a pattern of immediately matching their head and hands to the annotation. Confirming the position of both hands at the time they fixed their heads to the annotation position, and this caused a considerable change in the visual field. In summary, due to the obvious visualization of the criteria, the ANN participants tended to satisfy the condition. On the other hand, TUT participants were asked to watch tutor's demonstration. The TUT group was able to take action relatively quickly because the tutor showed the series of actions in step by step, which was not provided in ANN. This disparity difference would lead to the difference in the learning curve, especially for the interpretation of the longer recall time of TUT group (4.45 times) at the first attempt as the time for recalling the actions of the tutor in third-person perspective and converting it into the participant's own actions. We conclude that the visual instruction has no effect on neither accuracy nor error, but a significant effect on the time performance. In CM, no significant differences were found between ANN and TUT in time and accuracy, which can be interpreted as the choice between ANN and TUT does not have a significant impact on the training of complex technical system. The result of CM did not confirm both Hypothesis 5 and 6. However, there was an interesting difference in the participants' behavior in the experiment. During the tests, ANN participants tended to focus on the equipment while TUT participants tended to check the screen frequently as tutor did in the demonstration. We were able to infer from these behaviors that the TUT participants tried to mimic the tutor rather than understanding the manipulation. We conclude that visual instruction affects the user's experience, but does not affect the user's performance.

Task-wise, it was shown that the number of average errors increases in order of ME, SE, and CM, and thus it can be interpreted that there were increase of cognitive loads of participants in order of recalling terminus, action, and terminus-action. At the beginning of the test, many participants were embarrassed and did not recall a pattern despite the fact that they were expecting to be evaluated shortly after training. As the evaluation was repeated in a test, however, some performances were improved despite no intervention from the experimenter. All the tasks the execution speed tended to decrease as the attempt was repeated. However only CM tended to decrease in error according to the number of repetition, while others did not. The results of the retention test, which was performed a week after the performance of procedural learning using immersive VR, showed low knowledge loss over time, as demonstrated in [14, 19]. But the interpretation of the recalling pattern as the learning curve is limited due to a small number of measurements compared to [19] that has 120 steps.

The limit of our experiment is that the tasks were all virtual. Therefore, we could not measure the external validity that represents the learning transfer. Also, in the case of SE, only the static posture was in the scope, not the dynamic behavior due to the limitation of measurement. If the SE can perform a comparison of dynamic behavior, it would be possible to analyze the detailed accuracy of ANN and TUT. In our implementation of TUT in ME, due to the movement of the tutor was too sensitive to the distance from the user, the movement was too artificial not natural, which might affect negatively to user's experience. For the string comparison as an error metric, we were able to apply it because of the constraint on strictly sequential operations. But the tasks with operations that have a non-unique sequence will be not appropriate to this measurement.

			initial test			retention test		
			1	2	3	4	5	6
ME	Time	ANN	116.94 (49.82)	88.53 (37.21)	64.88 (17.97)	72.00 (19.91)	55.93 (15.56)	51.89 (9.05)
		TUT	90.47 (32.77)	99.29 (32.27)	66.32 (16.90)	72.45 (20.10)	57.95 (16.45)	56.11 (13.17)
	Accuracy	ANN	0.92 (0.29)	0.92 (0.29)	0.92 (0.29)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
		TUT	0.58 (0.51)	0.58 (0.51)	0.58 (0.51)	0.92 (0.29)	1.00 (0.00)	0.75 (0.45)
SE	Time	ANN	67.80 (15.43)	54.76 (13.66)	46.11 (12.59)	61.31 (16.85)	43.49 (15.25)	35.67 (2.87)
		TUT	301.78 (86.26)	45.74 (13.00)	44.14 (10.57)	43.40 (8.07)	37.59 (4.98)	35.10 (3.07))
	Accuracy	ANN	0.33 (0.49)	0.50 (0.52)	0.50 (0.52)	0.33 (0.49)	0.33 (0.49)	0.33 (0.49)
		TUT	0.33 (0.49)	0.67 (0.49)	0.67 (0.49)	0.45 (0.52)	0.42 (0.51)	0.42 (0.51)
CM	Time	ANN	109.97 (68.69)	65.42 (33.21)	36.74 (11.73)	79.48 (46.15)	44.88 (17.75)	38.74 (16.80)
		TUT	115.83 (44.48)	47.45 (11.44)	53.65 (57.46)	78.42 (27.20)	49.59 (17.73)	36.33 (10.89)
	Accuracy	ANN	0.50 (0.52)	0.58 (0.51)	0.83 (0.39)	0.58 (0.51)	0.67 (0.49)	0.83 (0.39)
		TUT	0.67 (0.49)	0.83 (0.39)	0.75 (0.45)	0.58 (0.51)	0.67 (0.49)	0.58 (0.51)

Table 2: Statistics of time and accuracy performances with regards to task and attempt in a form of 'mean (SD)'.

	Maze escape (ME)			Stretching exercise (SE)			Crane manipulation (CM)		
	Time	Accuracy	Error	Time	Accuracy	Error	Time	Accuracy	Error
ANN	75.02 (35.49)	**0.95 (0.2)**	**0.04 (0.2)**	**51.52 (17.11)**	0.38 (0.49)	1.3 (1.25)	62.53 (45.12)	0.66 (0.47)	3.09 (7)
TUT	73.76 (27.66)	0.73 (0.44)	0.34 (0.69)	84.62 (103.87)	0.49 (0.5)	1.08 (1.23)	63.54 (41.68)	0.68 (0.46)	1.59 (3.75)

Table 3: Overall task performance results in a form of 'mean (SD)'. Significant results are noted as bold.

6 Conclusion

In this paper, we conducted a comprehensive evaluation of ANN and TUT in procedural task learning using three tasks designed based on terminus and action pairs. The three performances of procedural learning (recall time, accuracy, and error) were measured by mapping the sequence of the participant's behaviors and events to a string and recording the time of each. The result revealed that ANN group outperformed to TUT group in the accuracy (error as well) at ME task and the time performance at SE task. In the user's behavior aspects, the ANN group tended to perform actions to satisfy the conditions given by the annotations, and the TUT group tended to mimic the behavior of the tutor. Our results led to the following conclusions: (i) For delivering procedural spatial knowledge through the terminus, ANN performed more accurate transfer than TUT, (ii) For delivering procedural actions, ANN showed better time performance even though TUT showed lower execution time after long initial recall, (iii) For delivering a mixture of procedural terminuses and actions, TUT and ANN showed no statistical differences. Even though ANN showed statistically better time performance, the use of TUT will be beneficial for the requirement of short recall time at retention phase. As a future study, more meaningful discoveries through more delicate and well-structured experimental designs of ME that is the most industrially impactful task should be investigated. Comparisons of the user experience according to the format of ANN (e.g., 3D text vs. 3D model or static vs. dynamic) will also be a valuable contribution.

Acknowledgments

The contents of this paper are the results of the research project of the Ministry of Oceans and Fisheries of Korea (A fundamental research on maritime accident prevention - phase 2, PMS3840, and H.N. Liang and D. Monteiro would like to acknowledge support from Xi'an Jiaotong-Liverpool University Key Program Special Fund (Grant KSF-A-03) and Research Development Fund (Grant RDF-16-02-40) for this research. The authors are sincerely thank the participants, the reviewers, and the members of VR/AR research community on safety, Virtuality for Safety.

References

[1] R. J. Adams, M. D. Lichter, E. T. Krepkovich, A. Ellington, M. White, and P. T. Diamond. Assessing upper extremity motor function in practice of virtual activities of daily living. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 23(2):287–296, 2015.

[2] A. Aich, T. Mallick, H. B. Bhuyan, P. P. Das, and A. K. Majumdar. Nrityaguru: A dance tutoring system for bharatanatyam using kinect. In *Computer Vision, Pattern Recognition, Image Processing, and Graphics: 6th National Conference, NCVPRIPG 2017, Mandi, India, December 16-19, 2017, Revised Selected Papers 6*, pp. 481–493. Springer, 2018.

[3] V. A. Aleven and K. R. Koedinger. An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive science*, 26(2):147–179, 2002.

[4] L. W. Anderson, D. R. Krathwohl, P. W. Airasian, K. A. Cruikshank, R. E. Mayer, P. R. Pintrich, J. Raths, and M. C. Wittrock. A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives, abridged edition. *White Plains, NY: Longman*, 2001.

[5] R. S. Astur, J. Tropp, S. Sava, R. T. Constable, and E. J. Markus. Sex differences and correlations in a virtual morris water task, a virtual radial arm maze, and mental rotation. *Behavioural brain research*, 151(1-2):103–115, 2004.

[6] S. Baldassi, G. T. Cheng, J. Chan, M. Tian, T. Christie, and M. T. Short. Exploring immersive ar instructions for procedural tasks: The role of depth, motion, and volumetric representations. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pp. 300–305. IEEE, 2016.

[7] A. C. Boud, D. J. Haniff, C. Baber, and S. Steiner. Virtual reality and augmented reality as a training tool for assembly tasks. In *iv*, p. 32. IEEE, 1999.

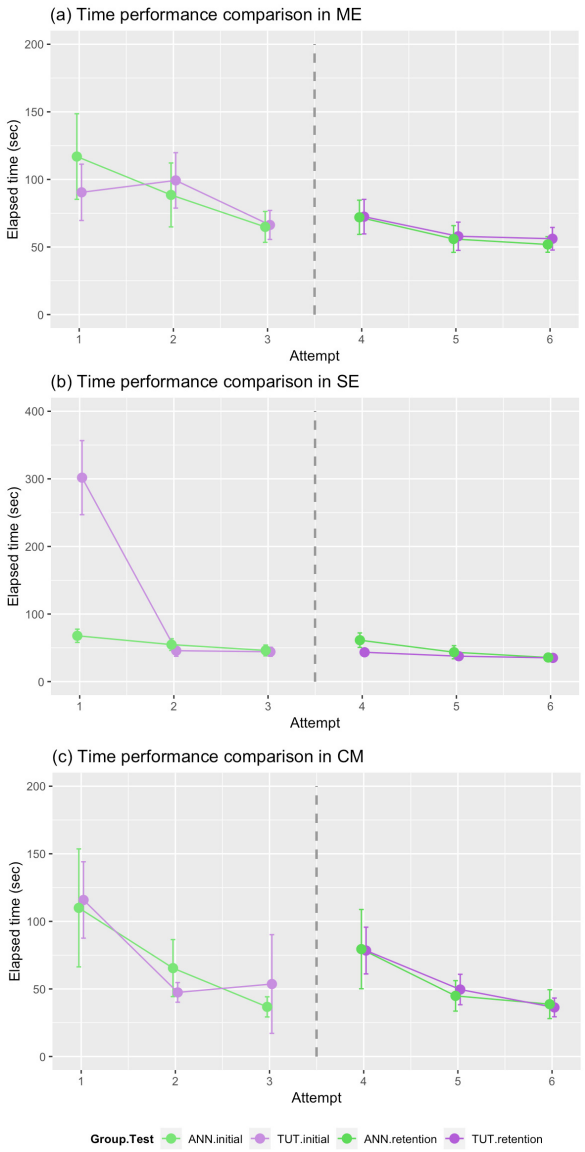

Fig. 8: Elapsed time results with a 95% of confidence interval.

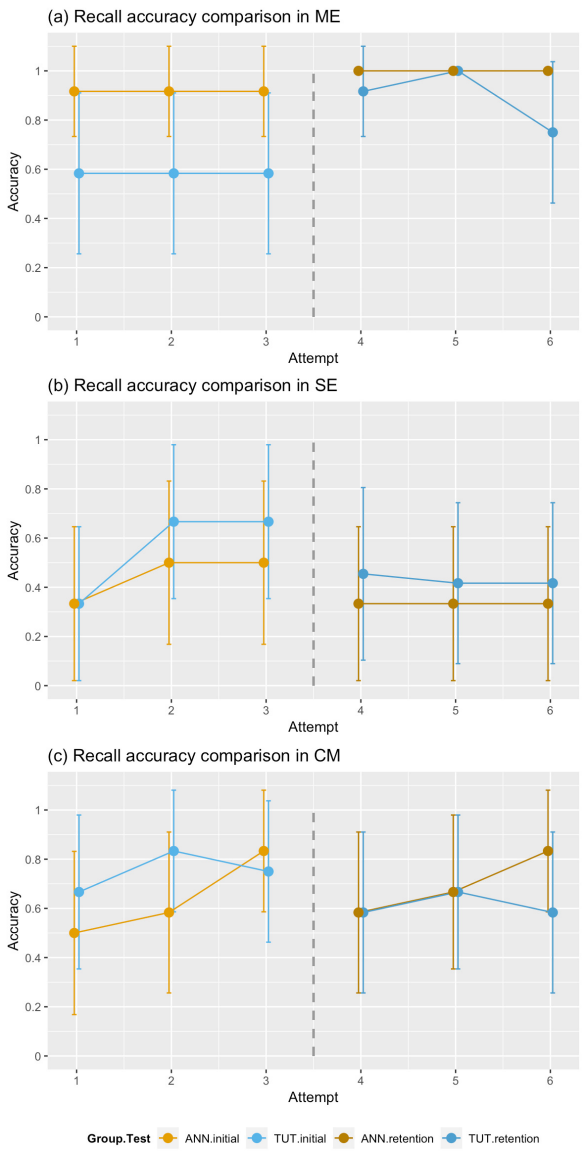

Fig. 9: Recall accuracy results with a 95% of confidence interval.

[8] P. Carlson, A. Peters, S. B. Gilbert, J. M. Vance, and A. Luse. Virtual training: Learning transfer of assembly tasks. *IEEE transactions on visualization and computer graphics*, 21(6):770–782, 2015.

[9] D. N. Cassenti. *Advances in Human Factors in Simulation and Modeling: Proceedings of the AHFE 2018 International Conferences on Human Factors and Simulation and Digital Human Modeling and Applied Optimization, Held on July 21–25, 2018, in Loews Sapphire Falls Resort at Universal Studios, Orlando, Florida, USA*, vol. 780. Springer, 2018.

[10] A. Covaci, A.-H. Olivier, and F. Multon. Visual perspective and feedback guidance for vr free-throw training. *IEEE computer graphics and applications*, (5):55–65, 2015.

[11] C. Donalek, S. G. Djorgovski, A. Cioc, A. Wang, J. Zhang, E. Lawler, S. Yeh, A. Mahabal, M. Graham, A. Drake, et al. Immersive and collaborative data visualization using virtual reality platforms. In *Big Data (Big Data), 2014 IEEE International Conference on*, pp. 609–614. IEEE, 2014.

[12] B. I. Edwards, K. S. Bielawski, R. Prada, and A. D. Cheok. Haptic virtual reality and immersive learning for enhanced organic chemistry instruction. *Virtual Reality*, pp. 1–11, 2018.

[13] A. Ferracani, D. Pezzatini, J. Bianchini, G. Biscini, and A. Del Bimbo. Locomotion by natural gestures for immersive virtual environments. In *Proceedings of the 1st International Workshop on Multimedia Alternate Realities*, pp. 21–24. ACM, 2016.

[14] F. Ganier, C. Hoareau, and J. Tisseau. Evaluation of procedural learning transfer from a virtual environment to a real situation: a case study on tank maintenance training. *Ergonomics*, 57(6):828–843, 2014.

[15] M. Garau, M. Slater, D.-P. Pertaub, and S. Razzaque. The responses of people to virtual humans in an immersive virtual environment. *Presence: Teleoperators & Virtual Environments*, 14(1):104–116, 2005.

[16] S. Gillner and H. A. Mallot. Navigation and acquisition of spatial knowledge in a virtual maze. *Journal of Cognitive Neuroscience*, 10(4):445–463, 1998.

[17] G. Gorisse, O. Christmann, E. A. Amato, and S. Richir. First-and third-person perspectives in immersive virtual environments: Presence and performance analysis of embodied users. *Frontiers in Robotics and AI*, 4:33, 2017.

[18] K. Hassani, A. Nahvi, and A. Ahmadi. Design and implementation of an intelligent virtual environment for improving speaking and listening skills. *Interactive Learning Environments*, 24(1):252–271, 2016.

[19] C. Hoareau, R. Querrec, C. Buche, and F. Ganier. Evaluation of internal and external validity of a virtual environment for learning a long procedure.

International Journal of Human–Computer Interaction, 33(10):786–798, 2017.

[20] C. Houston-Wilson, J. M. Dunn, H. v. d. Mars, and J. McCubbin. The effect of peer tutors on motor performance in integrated physical education classes. *Adapted Physical Activity Quarterly*, 14(4):298–313, 1997.

[21] W. L. Johnson and J. Rickel. Steve: An animated pedagogical agent for procedural training in virtual environments. *ACM SIGART Bulletin*, 8(1-4):16–21, 1997.

[22] J. Jung and Y. J. Ahn. Effects of interface on procedural skill transfer in virtual training: Lifeboat launching operation study. *Computer Animation and Virtual Worlds*, 29(3-4):e1812, 2018.

[23] R. L. Lammers, M. Davenport, F. Korley, S. Griswold-Theodorson, M. T. Fitch, A. T. Narang, L. V. Evans, A. Gross, E. Rodriguez, K. L. Dodge, et al. Teaching and assessing procedural skills using simulation: metrics and methodology. *Academic Emergency Medicine*, 15(11):1079–1087, 2008.

[24] J. Lee, M. Cheon, S.-E. Moon, and J.-S. Lee. Peripersonal space in virtual reality: navigating 3d space with different perspectives. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pp. 207–208. ACM, 2016.

[25] P. Maclean and M. Cahillane. Maintaining a human touch in the design of virtual part-task trainers (vppt): Lessons from cognitive psychology and learning design.

[26] K. Mania, T. Troscianko, R. Hawkes, and A. Chalmers. Fidelity metrics for virtual environment simulations based on spatial memory awareness states. *Presence: Teleoperators & Virtual Environments*, 12(3):296–310, 2003.

[27] K. Mania, D. Wooldridge, M. Coxon, and A. Robinson. The effect of visual and interaction fidelity on spatial cognition in immersive virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 12(3):396–404, 2006.

[28] R. P. McMahan, D. A. Bowman, D. J. Zielinski, and R. B. Brady. Evaluating display fidelity and interaction fidelity in a virtual reality game. *IEEE transactions on visualization and computer graphics*, 18(4):626–633, 2012.

[29] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, vol. 2351, pp. 282–293. International Society for Optics and Photonics, 1995.

[30] M. Naef, O. Staadt, and M. Gross. Spatialized audio rendering for immersive virtual environments. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pp. 65–72. ACM, 2002.

[31] T. Okuda, E. Tanaka, and T. Kasai. A method for the correction of garbled words based on the levenshtein metric. *IEEE Transactions on Computers*, 100(2):172–178, 1976.

[32] S. Pick, B. Weyers, B. Hentschel, and T. W. Kuhlen. Design and evaluation of data annotation workflows for cave-like virtual environments. *IEEE Trans. Vis. Comput. Graph.*, 22(4):1452–1461, 2016.

[33] J. E. Pompeu, F. A. dos Santos Mendes, K. G. da Silva, A. M. Lobo, T. de Paula Oliveira, A. P. Zomignani, and M. E. P. Piemonte. Effect of nintendo wii™-based motor and cognitive training on activities of daily living in patients with parkinson's disease: a randomised clinical trial. *Physiotherapy*, 98(3):196–204, 2012.

[34] R. Radkowski, J. Herrema, and J. Oliver. Augmented reality-based manual assembly support with visual features for different degrees of difficulty. *International Journal of Human-Computer Interaction*, 31(5):337–349, 2015.

[35] E. D. Ragan, D. A. Bowman, R. Kopper, C. Stinson, S. Scerbo, and R. P. McMahan. Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. *IEEE transactions on visualization and computer graphics*, 21(7):794–807, 2015.

[36] J. W. Regian, W. L. Shebilske, and J. M. Monk. Virtual reality: An instructional medium for visual-spatial tasks. *Journal of Communication*, 42(4):136–149, 1992.

[37] L. T. Ritz and A. R. Buss. A framework for aligning instructional design strategies with affordances of cave immersive virtual reality systems. *TechTrends*, 60(6):549–556, 2016.

[38] O. C. Santos. Training the body: The potential of aied to support personalized motor skills learning. *International Journal of Artificial Intelligence in Education*, 26(2):730–755, 2016.

[39] R. Skarbez, S. Neyret, F. P. Brooks, M. Slater, and M. C. Whitton. A psychophysical experiment regarding components of the plausibility illusion. *IEEE transactions on visualization and computer graphics*, 23(4):1369–1378, 2017.

[40] S. M. Smith and E. Vela. Environmental context-dependent memory: A review and meta-analysis. *Psychonomic bulletin & review*, 8(2):203–220, 2001.

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

[42] B. Volmer, J. Baumeister, S. Von Itzstein, I. Bornkessel-Schlesewsky, M. Schlesewsky, M. Billinghurst, and B. H. Thomas. A comparison of predictive spatial augmented reality cues for procedural tasks. *IEEE transactions on visualization and computer graphics*, 24(11):2846–2856, 2018.

[43] D. Waller, E. Hunt, and D. Knapp. The transfer of spatial knowledge in virtual environment training. *Presence*, 7(2):129–143, 1998.

[44] M. Waterman. Sequence alignments. mathematical methods for dna sequences, waterman ms ed, 1989.

[45] D. C. Wightman and G. Lintern. Part-task training for tracking and manual control. *Human Factors*, 27(3):267–283, 1985.