The Effect of Interaction Models on Multi-user Usability of Collaborative Virtual Environments

Aida Erfanian and Yaoping Hu

Department of Electrical and Computer Engineering
University of Calgary
Calgary, Alberta, CANADA
{aerfania@ucalgary.ca; huy@ucalgary.ca}

Abstract— **Multi-user usability of collaborative virtual environments (VEs) require the consideration of users' socio-human needs. However, most investigations of the usability have focused on either the technologies involved or individual user experience. Few have examined the effect of interaction models on multi-user usability, by taking account the socio-human needs. Thus, we have undertaken a study on such effect for collaborative tasks, comparing two interaction models: first-come-first-serve (FCFS) and dynamic priority (DP). Departing largely from our previous verification of these models for conflict resolution, the current study assessed multi-user usability for achieving a collaborative goal in a quasi-practical scenario. Based on the ISO/IEC 25010:2011 standard on usability, our observations revealed that multiple users completed the collaborative tasks statistically more effective, efficient, and satisfactory under the DP model than under the FCFS model. These observations indicate a potential practicality of the DP model for multi-user VEs, in which experts perform collaborative tasks.**

Keywords— interaction models; multi-user usability; multi-user collaborative virtual environments; socio-human needs

I. INTRODUCTION

The usability of collaborative virtual environments (VEs) requires to consider the socio-human needs of multiple users (i.e., multi-user experience). For many years, usability has been one of the key metrics to evaluate systems [1]. However, recent investigations have revealed a disconnection between usability defined in software engineering and its counterpart used in human-computer interaction [2]. Some studies have proposed solutions to eliminate this disconnection by taking account user experience. Primarily, the solutions have considered user experience in single-user systems for individual task, such as path finding, localization, and navigation [3, 4].

A few investigations have been undertaken in VEs to address the usability of multi-user systems. Aiming to develop the technologies of real-time learning VEs for multiple users, studies have either assessed the usability of the VEs based on individual learning [5]; or proposed usability guidelines to support individual learning within the VEs [6]. These studies measure the usability of these VEs by considering individual experience in learning, rather than multi-user experience in collaborative work. Other studies have attempted to improve usability of multi-user tabletop systems [7, 8]. By analyzing system-logged data, these studies emphasize on evaluating the

effect of technologies on the usability. The metrics of the usability includes the mostly used regions on a tabletop surface [7, 8] and the number of interaction in a certain region of the surface [8]. Nevertheless, the evaluation disregards multi-user experience in two aspects. One aspect arises from the lack of examining user perceived needs in collaboration. Another aspect comes from the oversight of interaction models, which govern multi-user interaction with shared objects to impact multi-user experience. To our best knowledge, there is no study on how interaction models affect multi-user usability of collaborative VEs, in which users work together in real time.

Thus, we undertake this current study to investigate the effect of interaction models on multi-user usability of collaborative VEs. The investigation utilizes both system-logged data and user perceived needs. The interaction models in investigation are first-come-first serve (FCFS) and dynamic priority (DP), which come from our previous work [9]. This previous work has utilized a well-controlled scenario, in which multiple users have performed a same task simultaneously to create conflicts; and has evaluated user perception of equality in interaction for conflict resolution within collaborative VEs. The observations of the evaluation have demonstrated that the DP model yields significantly an equality in interaction among multiple users, compared to the FCFS model. Under the DP model, haptic (pertinent to the sense of touch) cues are more intuitive for each user to perceive his/her gaining of interaction than visual cues. In contrast, no similar observations exist under the FCFS model.

Departing significantly from this previous work [9], the current study employs a quasi-practical scenario with three unique attributes. First, each user performs a task in his/her own pace with various time lengths. Secondly, each user undertakes a task differing from that of other users. Lastly, all users collaborate together to achieve a common goal. Hence, the quasi-practical scenario mimics actual engineering decision-making meetings carried out by a group of experts in various disciplines. For evaluation, the current study focuses on multi-user usability of collaborative VEs, considering metrics of both software engineering and multi-user experience.

To measure the multi-user usability of collaborative VEs, we propose a metrics for the current study. Based on the ISO/IEC 25010:2011 standard [10] and the socio-human needs of multiple users, the metrics include usability factors such as the effectiveness of achieving the goal of collaborative tasks, the

efficiency in achieving the goal, the perceived equality in interaction, and the cognitive workload of the users. The latter two factors reflect degrees of satisfaction for meeting the socio-human needs. Our observations reveal that all users performed collaborative tasks statistically more effective, efficient, and satisfactory under the DP model than under the FCFS model.

This paper is organized as follows: Section II presents the definition of the metrics used for multi-user usability; Section III describes the quasi-practical scenario of collaboration by experts, with an emphasis on its difference from an actual scenario; Section IV depicts an empirical study on multi-user usability, including its methods, observed results and discussion; and Section V concludes the study.

II. METRICS OF MULTI-USER USABILITY

The ISO/IEC 25010:2011 standard [10] defines usability as: "degree to which a product or system can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use." Specifically, the standard classifies the usability into three factors: *effectiveness* (*i.e.*, completeness and accuracy with which users achieve specified goals); *efficiency* (*i.e.*, resources expended in relation to the completeness and accuracy with which users achieve goals); and *satisfaction* (*i.e.*, degrees to which user needs are satisfied when a product or system is used in a specified context of use. This includes attitudes towards the use of product).

On one hand, the standard does not specify any rules on how to measure each of these factors. On another hand, the definitions of these factors are generic enough for any single- or multi-user systems. That is, a metrics of measuring these factors needs to be specified for applications involved.

For the tasks carried out within a multi-user collaborative VE, we thus detailed a metrics of these usability factors. As defined above, *effectiveness* is directly related to the goal of tasks. Consequently, the metrics of measuring effectiveness requires a clear description of the goal. In our quasi-practical scenario within a multi-user collaborative VE, the goal is to perform the whole set of tasks by all users together. Although each user has a unique task to perform, the individual contribution of the unique task is vitally important to achieve the goal. That is, the achievement of the goal is determined by combining each user's tasks, measured at a level of completion and a degree of accuracy. When there arises an opportunity of interaction for collaboration, the users might not always attempt to work together because of the human nature. As results, the completeness of achieving the goal ranges from non-completion at all (0%) to the total completion (100%); and the accuracy of this achievement depends on user involvement in collaboration, varying between non-user involvement (0%) and all-user involvement (100%). The multiplication of the completeness with the accuracy yields a measure of effectiveness [11].

Efficiency measures the effectiveness of achieving the goal with respect to the resources used. One vital resource is the time that all users work together to achieve the goal of tasks [11], because time is always constrained in the modern world.

To measure *satisfaction*, we focus on *socio-human needs* in collaborative VEs. These needs include perceived equality in interaction and cognitive workload [12]. As well, the cognitive workload embraces users' attitudes towards the use of the VE.

Table I summarizes the metrics of usability that we define for this current study. In Section IV below, we detail the methods of collecting raw data and of processing the data for computing the metrics.

III. QUASI-PRACTICAL SCENARIO

Aiming at facilitating users to undertake collaboration within multi-user VEs, we designed a quasi-practical scenario to mimic a decision-making meeting of experts (*i.e.*, users who are specialized in different disciplines) in petroleum industries. For this design, we interviewed two graduate students and one postdoc fellow in petroleum engineering at the University of Calgary. The interviews yielded the process of examining a geological grid (a dataset) for the production of an oil/gas reservoir. Conventionally, the process involves engineering and scientific experts. Specialized in certain properties (data/ideas) of an oil/gas reservoir, each type of experts possesses very different knowledge and expertise from those had by other types of experts. In addition, the interviews revealed potential ways of interacting with a shared geological grid by the experts and of presenting/communicating their data/ideas for collaborative decision-making.

Derived from the interviews, we identified three essential types of experts – reservoir engineers, production engineers and geologists. They normally conduct decision-making meetings whenever a problem arises in a given oil/gas reservoir for production. During the meetings, they routinely need to figure out what aspects contribute to the departure of an actual production rate from its predicted counterpart. Thus, they collaborate with each other to represent their data/ideas on a shared geological grid, which represents the known properties of the oil/gas reservoir. In other words, the collaboration attempts to complete a map of whole known properties of the reservoir on the geological grid. While the map (a goal) is common among all experts, their individual tasks are uniquely different to form a whole set of tasks. Thus, it is very crucial for each expert to actively contribute his/her efforts to achieve this common goal.

With current VE technologies, the experts are able to view the geological grid of a given oil/gas reservoir as a three-dimensional (3D) stereoscopic object on a wall-sized screen. This shared object has not only geometric shape and size in three dimensions, but also data representing certain known properties of the reservoir. However, there are three drawbacks of conducting the decision-making meetings using the current VE

TABLE I. METRICS OF MEASURING USABILITY

Usability Factors	Metrics
Effectiveness	Completeness: Percentage of collaboration to achieve a goal by all users.
	Accuracy: Percentage of user participation in attempting his/her tasks.
Efficiency	Effectiveness measured per time unit (minute).
Satisfaction (Socio-human needs)	Perceived equality in interaction
	Cognitive workload

technologies. At first, only one expert who presents the geological grid dominates the interaction with the shared object. Thus, the VE of the geological grid is a single-user system. Secondly, the interaction with the shared object is limited to two forms – to rotate the object for view and to point to a particular region (*i.e.*, a group of grid cells) of interest. The limited forms of interaction restrict the consultation among the experts. Lastly, the input data of the experts are separately recorded from the grid, due to the limited forms of interaction. After the meetings, one expert (mostly a reservoir engineer) combines these recorded data to update the properties of the grid. That is, this expert makes the sole decision about the oil/gas reservoir. Consequently, these drawbacks result in the duplication of efforts which negatively affects the efficiency of such decision-making meetings [13].

To overcome these drawbacks, we created a multi-user VE, in which each user (expert) could interact with the shared object (a geological grid) by means of a haptic device. This device not only functioned as a 3D mouse for interaction, but also provided force feedback (as haptic cues) to the user's hand when the user gained the control of interaction with the shared object. The interaction had four essential forms such as rotating, navigating, pointing, and highlighting. These forms of interaction permitted each user to input his/her data/ideas for real-time updating the properties of the gird. The real-time update enables the multiple users (experts) to make the decision collaboratively. The utilization of haptic cues is primarily due to the observation that haptic cues are more intuitive for the users to perceive interaction than visual cues [9]. As well, we implemented two interaction models – FCFS and DP – for interaction among multiple users, as derived from our previous work [9]. Thus, these models could influence the multi-user usability of the collaborative VE for decision-making.

Following the process of conducting a decision-making meeting by the experts, we designed a scenario to mimic the process within our multi-user VE. In this scenario, we assigned each user, as one type of pseudo experts, with a unique list of tasks. Three types of pseudo experts resembled reservoir engineers, production engineers and geologists, respectively. By executing the whole set of tasks in all lists, the goal of these pseudo experts (as users) was to complete a map of known properties of the grid. This scenario was quasi-practical because of the following two aspects. One aspect was the use of pseudo experts as users instead of actual experts. Another aspect was the constraint of no verbal communication among the users. This constraint eliminated the role of verbal communication in multi-user usability, while we investigated the effect of the interaction models on such usability of collaborative VEs. Under the quasi-practical scenario, we conducted an empirical study for the investigation.

IV. EMPIRICAL STUDY

A. Participants

Thirty participants (17 males and 13 females with the average age of 25.27 ± 5.12 years old) took part in this study. They all were naïve to the purpose of the study. The participants, who were different from those of our previous work [9], formed 10 groups of three participants. In each group, the three participants functioned as pseudo experts of a reservoir engineer,
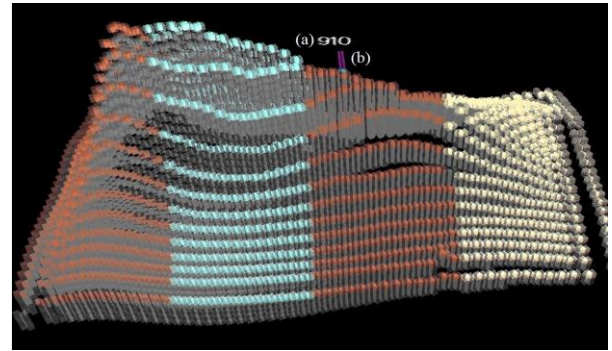

Figure 1. The shared geological grid partitioned into 4 regions: (a) a cell number to represent certain properties of the grid cell; and (b) a 3D cursor to correspond to the stylus motion of Omni device.

a product engineer and a geologist, respectively. The sample size of the 10 groups was greater than the minimal sample size (8) calculated by using the Lehr's formula [14]. All participants had normal or corrected-to-normal vision, and a stereo acuity of at least 40" of arc (determined by the Randot Stereo test). We applied the Ishihara color test for color blindness to all participants, and none of them was color blinded. The participants were right-handed, according to a modified version of the Edinburgh handedness inventory [15]. This study was approved under the Canadian Tri-Council Ethics guidelines.

B. Multi-user Collaborative VE

We developed a multi-user collaborative VE in C++ with OpenGL and OpenHaptics libraries for graphic and haptic rendering, respectively. As depicted in Fig. 1, the shared object for all participants was a geological grid of an oil/gas reservoir. This grid was partitioned into 4 different regions, visually separable by colors. Each grid cell had a unique number to indicate certain properties of the cell. A pair of stereoscopic goggles enabled each participant to view the shared grid in 3D stereoscopic vision. Due to the availability of haptic devices, we provided three PHANToM Omni devices (SensAble Technologies Inc., USA) to three participants in one group. Each Omni device had a stylus to function as a 3D mouse. The motion of the stylus was represented as a 3D cursor along with the grid. Via the stylus, the Omni device provided force feedback to the hand of one participant in 3 degrees of freedom.

As shown in Fig. 2, the VE presented the shared grid on a wall-sized screen to all participants in the group. Each participant used his/her right hand to manipulate the stylus of an Omni device to interact with the grid. The Omni device was aligned to the right hand of the participant. Each participant was required to rest his/her elbow on the arm of the seat for comfort. By means of the Omni stylus, each participant was able to perform all 4 essential forms of interaction with the grid. The forms of rotating and navigating were applied to the whole grid, whereas the forms of pointing and highlighting were executed on individual cells of the grid.

For operating this VE, a graphic computer had a 2.53GHz (dual quad core processors) Intel© Xeon© CPU, a Quadro FX 4800 NVidia® graphics card, a 4GB RAM and a 64-bit Windows® 7 operating system. The VE implementation used a multi-threading architecture: one thread was responsible for
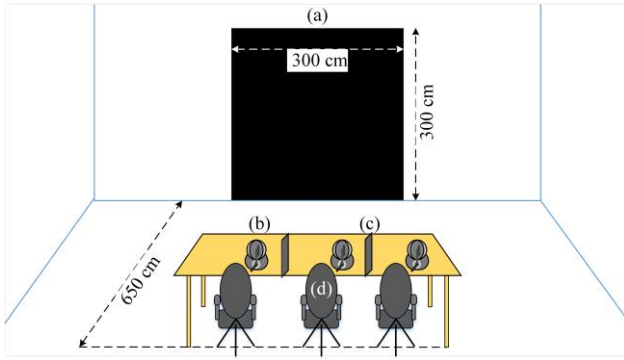
Figure 2. The setup of a VE: (a) a wall-sized display; (b) Omni devices; (c) vision blockers; and (d) participants.

visual rendering; and another thread managed interaction derived from the Omni devices. In the VE, all Omni devices could issue simultaneously interaction commands with the shared grid. Only one device, however, could gain the control of the interaction at one time instance.

C. Implementation of the Interaction Models

An interaction model is needed to handle simultaneous interaction commands issued by multiple participants. Thus, we implemented two interaction models – FCFS and DP – to govern simultaneous interaction with the shared grid.

All humans have different characteristics of visuomotor responses because of individual age, physical fitness and mental abilities. As a result, the humans would have different time lengths for the identical visuomotor response. Inevitably, some humans are agiler than others. Within the multi-user VE, the FCFS model encouraged "winner-takes-all" by allowing the agilest participant to gain the control of interaction, and the other participants to be deprived of interaction. Against "winner-takes-all", in contrast, the DP model promoted an equal opportunity of interaction. By taking account the historic interaction of each participant, the DP model computed a probability of his/her current interaction. On the basis of the probabilities of all participants, the DP model determined which participant should gain the control of interaction, if at least two participants were issuing interaction commands within an interval of 300 ms, as described in our previous work [9]. The interval was comparable to the human visuomotor response time [16]. Thus, the participants were unaware of a waiting gap to hinder their interaction. In our previous work [9], we detailed the definition and implementation of both interaction models.

D. Procedure

We carried out a within subject-design study in 9 blocks, including one practice block and 8 testing blocks. The practice block was prior to all testing blocks. In the practice block, all participants, who were randomly assigned to their tasks, had to meet three prerequisites necessary to become pseudo experts. For each participant, the prerequisites included three abilities of mastering interaction, collaborating with the peers, and acquiring assigned expertise. The ability of mastering interaction indicated that the participant could easily perform all forms of interaction by using an Omni device. The ability of collaborating with the peers meant that the participant could

observe and incorporate the activities of the peers on the display to assist his/her next task. The ability of acquiring assigned expertise revealed that the participant understood the organization pattern of grid cells for his/her designated tasks. The accomplishment of these abilities was measured using a logged task length – that is, each participant was able to complete a designated task in about 30 seconds. The practice block was terminated, when all participants completed 7 or 8 tasks consecutively in less than 30 seconds per task. The thresholds of the measurement and termination were from a pilot study for verifying the setup of the study. Each participant learned to fill out a questionnaire at the end of the practice block.

On the shared geological grid, the grid cells of a given region were colored to activate the region for a testing block. Each testing block consisted of 21 tasks, and each participant was encouraged to complete 1/3 of these tasks. At the end of a testing block, each participant filled out an identical questionnaire for the block. There was no time constraint for each task. Thus, the number of completed tasks varied in each testing block, because some participants performed their tasks faster than others. However, the length of each testing block was limited consistently in 5 minutes, plus 3 to 5 minutes of distributing, filling out and collecting the questionnaire.

All 8 testing blocks were divided into two sessions of four blocks. The two sessions corresponded to the two interaction models, respectively. The order of these sessions was counter-balanced for all participant groups. Each session lasted about 40 minutes, including some short breaks. For humans, the 40-minute length is a maximum upper boundary of the sustained concentration, which is revivable after a break of few minutes [17]. As a result, all testing blocks of the two sessions lasted about one and a half hours. The total length of the study for each participant group was within 2 hours, including the time needed for pre-screening the vision and handedness of each participant, the time spent in the practice and all testing blocks. This 2-hour length was roughly equal to a regular length for a decision-making meeting of experts in practice.

In a testing block, each participant of the group was given a unique list of designated tasks. The list contained the numbers of grid cells, which were located within the activated region of the grid. For completing the designated task, a participant performed certain combination of all 4 interaction forms to locate the cell in sequence and to highlight it. The cell numbers in the list were ordered pseudo-randomly, and differed from one testing block to another.

Following the given list, a participant carried out a designated task when taking the control of interaction with the shared grid. Before this control occurred, there was a transition state from one task to another. A signal light appeared on the display to indicate an opportunity of interaction. During this transition state, the signal light was in yellow for 3 seconds and then turned from yellow into green. Each participant was encouraged to press the dark grey button on his/her Omni stylus, as soon as he/she saw the green light. Although all participants of a group could initiate simultaneously their commands to gain the control of interaction, only one of them was actually chosen by the governing interaction model that was implemented in the block. As a haptic cue, a trapezoidal force feedback at maximum

0.5 N lasted for three seconds and was felt by the participant who gained the control of interaction. The implementation of the haptic cue was detailed in our previous work [9].

While the task was in progress, the signal light disappeared and other participants had to wait until the re-appearance of the signal light in yellow, signaling the completion of the task. This yellow light allowed all participants of the group to get ready for the next task on their lists. If all three participants missed to press the dark grey button on their Omni styluses following the green light for 500 ms, the signal light flashed in yellow again to indicate a new opportunity of gaining the control. The participants were instructed to accomplish the tasks on their lists as much as they could. This encouraged them to actively participate in achieve a common goal, reflecting a decision-making meeting of experts in practice.

For each group, the goal of the participants was to collaboratively complete a map on the shared grid in each session. To complete the map, the participants had to highlight K cells within 4 testing blocks of the FCFS session or the DP session. We set K equal to 28, determined by running a pilot study of each testing block during the setup of the current study. Since the participants were considered as pseudo experts to contribute equally to achieve the goal, each participant might accomplish at least 7 tasks in a testing block. This number of tasks per participant is much greater than 4 – the human capacity of short-term memory [18, 19]. By completing one session of 4 testing blocks, all participants of the group undertook 84 tasks to achieve the goal at 100% completeness.

E. Data Collection and Analyses

Two types of data were collected for each group of three participants: objective data and subjective data. The objective data was logged in files by the software application of the VE. The objective data collected information about every executed task including: the status of the dark grey button on an Omni stylus; the time instance that a participant tried to gain the control of interaction; the priority of each participant in the session governed by the DP model; and the identification number of the participant who gained the control for a task.

The subjective data was collected using questionnaires. An identical questionnaire was distributed to the participants after each testing block. This questionnaire captured the participant perception of equality in interaction and cognitive workload. We converted the answers solicited by the questionnaire into numeric for data analyses. This questionnaire and data conversion were same as those used in our previous work [9].

The analyses of both objective and subjective data consisted of four steps. The first step (Step 1) was a preliminary analysis on the objective data to verify whether there was any outlier among the participants of each group. The last three steps corresponded to examine the usability factors of effectiveness (Step 2), efficiency (Step 3) and satisfactory (socio-human needs, Step 4), as indicated in Table I. For analyses, we used the objective data in Steps 2 and 3; and the subjective data in Step 4. In addition, we used the objective data to compute a standard error, against which we verified the analysis of the perceived equality in interaction for Step 4.

In each step, we executed the statistical method of analysis of variance (ANOVA) [20] to compare each metrics of a usability factor between the FCFS session and the DP session. As well, the same analysis was performed for the factor. The ANOVA method was in one-way repeated measures and within subject-design. Before an ANOVA analysis, we conducted a normality analysis (normal probability density function [21]) on the metrics (or factor) to be compared. The normality analysis ensured that, for a testing block, the distribution of the metrics (or factor) was in normal distribution and thus is suitable for applying the ANOVA analysis.

Effectiveness: As defined in Table I, the usability factor of effectiveness is composed of two metrics: completeness and accuracy. To show how complete and accurate for each particpant group to achieve their goal, these two metrics were measured separately following the general guidelines given by Sandom and Harvey [11]. However, the guidelines do not specify how to measure the effectiveness for collaborative tasks. Therefore, we derived the mathematical description (equations) of effectiveness based on the guidelines and the metrics of usability in Section II.

For each participant of a group, the completeness of achieving the goal was calculated as:

$$C_{IM,i,j} = (\sum_{k=1}^{M} T_{i,j,k})/K \ , \qquad (1)$$

where $IM \in \{FCFS, DP\}$ and $C_{IM,i,j}$ is the completeness of the goal achievement by i-th participant of the j-th group in the M testing blocks of a FCFS session or a DP session. The parameter $T_{i,j,k}$ is the number of tasks accomplished by i-th participant of the j-th group in the k-th block of a session. K stands for the total number of tasks that each participant was supposed to accomplish within the M testing blocks of a session. As mentioned earlier, we set K equal to 28; and M represents the 4 testing blocks of each session.

For each group of multiple participants, we measured the completeness of achieving the goal following the equation:

$$C_{IM,j} = (\sum_{i=1}^{N} C_{IM,i,j})/N \ , \qquad (2)$$

where $C_{IM,j}$ is the average completeness for all $N\ (= 3)$ participants in j-th group for all M testing blocks of a session.

To measure the accuracy, we calculated the percentage of times that each participant of a group pressed the dark grey button of the Omni stylus within a 500-ms interval after the green signal light. This pressing of the button indicated an active attempt of gaining the control of interaction. If a participant did not press the button in one task, the others had a higher chance of gaining the control. We used the following equation to measure the accuracy:

$$A_{IM,j} = \sum_{i=1}^{N} (UA_{IM,i,j}/K_{IM,j})/N \ , \qquad (3)$$

where $A_{IM,j}$ is the average accuracy of achieving the goal for the j-th group during all testing blocks of the FCFS session or the DP session; $UA_{IM,i,j}$ is the number of the times that the i-th participant of the j-th group tried to gain the control; $K_{IM,j}$ is the

number of tasks that the *j*-th group accomplished in all blocks of the FCFS session or the DP session; and $N \, (= 3)$ is the number of the participants in the *j*-th group.

It is known that the effectiveness of achieving the goal is a multiplication, $f\left(C_{IM,j}, A_{IM,j}\right)$, of the completeness and the accuracy [11]. In other words, the more the participants are inactive, the less the effectiveness would be to achieve the goal. Therefore, we measured the effectiveness of each group using the following equation [11]:

$$E_{IM,j} = f\left(C_{IM,j}, A_{IM,j}\right) = \ C_{IM,j} \times A_{IM,j} \quad , \qquad (4)$$

where $E_{IM,j}$ is the effectiveness of the FCFS session or the DP session for the *j*-th group. $C_{IM,j}$ and $A_{IM,j}$ are the completeness and accuracy, respectively, for the *j*-th group.

Efficiency: We measured the usability factor of efficiency by calculating the effectiveness per unit of time [11]. We considered a minute as the unit, because a session was completed in about 40 minutes. The efficiency was derived using the following equation [11]:

$$Eff_{IM,j} = E_{IM,j}/T_{IM,j} \ , \qquad (5)$$

where $Eff_{IM,j}$ is the efficiency of achieving the goal for the *j*-th group, $E_{IM,j}$ is the effectiveness for the *j*-th group, and $T_{IM,j}$ is the time spent (in minutes) for the *M* blocks of a session. Because the time length of each testing block was constrained in 5 minutes, the time spent for all 4 blocks of the FCFS session or the DP session was a total of 20 minutes. This time length was equal for all participant groups.

Satisfaction (Socio-human Needs): We performed analysis on subjective data for both metrics of satisfaction: perceived equality in interaction and cognitive workload. The data treatment for both metrics were the same as those in our previous work [9]. In short, the treatment for perceived equality in interaction had 3 phases on the numeric derived from the first part of the questionnaire. These phases were precondition analyses, analysis based on standard deviations, and analysis based on the highest and lowest bounds of the perceived percentages. The data treatment for cognitive workload was numeric resulted the second part of the questionnaire.

In the current study, nevertheless, the statistical analyses for both metrics differed from those in our previous work [9]. We performed one-way ANOVA analyses on each metrics between the FCFS session and the DP session (one way = interaction models), whereas the previous work applied two-way ANOVA analyses (two ways = cues x interaction models).

F. Results

In Step 1 (preliminary analysis on objective data), we observed that the total percentage of the interaction among three participants of each group was 100.0% for each testing blocks. This shows that there was no outlier among the participants.

In Step 2 (Effectiveness), the normality analyses verified that the completeness calculated from Eq. (2) were normally distributed in each testing block for all groups. The ANOVA
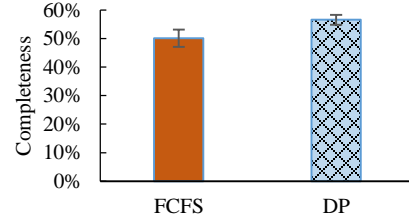

Figure 3. Average completeness of achieving the goal among all groups for both FCFS and DP sessions. [Error bars represent standard errors.]
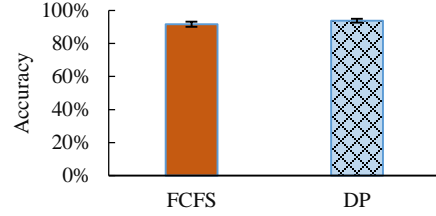

Figure 4. Average accuracy of achieving the goal among all groups for both FCFS and DP session. [Error bars represent standard errors.]
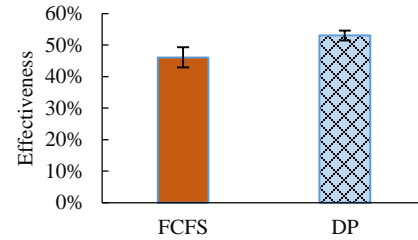

Figure 5. Average effectiveness of achieving the goal among all groups for both FCFS and DP sessions. [Error bars represent standard errors.]

analysis on the completeness revealed a significant difference between both interaction models [$F\,(1, 9) = 8.59; p < 0.05$]. As shown in Fig. 3, the average completeness among all groups for the FCFS session is about 7% less than that of the DP session. The length of error bars indicates that the completeness of the FCFS session is much more spread than that of the DP session.

The normality analyses on the accuracy derived from Eq. (3) confirmed the observation same as that on the completeness. However, there was no statistical significance of the accuracy between the interaction models [$F\,(1, 9) = 1.72; p > 0.05$]. Fig. 4 illustrates that the average accuracy among all groups is similar for the FCFS session (92%) and the DP session (94%). So do the samll error bars, inidicating the similar disparity of the accuracy for both interaction models. These observations confirmed that the participants of all groups performed their tasks in a consistent fashion in both sessions.

Not surprisingly, the normality analyses on the effectiveness agreed with what we observed for the completeness and the accuracy, as described above. The one-way ANOVA analysis on the effectiveness revealed a significant difference for the interaction models [$F\,(1, 9) = 6.11; p < 0.05$]. As depicted in Fig. 5, the average effectiveness among all groups were 46%
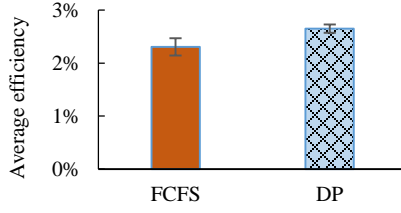
Figure 6. Average efficiency of achieving the goal among all groups for both FCFS and DP sessions. [Error bars represent standard errors.]

TABLE II.   USABILITY OF THE INTERACTION MODELS

Usability factor / Interaction model		FCFS	DP	Significant difference
Effectiveness		Low	High	yes
Efficiency		Low	High	yes
Satisfaction (Socio-human needs)	Perceived equality in interaction	Low	High	yes
	Cognitive workload	No differentiation	No differentiation	no

and 53% for the FCFS session and the DP session, respectively. That is, the FCFS session is averagely 7% less effective than the DP session. As well, the error bars indicate a much wider spread for the FCFS session than for the DP sessions.

Again, the normality analyses in Step 3 (Efficiency) verified that the efficiency calculated by Eq. (5) were in a normal distribution in each testing block for all groups. Following these analyses, the one-way ANOVA analysis confirmed that there was a significant difference of efficiency for the interaction models [$F (1, 9) = 6.37$; $p < 0.05$]. As illustrated in Fig. 6, the average efficiency of the FCFS session (2.31%) is lower than that of the DP session (2.65%). Moreover, the error bars shows a larger spread of the efficiency for the FCFS session than that for the DP session.

The analyses in Step 4 follow the same sequence as that in our previous work [9]. In this current study, the results of all three phases were in agreement with those in the previous work. At first, the mean perceived percentages of gaining the control of interaction was 47.0% for the FCFS sessions and 49.0% for the DP session – being close to 50%. There was also no statistical significance, compared the perceived percentage between the interaction models [$F (1, 9) = 0.71$; $p > 0.05$]. Secondly, analyzing the standard deviations of these percentages revealed that the FCFS session had larger standard deviations (28%) than the DP session (15%), followed by a statistical significance for these standard deviations between the interaction models [$F (1, 9) = 41.16$; $p < 0.05$]. This observation confirms the same trend as the error bars presented above. Thirdly, the analysis of the standard deviations for the highest bounds (average + standard deviation) and lowest bounds (average − standard deviation) of these percentages revealed a significant difference between the interaction models: [$F (1, 9) = 20.68$; $p < 0.05$] for the highest bounds, and [$F (1, 9) = 15.62$; $p < 0.05$] for the lowest bounds. Lastly, the analysis of cognitive workload showed no significant difference between the interaction models [$F (1, 9) = 1.05$; $p > 0.05$]. The average cognitive workload among all groups was similar for the FCFS session (117.10) and the DP session (120.31). The error bars had also a similar disparity for both interaction models.

G. Discussion

Following the ISO/IEC 25010:2011 standard [10], we studied the multi-user usability of a collaborative VE. This study investigated the effect of two interaction models – FCFS and DP – on all usability factors: effectiveness, efficiency and satisfaction. The quantification of the effectiveness and efficiency followed the definitions by Sandom and Harvey [11]. In measuring the factor of satisfaction, we considered socio-

human needs in collaboration such as perceived equality in interaction and cognitive workload. This consideration attempts to bridge the gap of usability's definitions in both software engineering and human-computer interaction, as noticed by Law and Abrahão [2]. The bridge is built upon two pillars: interaction models for conflict resolution and socio-human needs in collaboration. Thus, the current study fills a void left by existing work on multi-user usability [5-8].

Specifically, we observed that multiple users in a quasi-practical scenario achieved their collaborative goal statistically less effective, efficient and satisfactory under the FCFS model than under the DP model, as summarized in Table II. This significant difference might be correlated to the characteristics of both interaction models for conflict resolution. The FCFS model promotes a solution of "winner-takes-all" – an unfairness in gaining the control of interaction among multiple users whose agilities are inevitably unequal. In contrast, the DP model provides a solution of an equality in interaction among these users. Moreover, this statistical difference is significant; when all users have a similar distribution of trying to gain interaction, as indicated by the insignificance of the task accuracy between both models. With an insignificant difference of cognitive workload between both interaction models, the DP model meets better socio-human needs in multi-user collaboration than the FCFS model. The consideration of socio-human needs, in turn, might contribute to the higher effectiveness and efficiency yielded by the DP model in multi-user collaboration, compared to by the FCFS model. Therefore, how an interaction model resolves conflicts for socio-human needs could impact greatly on achieving a common goal in a collaborative VE.

Although examining the effect of both FCFS and DP interaction models within a collaborative VE, two aspects make the large departure of this current study from our previous work [9]. One aspect is the distinct purposes of both current study and previous work. The purpose of the current study aims at understanding which of both models could be more facilitative to multi-user usability; whereas the purpose of our previous work targets on providing an equality in interaction for multi-user conflict resolution. Another aspect is the varying scenarios

of multi-user collaboration involved in this study and our previous work. The current study concerns the collaboration in a quasi-practical scenario, which mimics the decision-making meetings by industrial experts. In contrast, our previous work considers the collaboration in a well-controlled scenario, which provokes simultaneous interactive commands to be resolved.

Despite of these aspects, we observed one common outcome from both current study and previous work. That is, the DP model functions always better than the FCFS model for both multi-user usability and conflict resolution. This outcome might reflect the advantage of the DP model over its counterpart to meet socio-human needs, which are inevitably essential in multi-user collaboration. That is, the DP model is more usable than the FCFS model in scenarios, which mimic an actual engineering practice. This advantage might be caused by the fact that the DP model promotes achieving a common goal rather than showing competiveness, as driven by the FCFS model. Nevertheless, studies are needed to further investigate this cause-effect phenomenon.

Over the FCFS model, the advantage of the DP model is evident by a statistically significant difference at about 7% higher for effectiveness (in turn, efficiency) in the current study. Currently, it is insufficient to extrapolate how this number of difference could impact the effectiveness (efficiency) of multi-user usability in practice. This is mainly due to the following two aspects. At first, the time length of one session in the study was limited to about 40 minutes, due to the maximum upper boundary of the human sustainable concentration [17]. In practice, the actual decision-making meetings of experts usually last about one to two hours – roughly 2 times longer. Secondly, the elements of social human communication (for example, verbal communication, objects, and environments [22]) might also contribute to the multi-user usability. We have not considered these elements yet, because the current study serves as a necessary pre-requisite towards generalization. Further studies are needed for both aspects.

V. Conclusion

We studied the effect of interaction models on multi-user usability within collaborative VEs. The study was based upon a metrics, which is derived from the ISO/IEC 25010:2011 standard [10] to include socio-human needs. By comparing each factor of the multi-user usability, our preliminary results indicate that the DP model yielded statistically more effective, efficient and satisfactory in achieving the goal of collaboration in a quasi-practical scenario than the FCFS model. These results indicate a potential practicality of the DP model for multi-user VEs, in which experts perform collaborative tasks.

References

[1] F. E. Ritter, G. D. Baxter and E. F. Churchill, "User-centered systems design: A brief history," *Foundations for Designing User-Centered Systems*, Springer Verlag, 2014, pp. 33-54.

[2] E. L. Law and S. Abrahão, "Interplay between User Experience (UX) evaluation and system development*," Int. J. of Human-Computer Studies*, vol. 72, pp. 523-525, Jun. 2014.

[3] A. Følstad and K. Hornbæk, "Work-domain knowledge in usability evaluation: Experiences with Cooperative Usability Testing," *J. Syst. Software*, vol. 83, pp. 2019-2030, 2010.

[4] L. Reis, E. Duarte and F. Rebelo, "Main usability issues in using virtual environments for older population warning studies," *Design, User Experience, and Usability: User Experience in Novel Technological Environments*, Springer Verlag, vol. 8014, pp. 189-198, 2013.

[5] S. Minocha and A. J. Reeves, "Interaction design and usability of learning spaces in 3D multi-user virtual worlds," *Human Work Interaction Design: Usability in Social, Cultural and Organizational Contexts*, vol.316, pp. 157-167, 2010.

[6] I. Perera, C. Allison and A. Miller, "Policy considerations for managing 3D multi user learning environments – Achieving usability and trust for learning," *Proc. 6th ICVL*, Bucharest, Romania, Sept. 2011, pp. 106-112.

[7] P. Marshall, R. Morris, Y. Rogers, S. Kreitmayer and M. Davies, "Rethinking 'multi-user': An in-the-wild study of how groups approach a walk-up-and-use tabletop interface," *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, Vancouver, BC, Canada, May 2011, pp. 3033-3042.

[8] J. Rick, A. Harris, P. Marshall, R. Fleck, N. Yuill and Y. Rogers, "Children designing together on a multi-touch tabletop: An analysis of spatial orientation and user interactions," *Proc. of the 8th Int. Conf. on Interaction Design and Children*, Como, Italy, Jun. 2009, pp. 106-114.

[9] A. Erfanian, T. Zeng and Y. Hu, "Dynamic strategies of conflict resolution on human perception of equality within multi-user collaborative virtual environments," *Proc. of 9th Int. Conf. on Collaborative Computing: Networking, Applications and Worksharing*, Austin, TX, USA, Oct. 2013, pp. 363-370.

[10] B. BSI, *ISO/IEC 25010: 2011 Systems and software engineering— Systems and software Quality Requirements and Evaluation (SQuaRE)—System and software quality models,* 2011.

[11] C. Sandom and R. S. Harvey, *Human Factors for Engineers*, IET, 2004.

[12] N. Yuill and Y. Rogers, "Mechanisms for collaboration: A design and evaluation framework for multi-user interfaces," *ACM Trans. on Computer-Human Interaction*, vol. 19, pp. 1-25, 2012.

[13] K. S. Quick and M. S. Feldman, "Boundaries as junctures: Collaborative boundary work for building efficient resilience," *J. Public Administr. Research and Theory*, vol. 24, pp. 673-695, 2014.

[14] G. Van Belle, *Statistical Rules of Thumb*, John Wiley & Sons, 2011.

[15] R. C. Oldfield, "The assessment and analysis of handedness: the Edinburgh inventory," *Neuropsychologia*, vol. 9, pp. 97-113, 1971.

[16] A. Welford, "Choice reaction time: Basic concepts," *Reaction Times*, pp. 73-128, 1980.

[17] M. David Cornish, D. Cornish and D. Dukette, *The Essential 20: Twenty Components of an Excellent Health Care Team"*, Dorrance Publishing, 2009.

[18] N. Cowan, "The magical mystery four: How is working memory capacity limited, and why?" *Current Directions in Psychological Sci.*, vol. 19, pp. 51-57, Feb. 2010.

[19] G. A. Miller, "Human memory and the storage of information," *IRE Trans. On Information Theory*, vol. 2, pp. 129-137, Sep. 1956.

[20] K. Hinkelmann and O. Kempthorne, *Design and analysis of experiments: Introduction to experimental design*, Hoboken, 2008.

[21] C. Forbes, M. Evans, N. Hastings and B. Peacock, *Statistical Distributions*, John Wiley & Sons, 2011.

[22] D. Roberts, R. Wolff, O. Otto, D. Kranzlmueller, C. Anthes and A. Steed, "Supporting social human communication between distributed walk-in displays," Proc. of the ACM Symp.on Virtual Reality Software and Technology, Hong Kong, China, Nov. 2004, pp. 81-88.