Digital Object Identifier: 10.1109/THMS.2017.2700431

Framework of Multi-User Satisfaction for Assessing Interaction Models Within Collaborative Virtual Environments

Aida Erfanian, *Student Member, IEEE*, Yaoping Hu, *Member, IEEE*, and Tao Zeng

***Abstract*—Collaborative virtual environments (VEs) require interaction models for resolving conflicts and promoting multi-user collaboration. Common models, such as the first-come-first-serve (FCFS) model, which grants interaction opportunities to the most agile user, and the static priority model, which gives interaction opportunities to the user with the highest predefined priority, disregard the importance of perceiving equality in interaction (EII) among all users. One exception is the dynamic priority (DP) model, as proposed in our earlier work, which grants interaction opportunities to a user based on the recency of his/her gained opportunities. To date, few research efforts have investigated the effect of interaction models on multi-user satisfaction. This paper hence presents an assessment of the DP model's effect on multi-user satisfaction within a collaborative VE. We first verified that the DP model allowed multiple users to perceive EII. We then conducted an experiment to examine the effect of the DP and FCFS models on multi-user satisfaction under a quasi-practical scenario that mimicked a decision-making meeting of experts. The framework of the examination was based on several metrics, which we proposed for the components of the ISO/IEC 25010:2011 standard. This framework resolved issues with existing metrics that measure user satisfaction by analyzing individual experience, thus omitting EII desired by multiple users. The results of the experiment indicated that the DP model fulfilled the metrics of the framework significantly better than the FCFS model. This observation implies a potential application of the DP model in collaborative VEs where multi-user satisfaction is the key to productive collaboration.**

***Index Terms*— Framework of multi-user satisfaction, dynamic priority model, cognitive needs in collaborative work, collaborative virtual environments**

I. Introduction

MULTI-USER virtual environments (VEs) have a potential to promote collaboration in a group work of experts (e.g., users). In the oil/gas industry, for example, experts of various disciplines routinely work together to view and manipulate shared 3D objects (e.g., geological and geophysical information) on a computer-based large display for communicating their ideas and solutions. Such multi-user collaborative work enriches the understanding of an engineering problem [1]. Currently, there are two typical settings for collaborative work: distributed and co-located. Over networks, a distributed setting encounters issues of network security and little awareness among users in disparate locations [2]. Targeting awareness among users, research activities attempt developing techniques of projecting the avatars of remote users to be co-located with other users [3]. In contrast, a co-located setting requires all users to be physically proximate, yielding mutual awareness among the users. The awareness establishes a foundation toward user satisfaction — an essential factor to promote genuine collaboration [2]. However, few multi-user VEs in practice elevate user satisfaction for co-located collaborative work. This results largely from an inconsideration of human cognitive needs (such as equality) in designing interaction models for multi-user collaborative VEs.

An interaction model is a policy to coordinate interactive commands issued by multiple users when they intend to interact with a shared object. A conflict occurs inevitably when the users attempt to gain the interaction with the object simultaneously. The interaction model resolves the conflict by granting one of the users an access to the shared object. In existing collaborative VEs, interaction models lack the full capacity of conflict resolution [4]. Many models devote to conflict avoidance by assigning each user a distinctive, static, or dynamic region of interaction [5]–[7]. When a conflict arises, the models impose social protocols (behavioral policies among the users) to partially resolve the conflict. This partial solution is generally ineffective due to the overhead of imposing the policies [8]. For conflict resolution, common models are the first-come-first-serve (FCFS) model and the static priority (SP) model [9]. The FCFS model always grants the agilest user an access to the shared object. The SP model assigns the access to the user with a predefined priority higher than other users or applies the FCFS model to determine the access among users with a same priority. Thus, both the FCFS and SP models promote an inequality of interaction among the users. In general, both these models suffer from a deficiency of

Manuscript submitted October XX, 2015. This work was supported by the NSERC grants (STPGP/364847-2008, STPGP 478870) to Y. Hu.

A. Erfanian and Y. Hu are with the Department of Electrical and Computer Engineering, Schulich School of Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada (e-mail: aerfania@ucalgary.ca; huy@ucalgary.ca).

T. Zeng was with the University of Calgary, Calgary, AB, Canada. He is now with the Department of Electrical Engineering, Xiamen University, Xiamen, Fujian, China (e-mail: tao.zeng@xmu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

disregarding cognitive needs among the users for collaborative work.

Such deficiency usually leads to dissatisfaction among the users, causing an inefficiency of collaborative work [4][8]–[10]. An inefficiency of collaborative work is a key concern of many businesses [11]. To overcome the deficiency, we proposed a novel interaction model, the dynamic priority (DP) model in our earlier work [12], to offer all users equal opportunities in interaction. The perception of equality in interaction (EII) is a vital cognitive need [13]. This need is crucial to underlie user satisfaction in collaborative work [14]. However, few research efforts have investigated how the perception of EII affects multi-user satisfaction within collaborative VEs. Some metrics exist to measure user satisfaction by analyzing individual experience in single-user tasks [15]–[19]. These metrics, in general, omit cognitive needs among multiple users of a group, who undertake a collaborative work. Thus, a comprehensive set of metrics to encompass the cognitive needs of multiple users is imperative for measuring multi-user satisfaction in collaborative VEs.

Considering cognitive needs, our earlier work evaluated users' perceived EII for conflict resolution within a collaborative VE. This work used a well-controlled scenario, in which multiple users employed identical devices to interact with a shared object simultaneously. The evaluation revealed that the DP model yielded significantly perceived EII among multiple users compared to the FCFS model. Under the DP model, haptic (pertinent to the sense of touch) cues were more intuitive for each user to perceive his/her gaining of interaction than visual cues. These findings were accompanied with similar levels of perceived workload under both models. Nevertheless, this earlier work did not investigate the role of the DP model in underlying multi-user satisfaction within a collaborative VE.

In this paper, we propose a framework of multi-user satisfaction and present an experiment on how the DP model affects multi-user satisfaction within a collaborative VE. The framework incorporates both the ISO/IEC 25010:2011 standard [19] and the cognitive needs of multiple users. For the experiment, we verify at first that the DP model offers perceived EII even if users employ heterogeneous devices (such as a mouse, a haptic device, etc.) for their interaction. The verification and our earlier work serve as two-pillared prerequisites of the experiment. Under the proposed framework of multi-user satisfaction, the experiment compares both the DP and FCFS models to affect multi-user satisfaction in a quasi-practical scenario [20], which mimics a decision-making meeting of experts (i.e., users who are specialized in different disciplines) in industrial settings. This experiment aims to provide an interaction model for improving multi-user satisfaction within a VE, in which experts undertake their collaborations.

II. Framework of Multi-User Satisfaction

The ISO/IEC 25010:2011 standard [19] defines user satisfaction as "degrees to which user needs are satisfied when a product or system is used in a specified context of use." The standard classifies user satisfaction into four factors, *trust*, *usefulness*, *pleasure*, and *comfort*, but is unspecific about metrics to assess each factor. The elucidation of the metrics depends herein upon the particular context of use for a system. Various metrics exist to evaluate user satisfaction about collaborative systems [16]–[18]. These metrics are commonly based on subjective data, which are acquired through questionnaires to capture the users' perception of using a system. Subjective data might be unreliable due to individual variations in interpreting the questionnaires [21]. A remedy is to use objective data, which are logged by the system. Although recording the behaviors of a system and its users at certain degrees, the logging could not fully capture the users' perception. Combining subjective and objective data is thus necessary to increase the reliability and to capture the perception [21].

TABLE I.
Framework of Multi-User Satisfaction

Factors	Metrics	Definitions	Analysis
Trust	Real-time response	System response to users' interactive commands without invoking their notice of needed processing time	Eq. (1)
	Simultaneous interaction	Capability of users' interaction with a shared object	Eq. (2)
	Conflict resolution	Treatment of interactive commands issued simultaneously by users	Eq. (3)
Usefulness	Task focus	*TF1*: Degree of task completion by all users	Eq. (4)
		TF2: Degree of user participation in all interaction opportunities	Eq. (5)
	Decision time	Time used to reach a common goal	Eq. (6)
	Consensus	Degree of similarity in the task behavior of all users to reach a common goal	Eq. (7)
Pleasure	Equality in interaction	Perception of having equal opportunities in collaboration	Eq. (8) Eq. (9)
Comfort	Perceived Workload	Perception of workload in Collaboration	Eq. (10)

Few reports on collaborative VEs have measured multi-user satisfaction by considering all factors. For developing and evaluating interaction models used in collaborative VEs, there is also a lack of comprehensive metrics to assess multi-user satisfaction based on the standard. Focusing on EII (a vital cognitive need) in collaborative VEs, we thus defined a set of metrics for each factor as detailed in Table I. The metrics for *trust* evaluate whether a multi-user collaborative VE behaves as intended (i.e., the system's behaviors). The metrics for *usefulness* reveal at what level the pragmatic aims of using the VE are achieved by multiple users (i.e., the users' behaviors). The metrics for both *trust* and *usefulness* rely on objective data

for reliability. In contrast, the metrics for both *pleasure* and *comfort* could apply subjective data to capture users' perception. Considering cognitive needs, the metrics for *pleasure* indicate whether multiple users perceive EII during their collaboration. The metrics for *comfort* yield how the users perceive workload using the VE. These metrics need to minimize inconsistency due to individual variations.

For *trust*, three metrics are commonly used to assess intended behaviors of a multi-user collaborative VE. These metrics are real-time response (*RTR*), simultaneous interaction (*SI*), and conflict resolution (*CR*) [4][22]. *RTR* is a general requirement for interactive systems [22]. *SI* and *CR* are specific requirements for interactive systems to support multi-user collaboration [4]. The fulfillment of these requirements is measurable and paves a foundation of user satisfaction [22].

The metrics of *RTR* determines an upper boundary of system response time, $R(t)$, to users' initiations of interactive commands. Both *RTR* and $R(t)$ have the following relationship:

$$R(t) \leq RTR \quad . \tag{1}$$

Within this boundary, all users of a collaborative group shall not observe a delay between initiating an interactive command and enacting the interaction under the command.

The metrics of *SI* is a system capacity of permitting multiple users to initiate their interactive commands at the same time. Using the logged data of interactive commands, this metrics is fulfilled when at least two command initiations occur among a group of *N* users for an interaction opportunity (e.g., *o*th interaction opportunity). An interaction opportunity is signaled to the users by a collaborative VE. That is,

$$\forall o \ [(SI = \textstyle\sum_{k=1}^{N} ICM_k) \geq 2] \ , \quad ICM_k \in \{0, 1\}; \tag{2}$$

where ICM_k is a binary value of 0 or 1 to register the *k*th user's ($k \in [1, N]$) status of initiating an interactive command at *o*th interaction opportunity. If the *k*th user has initiated an interactive command, ICM_k is 1; otherwise, ICM_k is 0.

The metrics of *CR* needs a policy (i.e., an interaction model) to coordinate interactive commands issued by multiple users. Given an interaction opportunity, the policy grants one user to access a shared object when the system logs at least two command initiations:

$$\{\forall o \wedge \forall \ (\textstyle\sum_{k=1}^{N} ICM_k) \geq 2\} \ [CR = \ \textstyle\sum_{k=1}^{N} EA_k = 1], \quad EA_k \in \{0, 1\} \ ; \tag{3}$$

where EA_k is a binary value of 0 or 1 to represent the *k*th user's status of gaining the exclusive access. If the *k*th user gains the access, EA_k is 1; otherwise, EA_k is 0. All metrics of *RTR*, *SI*, and *CR* shall accord with each other, although they are assessed independently.

For *usefulness*, three metrics of task focus (*TF*), decision time (*DT*), and consensus (*CS*) are the pragmatic aims of using the collaborative VE by multiple users [25]. The metrics of *TF* is measured by a degree of task completion by all users [2], *TF1*, and a degree of user participation in all interaction opportunities [23], *TF2*, as

$$TF1 = TCT/(\textstyle\sum_{k=1}^{N} \overline{TCT}_k/N) \ , \tag{4}$$

$$TF2 = \textstyle\sum_{k=1}^{N} (OPP_k/OPP)/N \ ; \tag{5}$$

where *TCT* is the task completion time by a collaborative group of *N* users, $\overline{TCT}_k$ is the average task completion time by the *k*th user of the group, OPP_k is the number of interaction opportunities that the *k*th user participated in, and *OPP* is the total number of interaction opportunities. The VE determines the *k*th user's task completion. Because *TF* measures the conjunction of both task completion (i.e., *TF1*) and user participation (i.e., *TF2*), a high degree of *TF* should reflect high degrees of both *TF1* and *TF2*. Thus, TF is expressed as $TF = TF1 \wedge TF2$.

The metrics of *DT* is the time that the group uses to achieve a common goal [24]. This metrics is an average of two components, *D1* and *D2*, as indicated below:

$$D1 = \Delta t/PG \ , \quad D2 = \Delta t + \textstyle\sum_{k=1}^{N} (\overline{TCT}_k \times RT_k) \ ,$$
$$RT_k = \begin{cases} T_k - TA_k, & T_k \geq TA_k \\ 0 & , \ T_k < TA_k \end{cases} \ , \tag{6}$$
$$DT = Average(D1, D2) \ ;$$

where Δt is the time length of a collaborative session, *PG* is the percentage of tasks accomplished toward the common goal, $\overline{TCT}_k$ is the same parameter used in Eq. (4), RT_k is the number of remaining tasks to be accomplished by the *k*th user, T_k is the number of tasks for the *k*th user to complete, and TA_k is the number of tasks that the *k*th user has accomplished. That is, RT_k is the difference between T_k and TA_k. RT_k is set to zero if the *k*th user accomplishes a greater number of tasks than needed by T_k. In other words, there are no remaining tasks for the *k*th user to complete. A smaller value of *DT* indicates shorter time used to achieve the goal.

The metrics of *CS* is a degree of agreement among the behaviors of all *N* users in the group [26], as measured by

$$CS = 1 - (\textstyle\sum_{k=1}^{N} (|T_k - TA_k|/T_k))/N, \tag{7}$$

where the absolute difference between T_k and TA_k covers two behavioral situations of the *k*th user. In one situation, the user has accomplished a fewer number of tasks than T_k. In another situation, the user accomplishes a greater number of tasks than T_k. Both situations cause disagreement among the behaviors of all users, leading to the decrease in the degree of consensus. Expressed in percentage, a higher degree of *CS* denotes more consensus among the users [26]. A higher degree of *usefulness* is thus associated with higher *TF*, lower *DT*, and higher *CS*, although they are assessed independently.

The metrics for both *pleasure* and *comfort* focus on cognitive needs among multiple users to undertake collaborative work. The metrics used to assess *pleasure* is *EII*, because it promotes content and delight among users to underlie user satisfaction [14]. Commonly, Likert scales or variations are used to record users' subjective responses to questions [27]. To obtain consistency in such responses, the questions should relate directly to the users' actions and be constructive [21]. An

example of the questions is to solicit a percentage of interaction gained by each user during the users' collaboration. The solicitation shall be constructive enough to derive EII perceived by the users. Given that the *k*th user indicates a percentage of his/her gained interaction, $X_{PIn,k}$, during a collaborative session, the standard deviation of the perceived percentages among all users, $SD(X_{PIn})$, is as follows:

$$\mu_{PIn} = \sum_{k=1}^{N} X_{PIn,k}/N \ ,$$
$$SD(X_{PIn}) = \sqrt{\frac{1}{N-1}\left[\sum_{k=1}^{N}\left(X_{PIn,k} - \mu_{PIn}\right)^2\right]} \ ; \tag{8}$$

where μ_{PIn} is the mean of perceived percentages of interaction among all users. The standard deviation, $SD(X_{PIn})$, represents a dispersion from this mean. A smaller standard deviation represents a narrower dispersion of the perceived percentage of interaction, indicating a better trend of EII. Thus, *EII* is inversely related to $SD\ (X_{PIn})$ as

$$EII = 1/SD(X_{PIn}) \ . \tag{9}$$

Eq. (9) agrees well with approaches commonly used for assessing equality in economics, such as the coefficient of variation and Gini index [28].

The metrics for assessing *comfort* is perceived workload (*PW*) [29]. This metrics of *PW* embraces users' attitudes toward their use of the VE, considering both physical and cognitive efforts. The most common method of evaluating this metrics is the NASA Task Load Index (TLX) [30]. The following equation depicts the computation of *PW*:

$$PW = \left(\sum_{k=1}^{N} PW_k\right)/N \ ,$$
$$PW_k = \sum_{f=1}^{6} WF_{f,k} = \sum_{f=1}^{6}\left(SR_{f,k}/RI_{f,k}\right) \ ; \tag{10}$$

where PW_k is the TLX score of the *k*th user; $WF_{f,k}$ is the score of the *k*th user for the *f*th component among six workload components (i.e., mental demand, physical demand, time demand, effort, frustration level, and performance); and $SR_{f,k}$ and $RI_{f,k}$ are the *k*th user's perceived subscale rating and relative importance of the *f*th component, respectively. The individual variations of *PW* are minimized in Eq. (10) because of using $RI_{f,k}$ to reflect the *k*th user's weighting on the *f*th component. The higher score of *PW* is, the more strenuous workload (i.e., less comfort) is for the *N* users.

In short, Table I describes a framework of measuring multi-user satisfaction within a collaborative system. This framework provides a foundation to our experiment of assessing interaction models for conflict resolution to affect multi-user satisfaction within a collaborative VE.

III. Multi-User Interaction models

In this paper, the interaction models for conflict resolution are the DP and FCFS models. Both models grant an exclusive access to a shared object among simultaneously issued interactive commands of multiple users. However, an understanding of the VE's domain properties is vital for implementing the models in our collaborative VE.

A. Domain properties

The multi-user collaborative VE has stochastic properties. Within the VE, users can initiate randomly their interactive commands to gain an access to a shared object. Once a user gains the access, it is stochastic for the user to complete his/her task of interacting with the object. That is, both command initiation and task completion can be characterized as Markovian processes [31]. Thus, we used a queue to hold and manage command initiations of all users by applying the following Kendall's notation [32]:

$$M/M/S/QC/N/IM \ , \tag{11}$$

where the first *M* depicts a Markovian process of the command initiations by the users at an interaction opportunity, the second *M* denotes a Markovian process of the task completion by the user who interacts with the shared object, the parameter *S* indicates the number of shared objects in the VE, the parameter *QC* describes the capacity of the queue to hold command initiations of *N* users, and the parameter *IM* represents an interaction model to select one of the command initiations from the queue. At an interaction opportunity, the queue holds all initiations in sequence. That is, the initiation of the agilest user arrives to the queue first and thus is placed at the front of the queue and so on. When an *IM* selects an initiation from the queue, the user of this initiation gains the exclusive access to a shared object for his/her task of interaction.

The prediction of queuing command initiations is necessary for implementing an *IM* in our collaborative VE. According to the first *M* in Eq. (11), the sequence of the command initiations is discrete and mostly predicted by a Poisson distribution [31]. That is, the probability, $\Pr(X)$, of a user to initiate a command after q other initiations already in the queue is given as

$$\Pr(X = q) = \lambda^q e^{-\lambda}/q! \ , \tag{12}$$

where X is a random variable representing the number of initiations in the queue and λ is the average initiation rate of all users. Similarly, the second *M* in Eq. (11) is representable by a continuous form of Eq. (12) for task completion, where □ denotes the average time of task completion by all users.

For the two-pillared prerequisites described in our earlier work [12] and in Section IV, we considered a well-controlled scenario. The scenario had a stochastic process of command initiation but a fixed (deterministic) process of task completion. The scenario becomes a special case of Eq. (11) using a *D* (deterministic) to replace the second *M*. Focusing on conflict resolution for multi-user satisfaction, the experiment in Section V used a quasi-practical scenario. The scenario was stochastic for both command initiation and task completion. The time length of a collaborative session was fixed to encompass a varying number of tasks. Hence, the quasi-practical scenario was represented by Eq. (11). In both scenarios, we took account Eq. (12) to queue command initiations at each interaction opportunity.

For the prerequisite verification in Section IV and the experiment in Section V, we specified the other parameters in Eq. (11) as $S = 1$ to inflict conflicts, $QC = N = 3$ to consider the availability of interaction devices, and $IM \in \{DP, FCFS\}$ to be interaction models. The following sub-sections describe the *IM* for conflict resolution among the initiations in the queue.

B. DP model

The DP model incorporates the individual variation of agility (i.e., manifestation of physical fitness, mental ability, willingness, and work pace) to respond to a visual stimuli displayed in the VE. This incorporation is characterized by an *x-interval*, which is extra to the average value of the human visuomotor response time (VRT). For all command initiations queued within the *x-interval*, the DP model considers them as simultaneous. This consideration gives the initiations the same probability as:

$$\Pr(X = 0) = e^{-\lambda}; \quad \text{for } q = 0 \quad . \tag{13}$$

Thus, the individual agility and the sequence of the initiations are irrelevant to determining the priority of the users — their probability of gaining exclusive access to the shared object.

Without penalty, the priority of the users is determined using the following computation. At the beginning of a collaborative session (at the 0th interaction opportunity), the DP model assigns the *k*th user a priority of $P(k, 0) = 1/N$. At the *o*th interaction opportunity, the priority of the *k*th user is dynamically updated based on his/her historical interaction. This update is formulated as

$$P(k, o) = \frac{PN(k, o)}{PD(k, o)} = \frac{PN(k, o-1) + f(k, o-1)}{[(N-1) \times o] + 1} \ ,$$

$$f(k, o-1) = \begin{cases} 0, & \text{gaining } (o-1)\text{th opportunity} \\ 1, & \text{otherwise} \end{cases} \ ; \tag{14}$$

where $PN(k, o)$ and $PD(k, o)$ are the numerator and denominator of the *k*th user's priority, respectively, and $f(k, o-1)$ is the update function based on the previous interaction opportunity. If the *k*th user gains the access to the shared object in the previous opportunity, the numerator $PN(k, o)$ of the current opportunity remains unchanged; otherwise, this numerator is incremented by 1. In the meantime, the denominator $PD(k, o)$ is updated to fulfill the condition as

$$\sum_{k=1}^{N} P(k, o) = 1 \ , \quad o \geq 1 \ . \tag{15}$$

Hence, the DP model selects the command initiation of the user with the highest priority among all entries of the queue.

C. FCFS model

Although being commonly implemented in a collaborative VE, the FCFS model is known to promote "winner-takes-all." The model favors the agilest user to enact the interaction with a shared object. That is, the FCFS model selects the command initiation at the front of the queue, for $q = 0$. This selected initiation has a probability represented in Eq. (13). The probabilities of other command initiations in the queue (for $q > 0$) are as follows:

$$\begin{aligned} \Pr(X > 0) = 1 - \Pr(X = 0) &= \textstyle\sum_{q=1}^{N-1} \Pr(X = q) \\ &= \lambda e^{-\lambda} + \frac{\lambda^2}{2} e^{-\lambda} + \cdots = e^{-\lambda}\left(\lambda + \frac{\lambda^2}{2} + \cdots\right) \\ &= \Pr(X = 0)\left(\lambda + \frac{\lambda^2}{2} + \cdots\right) . \end{aligned} \tag{16}$$

Comparing Eq. (16) to Eq. (13), the probability of the selected initiation is less than the probabilities of the unselected ones. That is, the higher the probability of a command initiation is, the lower is its chance of being selected by the FCFS model. Under the FCFS model, users with an averaged VRT (or below) have less chance of being granted access to a shared object.

For comparison, we used both the DP and FCFS models for conflict resolutions in the verification and the experiment described below.

IV. Test Environment and Prerequisite Verification

Two pillars are necessary as prerequisites to support our experiment described in Section V. One pillar is our earlier work [12] to evaluate the effect of the DP model on the perceived EII using homogeneous haptic devices, as summarized in Section I. Another pillar is a needed verification that this effect holds even if users employ heterogeneous devices (such as a mouse, a haptic devic, etc.) for their interaction. That is, the perceived EII under the DP model is independent of interaction devices in use. Thus, we undertook this verification.

A. Architecture

As illustrated in Fig. 1, we developed a five-layered and multi-threaded architecture for a collaborative VE. Within this VE, multiple co-located users interacted with a shared object for their collaboration. Although the VE incorporated three interaction devices due to their availability, the architecture of the VE is expandable to accommodate more devices as described in Section III. Within the interaction space, Fig. 1a depicts a heterogeneous setup of different devices for user interaction, whereas Fig. 1b illustrates a homogeneous setup of identical haptic devices. Its illustration here serves for clarity to compare to the heterogeneous setup and to aid the description of the experiment in Section V. Thus, the context below focuses on the architecture of the heterogeneous setup in Fig. 1, including the elements of the homogeneous setup.

1) Interaction space: Three different input tools (one per user), as interaction devices, were employed by the users to interact with a shared object. These devices were three types such as a mouse, a PHANToM® Omni device, and a crafted tool with a time-of-flight range camera (Swiss Ranger SR4000, MESA Imaging AG, Zurich, Switzerland). The Omni device could reflect force to its user's hand as haptic cue. The crafted tool was made of an elongated stick with a small black ball (3 cm in diameter) on the top of the stick, as illustrated in Fig. 1a. For all users of a collaborative group, the VE displayed a geological grid as the shared object on a wall-sized screen. To provide a 3D stereoscopic view to the users, we used the center screen of a computer-aided VE (CAVE). Each user used a pair of stereoscopic goggles to view the shared object. Fig. 2

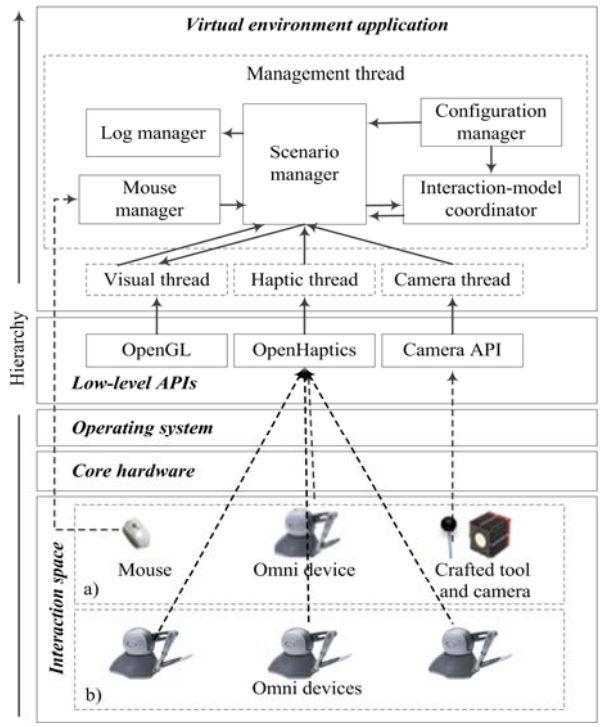

Figure 1: Architecture of a collaborative VE with two setups in the interaction space: (a) heterogeneous setup and (b) homogeneous setup.

exemplifies the interaction space.

2) Core hardware: We connected the three interaction devices to a graphic computer with a 2.53 GHz (dual quad core processors) Intel® Xeon® CPU, a 4 GB RAM, and a Quadro FX 4800 NVidia® graphics card.

3) Operating system: A copy of 64-bit Windows 7 Enterprise was the operating system installed on the computer.

4) Low-level APIs (application program interfaces): We used corresponding API to interface each interaction device with the layer of VE application. That is, OpenHaptics API and Camera API served to interface the Omni device(s) and the camera, respectively. For visual display, OpenGL API was used to render the shared object in 3D stereoscopic view.

5) VE application: Using C++, we implemented a software application of the collaborative VE for multiple users. The implementation was multi-threaded. One visual thread displayed virtual objects in the VE, one haptic thread permitted interaction with the shared object via the Omni device(s), and one camera thread employed the crafted tool via the range camera, if necessary. These threads cooperated along a management thread, which was responsible for handling the following functionalities:

- *Configuration manager:* inputted from a configuration file the scenario settings, including the number of collaborative sessions and the blocks of an experimental procedure, an interaction model used to resolve conflicts, the number of interaction opportunities, etc.
- *Interaction-model coordinator:* employed the DP or FCFS model based on the settings of a scenario inputted by the configuration manager.
- *Log manager:* logged data about user interactions.

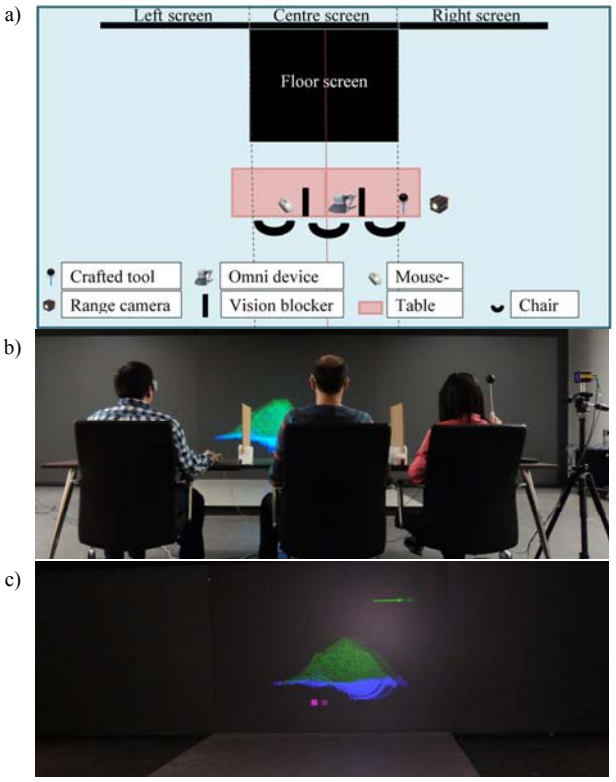

Figure 2: Interaction space of the heterogeneous setup for three human participants as users to interact with a shared object: (a) the layout of the VE; (b) the participants using different devices for interaction; and (c) a visual display for interaction, including a 3D geological grid as the shared object, a visual cue, and an arrow.

- *Mouse manager:* detected and managed the mouse events, if necessary.
- *Scenario manager:* ran an experimental procedure based on the steps inputted by the configuration manager.

B. Implementation of interaction models

The management thread handled the timing relationship among the visual, haptic, and camera threads, as indicated in Fig. 3. This timing relationship was determined by the frame rate of the OpenGL rendering (66 Hz or ~15 ms), the updating rate of the OpenHaptics scheduler for the Omni device (1 kHz or 1 ms), and the sampling rate of the camera (50 Hz or 20 ms). Governed by the operating system, the mouse event ran at a rate of about 1 kHz (or 1 ms) in the software application. This rate was equivalent to the updating rate of the Omni device. Due to the low sampling rate of the camera, the software application detected a command initiation issued by the crafted tool at a much slower pace than by the mouse event or the Omni device. To synchronize the detection of command initiations by all interaction devices, the interaction-model coordinator added a penalty to the detection time of the mouse event and the Omni device. The penalty was 19 ms, equal to the difference between the sampling rate of the camera and the updating rate of the mouse event or the Omni device. In addition, we considered the VRT of the users as 200 ms, the upper boundary of the VRT in neuroscience literature [33].

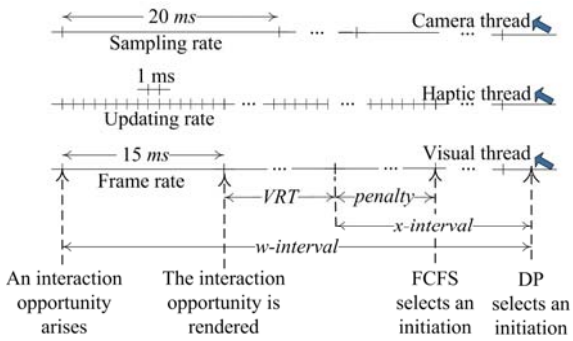

Figure 3: Timing relationship among the visual, haptic, and camera threads under both the DP and FCFS models for the heterogeneous setup.

At an interaction opportunity, the FCFS model selected the command initiation at the front of the queue. As illustrated in Fig. 3, the selection occurred at 219 ms (=VRT + penalty) after the rendering of the opportunity. In contrast, the DP model chose a command initiation among all entries of the queue arrived within the *x-interval*. As depicted in Fig. 3, the choice took place at 485 ms (=VRT + *x-interval*) following the rendering of the opportunity. Spreading from the emergence of the opportunity to the DP model's choice of an initiation, the *w-interval* was larger than one VRT to equal to 500 ms. We determined the length of the *w-interval* through several pilot tests to ensure that a user was unaware of a latency between issuing a command and the movement of the shared object within the *w-interval*. Thus, the *x-interval* was 285 ms, covering the much shorter penalty of 19 ms.

C. Participants

The verification had a total of 30 participants (16 males and 14 females with the average age of 24.73 ± 3.61 years), who differed from those participated in our earlier work using the homogenous setup. We conducted a baseline check to confirm the eligibility of all participants, including their ages, historical participation in our studies, vision, and handedness. As results, all participants were over 18 years old and naïve to the purpose of the verification. They had normal to corrected-to-normal vision with a stereo acuity of at least 40° arc as determined using the Randot Stereo-test. They passed color testing using the Ishihara color-blindness test. They were all right-handed and had no impairment for holding a stylus in an elongated shape and a mouse. We undertook handedness test using a modified version of the Edinburgh Handedness Inventory. These participants formed 10 groups of three participants (users). Each group undertook co-located multi-user collaboration within the VE. This verification was a within-subject-design and had the sample size of the 10 groups, which was larger than the minimal size (8) calculated using the Lehr's formula [34]. The verification had an ethics approval.

D. Procedure

As shown in Fig. 2a, three seats for the three participants of a group were placed at 650 cm in front of the center screen (10′ × 10′). Each interaction device was aligned with the right arm of a seat for a participant's right (dominant) hand. As depicted in Fig 2b, the range camera sampled the movement of the crafted tool. The camera was placed at a distance of 75 cm from the initial position of the crafted tool and calibrated according to this position. One flat obstacle was set between a pair of the participants. The obstacle prevented distractions from glimpsing other participants' hand movement. Using a pair of shuttle goggles, each participant saw a shared geological grid in 3D stereoscopic view, as shown in Fig. 2c.

All groups underwent the procedure of the well-controlled scenario, as described in Section III. At an interaction opportunity, the VE application presented the shared object and a yellow arrow pointing to one of six directions (left/right, up/down, and inward/outward), corresponding to a Cartesian coordinate system. As soon as the arrow turned into green, the task of the participants was to translate the object along the arrow-pointed direction simultaneously. For the task, one participant moved the mouse while holding down its left button for translating the object left/right and up/down and its right button for translating the object inward/outward. Another participant pressed and held the dark gray button of the Omni device while using its stylus to translate the object along the arrow-pointed direction. The third participant moved the crafted tool from its initial position along the specified direction to translate the object. However, only one of the participants could actually move the object, because the DP or FCFS model selected his/her command initiation among all queued initiations. A visual cue appeared to indicate the selection, while all participants viewed the visual cue and the movement of the object. The visual cue was a unique sphere, cube, and torus to represent the mouse, Omni device, and crafted tool, respectively. The interaction opportunity lasted 10 s, as ended by the reappearance of the yellow arrow. Being a special case of Eq. (11), the well-controlled scenario inflicted conflicts for a worst case of collaboration, similar as the homogenous setup.

There were totally one practice block and six testing blocks. The practice block consisted of three sections. In each section, a total of 30 interaction opportunities replicated six arrow-pointed directions for five times. The order of these opportunities was randomized. Each participant used a different interaction device for the opportunities in one section. Thus, the practice block ensured each participant to acquire the proper use of all interaction devices, to familiarize with the task, and to learn how to complete an identical questionnaire. Following the practice block, the six testing blocks were organized into three pairs. Each pair had one block under the DP model and another block under the FCFS model. The order of the models was counter-balanced. After each pair, the three participants switched their seats on a clockwise basis. Thus, all participants used each device for the task under both models. Each testing block included 30 randomized interaction opportunities, replicating each arrow-pointed direction for five times. At the end of each block, all three participants completed the identical questionnaire. The whole procedure lasted about 75 min, including the completion of the questionnaire and short breaks of 3 to 5 min between two blocks.

E. Data collection

In each testing block, we used two methods to collect data. One method was data logging. At each interaction opportunity, the VE application automatically logged the following information: the user who gained the control of the shared object and the history of interaction for all users. The latter was applied to compute the priority under the DP model but discarded under the FCFS model. The log recorded objective data about the actual interaction with the shared object for all participants in all testing blocks.

Another method collected subjective data through a two-component questionnaire. The first component requested each participant to mark a vertical line on a horizontal bar, which was bounded from 0% to 100%. The vertical line on the bar indicated the participant's perception (that arose in his/her mind) of gaining the interaction with the shared object. The second component applied the NASA TLX [30] to assess the workload perceived by each participant during a block. We converted all answers into numeric for analyses.

F. Data analyses

We applied the objective data to verify whether there was any misbehaver (as an outlier) among the participants of each group. Because the well-controlled scenario inflicted conflicts for a worst case of collaboration, each participant was required to initiate an interactive command at an interaction opportunity. From these logged data, we derived the percentages of the interaction with the shared object for all participants in the group and computed the mean and standard deviation of the percentages for the group. Although the mean was similar (about 33.3%) for each group under the DP and FCFS models, the standard deviation deserved an attention of investigation. The standard deviation under the DP model should be zero due to its theoretical definition described in Section III. In contrast, the standard deviation under the FCFS model should be non-zero, implying that a participant was more agile than the others. Any departure from these means and standard deviations suggested misbehavers among the participants. Their subjective data were to be consequently disqualified for analysis.

To analyze the subjective data for perceived EII and workload, we applied the statistical methods of two-way analysis of variance (ANOVA) and two-tailed paired *t*-test [35] for repeated measurements. Two pre-tests needed before these analyses. One pre-test evaluated normality (normal probability density function [36]) of the data to verify their validity for the statistical analyses. Based on the perceived percentages of interaction under the DP and FCFS models, we used Eq. (8) to compute their means and standard deviations. Another pre-test was one-way ANOVA to assess the indifference of the means between the DP and FCFS models. This pre-test ensured the comparability of the standard deviations under the DP and FCFS models.

To investigate the participants' perception of EII, we analyzed the standard deviations of the perceived percentage of interaction. Within the heterogeneous setup, each participant employed a different device for interaction in a testing block. Under the DP or FCFS model, we thus calculated the standard deviation of the perceived percentages for each device, SD_{deivce}, as follows:

$$SD_{deivce} = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}(p_{device,k} - \bar{p}_k)^2}\ , \qquad (17)$$

where $p_{device,k}$ is the perceived percentage of interaction by the *k*th participant of a group to use a particular device and $\bar{p}_k$ is the average perceived interactions by the *k*th participant. Based on the standard deviations derived from Eq. (17), we executed a two-way ANOVA (models × devices). To validate the results of analyzing the standard deviations, we repeated this analysis for the highest and lowest bounds of the perceived percentages, respectively. When the analyses presented any significant effect of differences, we performed two-tail paired *t*-test analyses to further examine the sources of the effect.

To analyze perceived worload, we used Eq. (10) to compute the workload from the data collected by the second component of the questionnaire (NASA TLX). Then, we conducted a two-way ANOVA on the workload (models × devices).

G. Results and discussion

The analyses of the objective data revealed that the total percentage of interaction opportunities with three command initiations within each group was 100% for every testing block. For all groups, the means of the percentages of the interaction were the same at about 33.3% in all testing blocks. The standard deviations of the percentages were indeed zero under the DP model and non-zero under the FCFS model. These results validated the participants' behavior for analyzing their subjective data. None of the participants was an outlier.

As a pre-test, normality tests verified the normal distribution of the subjective data for all testing blocks. As depicted in Fig. 4a, the means of the perceived percentages of the interaction under both DP and FCFS models were below 50.0% and within the range from 35.0% to 40.0% — close enough to the theoretical 33.3%. As another pre-test, one-way ANOVA on these means revealed that there was no significant difference between the DP and FCFS models [$F(1,9) = 0.29$; $p > 0.05$]. Thus, the standard deviations of the percentages were suitable for acquiring the perceived EII.

Fig. 4b illustrates the average standard deviations under the DP and FCFS models for all devices (i.e., the mouse, the Omni device, and the crafted tool). The average standard deviations of the DP model were smaller than those of the FCFS model. Two-way ANOVA (models × devices) revealed a significant difference between the models [$F(1,9) = 24.62$; $p < 0.001$] but indifference among the devices [$F(2,18) = 0.11$; $p > 0.05$]. There was no interaction between models and devices [$F(2,27) = 0.71$; $p > 0.05$]. These observations indicate that the differentiability of the averaged standard deviations between the DP and FCFS models is independent of the devices.

On the highest bound (mean + standard deviation) of the perceived percentages, a two-way ANOVA found that there was a significant difference between the models [$F(1,9) = 8.68$; $p < 0.05$] but no differentiability among the devices [$F(2,18) =$

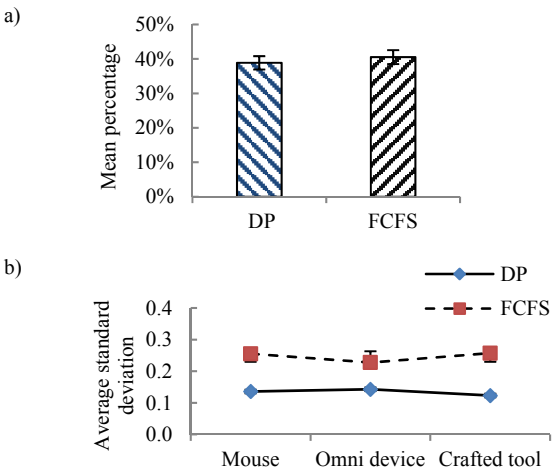

Figure 4: Analyses of means and standard deviations of the perceived percentages of the interaction: a) mean perceived percentages of the interaction; and b) average standard deviations. [Error bars represent standard errors.]

0.10; $p > 0.05$]. No interaction existed between the models and devices [$F(2,27) = 1.15$; $p > 0.05$]. The mean of the highest bound under the DP interaction model (52.97%) was much lower than its counterpart under the FCFS model (64.3%). The same analyses on the lowest bound of the perceived percentages (mean - standard deviation) yielded similar observations: significant difference between the models [$F(1,9) = 7.89$; $p < 0.05$], no differentiability among the devices [$F(2,18) = 0.04$; $p > 0.05$], and no interaction between the models and devices [$F(2,27) = 0.03$; $p > 0.05$]. However, the difference of the means of the lowest bound between the DP (24.6%) and FCFS (15.7%) models was much smaller than that of the means of the highest bound.

Two-way ANOVA (models × devices) on the data of *PW* indicated no significant difference between the models [$F(1,9) = 0.85$; $p > 0.05$] and among the devices [$F(2,18) = 0.40$; $p > 0.05$]. No interaction existed either between the models and devices [$F(2,27) = 0.71$; $p > 0.05$]. Therefore, the participants' *PW* was relatively indifferent under both interaction models and for three interaction devices.

The above results indicate that, without incurring extra *PW*, the DP model provided the participants a much better perceived EII than the FCFS model. This finding agreed with the effect observed in the homogeneous setup. Importantly, the perceived EII under the DP model is independent of interaction devices in use. Thus, the findings from the homogeneous setup and this prerequisite verification confirm that the DP model has the ability of resolving conflicts and yielding perceived EII under the well-controlled scenario. However, this ability remains unclear under a quasi-practical scenario, which permits users to work at their own pace. As confirmed by our previous work on multi-user usability [20], the quasi-practical scenario conformed to Eq. (11) to inflict occasional conflicts at some interaction opportunities. This paves the way for our experiment on examining multi-user satisfaction under the quasi-practical scenario.

V. Experiment

Using the framework of multi-user satisfaction presented in Section II, the experiment compared the DP and FCFS models to affect multi-user satisfaction. We conducted the experiment in a quasi-practical scenario within a collaborative VE. Derived from interviews with petroleum engineers, the scenario emulated a decision-making process of experts in petroleum industry. The process involves routinely three types of experts (i.e., reservoir engineers, production engineers, and geologists) whenever a problem arises from production of an oil/gas reservoir. Each expert consults certain changes and consequences related to part of the reservoir's properties (e.g., viscosity, pressure, and permeability). Together, they collaborate to identify factors that cause the departure of an actual production from its predicted counterpart. Their collaboration attempts to complete a property map of a geological grid. Although the map is a common goal to all experts, the individual task of each expert is uniquely different in the collaboration. As well, the task initiation and completion depend on the expert's own pace. Thus, the active contribution of each expert is crucial for achieving the common goal.

In brief, the quasi-practical scenario encompassed three unique attributes: (a) each user initiated and performed a task at his/her own pace with various lengths of completion time, (b) each user undertook a different task, and (c) all users collaborated to achieve a common goal. Hence, the scenario simulated a decision-making meeting of experts, who are peer users but specialized in various knowledge domains. The scenario was stochastic for both command initiation and task completion, as theorized by Eq. (11), to provide a relatively realistic situation of collaboration. That is, the scenario might not inflict conflicts at some interaction opportunities. In contrast, the well-controlled scenario used for the two-pillared prerequisites required all users to perform an identical task simultaneously to inflict conflicts. Due to its fixed processes of both command initiation and task completion, the well-controlled scenario was a special case of Eq. (11).

We assessed the multi-user usability of the DP and FCFS models under the quasi-practical scenario [20]. This work not only revealed that the DP model promoted effective, efficient, and satisfactory completion of collaborative tasks but also validated that the setup of the quasi-practical scenario was suitable for a collaborative VE. Hence, we employed the quasi-practical scenario for this experiment. Due to the fulfillment of the two-pillared prerequisites, we provided all users in the experiment with identical Omni devices for interacting with a shared object. The devices provided the users haptic cues to recognize their gained access to the object. The context below details the experiment that had a within-subject design.

A. Architecture

For the experiment, we implemented the collaborative VE with a five-layered and multi-threaded architecture as the homogeneous setup illustrated in Fig. 1b. This architecture employed three identical Omni devices and used the haptic thread to handle the events of these devices. As there was no

Figure 5: Corresponding placement between three Omni devices and three participants in the collaborative VE of the experiment.

need of using the mouse and camera-based crafted tool, we disabled the camera thread and mouse manager of Fig. 1. To simplify the timing relationship between the visual and haptic threads, the camera thread and penalty in Fig. 3 were disabled and set as 0 ms, respectively. Other hardware and software components of the architecture for the experiment were the same as those in Fig. 1 for the prerequisite verification.

B. Participants

Thirty right-handed participants (17 males and 13 females with the average age of 25.27 ± 5.12 years old) were involved in the experiment to form 10 groups. Being naïve to the purpose of the study, they underwent the same baseline check of eligible ages, historical participation in our studies, stereo acuity, color blindness, the use of a stylus, and handedness as in the prerequisite verification. These participants differed from those in our earlier work and the prerequisite verification. For the experiment, the 10 groups of the participants were more than the minimal 8 groups required by the Lehr's formula [34]. The experiment had an ethics approval.

C. Procedure

To carry out their collaboration, three participants of each group followed the same procedure under the quasi-practical scenario. The VE and participants' seating had the identical layout as depicted in Fig. 2a. Playing the role of an expert, each participant used an Omni device for interaction with the shared geological grid that possessed numerous property cells, as illustrated in Fig. 2c. Fig. 5 presents the placement of the Omni devices and the participants. Two flat obstacles among the devices blocked viewing the hand movements of the participants. Using shutter goggles, the 3D stereoscopic view was enabled for all participants.

There were three sessions: a practice session before two testing sessions. The practice session consisted of interaction opportunities, which were arbitrarily in order and examples of two testing sessions. In the practice session, we randomly assigned a pseudo-expert role to a participant of each group and trained him/her to become the pseudo-expert. Performing a unique list of tasks, each participant of the group needed to master three skills. The first skill was for interaction using an Omni device to rotate, translate, point, and highlight the shared object. The second skill was for collaboration by observing the activities of others on the display screen and using this observation to assist his/her next interaction. The third skill was for being an expert by understanding the grid organization of

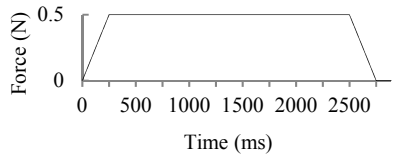

Figure 6: Force profile of the haptic cue.

the shared object to perform his/her designated tasks. To measure whether the participants acquired these skills, we logged data of task completion time and the number of accomplished tasks. The procedure of qualifying a pseudo-expert included two steps. At first, we used the logged data to ensure that a participant was able to complete about seven designated tasks consecutively in less than 30 s per task. Second, we asked the participant about his/her accomplishing the designated tasks. The outcomes of these steps were in agreement to qualify the participant as a pseudo-expert. The practice session lasted about 20 min, ended after each participant filled out an identical questionnaire.

Two testing sessions corresponded to the DP and FCFS models, respectively. The order of the sessions was counter-balanced for all groups. In each session, all pseudo-experts of a group played their trained roles to complete collaboratively a property map on the shared object (the common goal). According to his/her role, each pseudo-expert was assigned a unique list of designated tasks, corresponding to particular properties for a set of cells. A cell with a particular property needed to be found and highlighted with a color to label the property in consultation. Complementing each other, all lists together formed the property map on the shared object. Each pseudo-expert could decide whether or not to take part in an interaction opportunity. The participation was logged by his/her pressing the dark gray button on the stylus of his/her Omni device. When a pseudo-expert gained access to the shared object at an opportunity, his/her hand could feel a haptic cue via his/her Omni device. The haptic cue was a trapezoidal force, as depicted in Fig. 6. The access permitted the pseudo-expert at his/her own pace to accomplish one designated task from his/her list. The accomplishment of the task ended one interaction opportunity and began the next opportunity. As described in Section IV, the same visual signals indicated the beginning and end of an opportunity. A consent decision of the map marked the achievement of the goal.

Each testing session was divided into four blocks, with 30 interaction opportunities per block. Within a block, the number of accomplished tasks was not constant however. We constrained the time length of each block to be 5 min. This constraint yielded the total length of a testing session to be 40 min, including the time of completing the four blocks, the time of filling the questionnaires and short breaks. The length of the session met the upper threshold of human sustained concentration [37]. Thus, each group of three participants took at most 2 h in the study, including the baseline check, the practice session, and both testing sessions. Notably, the length of 2 h complied with that of regular meetings of experts in petroleum industry.

D. Data collection and analyses

Same as in the prerequisite verification, we gathered objective and subjective data for each group. The objective data were logged information about each pseudo-expert at an interaction opportunity, such as the initiation of his/her interactive command, the timing of the initiation, the identification number of the pseudo-expert who gained access to the shared object, the time length of accomplishing a task, and the history of interaction for all pseudo-experts. The subjective data were the same as those in the prerequisite verification.

Under the framework of multi-user satisfaction, we used all metrics presented in Table I for data analyses. From the logged data, we applied Eqs. (1)–(3) to compute *trust* and Eqs. (4)–(7) to assess *TF*, *DT*, and *CS* for measuring *usefulness*. Based on the subjective data, we used Eq. (9) to calculate the perceived percentages of interaction for measuring *pleasure* and Eq. (10) to obtain *PW* for measuring *comfort*. We undertook these measurements from collected data of both testing sessions, respectively. Because all metrics of *trust* were requirements to ensure intended behaviors of the VE, there would be unnecessary to compare each metrics between the DP and FCFS sessions. To compare each metrics of *usefulness*, *pleasure*, and *comfort* of the DP session to those of the FCFS session, we used the statistical method of one-way ANOVA (repeated measures).

Again, we performed pre-tests on the data to ensure the absence of outliers and the verification of normality before ANOVA. However, the criterion of determining outliers differed from that used in Section IV. Under the quasi-practical scenario, there might be no conflicts at some interaction opportunities. Each pseudo-expert of a group needed to participate in collaboration because of his/her unique task list to complement his/her peers' lists. Thus, the number of commands initiated by a pseudo-expert below a minimal threshold for a testing session indicated him/her to be an outlier. We set the threshold to be 25%, much lower than the average participation of 33% expected for the group. Both objective and subjective data of misbehavers were to be consequently disqualified for further analyses.

E. Results and discussion

There was no outlier among the participants of each group. For all groups, the data obtained for the metrics of *usefulness*, *pleasure*, and *comfort* in Table I were normally distributed under both testing sessions. These pre-test outcomes ensured ANOVA on the data.

For *trust*, the analyses of logged data revealed a consistency of *RTR*, *SI*, and *CR* under both DP and FCFS models. Table II gives the logged data averaged over all groups of participants for each of these metrics. System response time, *R(t)*, was averaged at 300.00 ± 8.41 ms and 15.50 ± 0.48 ms for the DP and FCFS models, respectively. This response time was less than the allowed upper boundary of 500 ms —— the *w-interval* as depicted in Fig. 3. When questioned after all sessions, none of the participants noticed a delay between initiating an interactive command and enacting the interaction. These met the condition of *RTR*, as required in Eq. (1). The number of logged command initiations per group, $\sum ICM$, was more than 2 at each interaction opportunity. This indicates the realization of *SI*, as specified in Eq. (2). Among the multiple initiations, the VE granted one participant an exclusive access, $\sum EA$, to the shared object. This validates the fulfillment of *CR*, as indicated in Eq. (3). Thus, the collaborative VE met all requirements to ensure its intended behaviors at all interaction opportunities under both DP and FCFS models.

TABLE II.
THE FULFILLMENT OF THREE METRICS FOR TRUST AVERAGED OVER ALL GROUPS OF PARTICIPANTS.

Logged Data	DP	FCFS	Criteria	Metrics Fulfillment
R(t)	300.00 ± 8.41 ms	15.50 ± 0.48 ms	≤500 ms	*RTR*: **Yes**
$\sum ICM$	2.82 ± 0.09	2.71 ± 0.18	≥2	*SI*: **Yes**
$\sum EA$	1	1	=1	*CR*: **Yes**

For *usefulness*, the ANOVA of *TF* revealed no significant difference between the DP and FCFS models. This was evident by [$F(1,9) = 0.06$; $p > 0.05$] for *TF*1 and [$F(1,9) = 2.72$; $p > 0.05$] for *TF*2. Being unit-less, *TF*1 was averaged as 55.17 ± 5.65% and 57.81 ± 11.36% under the DP and FCFS models, respectively. Similarly, *TF*2 had a mean of 93.85 ± 3.88% under the DP model versus 91.70 ± 4.70% under the FCFS model. However, the ANOVA of *DT* yielded a significant difference between both models [$F(1,9) = 5.55$; $p < 0.05$]. The average *D*T was 35.61 ± 1.15 min for the DP model and 39.57 ± 2.31 min for the FCFS model. As illustrated in Fig. 7, the average *DT* was much less spread among all groups under the DP model than under the FCFS model. The ANOVA on *CS* indicated a significant difference between the DP and FCFS models ([$F(1,9) = 72.38$; $p < 0.05$]). The average value of *CS* was 95.40 ± 1.28% under the DP model compared to 74.10 ± 7.57% under the FCFS model. The standard deviation was much smaller under the DP model than under the FCFS model, as depicted in Fig. 8.

Unsurprisingly, the results of analyses on *pleasure* and *comfort* were in agreement with those of the homogeneous setup and the prerequisite verification (i.e., the heterogeneous setup). That is, ANOVA on the standard deviation of perceived percentages of interaction revealed a significant difference between both models ([$F(1,9) = 41.16$; $p > 0.05$]). Further analyses confirmed this observation with [$F(1,9) = 15.62$; $p < 0.05$] for the lowest bound and [$F(1,9) = 20.68$; $p < 0.05$] for the highest bound. Again, there was no significant difference of *PW* between both models [$F(1,9) = 1.05$; $p > 0.05$].

These findings imply that the DP model offers an advantage over the FCFS model on multi-user satisfaction. Compared to the FCFS model, the DP model promotes *usefulness* of the collaborative VE by offering the similar level of *TF*, reducing *DT* and enhancing *CS*. With the VE's intended behaviors ensured by *trust*, the *usefulness* was accompanied with fostering

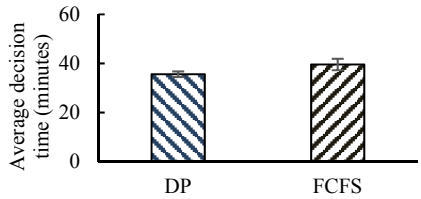

Figure 7: Average decision time among all groups under the DP and FCFS models. [Error bars represent standard errors.]

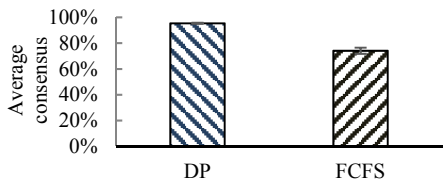

Figure 8: Average consensus among all groups under the DP and FCFS models. [Error bars represent standard errors.]

better *pleasure* for the similar level of *comfort* in the VE. This advantage might be achieved by providing the perception of EII to fulfill cognitive needs among multiple users as *pleasure*. After all, the perception could increase the awareness of multiple users for establishing a foundation toward user satisfaction [2].

VI. General Discussion

Existing research efforts of improving collaborative VEs disregard the cognitive needs of multiple users and employ various metrics of user satisfaction [15][16][38]–[42]. Measuring user satisfaction experienced in single-user tasks, the existing metrics are unfit to gauge multi-user satisfaction in collaborative work. In contrast, our framework of multi-user satisfaction has two advantages: (a) to champion the combination of both ISO/IEC 25010:2011 standard [19] and multi-user cognitive needs and (b) to foster the collaboration of multiple users. Although the framework takes account objective and subjective data to remedy individual variations in interpreting questionnaires, future work is needed to validate each metrics by comparing its objective and subjective data.

The framework is generic and applicable to both distributed and co-located settings as long as a VE underlying a collaboration fulfills all three metrics of *trust* and provide proper awareness among the users. Derived from ISO/IEC 25010:2011 standard [19], the metrics of *trust* reflect the nature of collaboration, the metrics of *usefulness* indicate the achieved degrees of pragmatic aims under collaboration, and the metrics of *pleasure* and *comfort* give levels of cognitive needs fulfilled and of workload required for collaboration. In other words, the metrics of each factor is related to collaboration but independent of users' tasks. The framework would thus be potential for gauging multi-user satisfaction under a collaborative system, which supports users' interaction with shared objects/information.

Based on the framework, our experiment demonstrated the effect of the DP model on multi-user satisfaction. Targeting multi-user cognitive needs (i.e., EII) for collaboration, the DP model resolves conflicts to fulfill Eq. (3) and thus to promote multi-user satisfaction within collaborative VEs. This encourages genuine collaboration among multiple users to cultivate the efficiency of a collaboration [2]. To our best knowledge, this experiment advocates uniquely the role of interaction models in affecting multi-user satisfaction. Although the experiment was undertaken in a co-located setting, the outcomes of the study are serviceable to collaborative work in a distributed setting. This versatile serviceability needs to be warranted through the implementation and verification of *RTR*, *SI*, and *CR* for a multi-user collaborative system.

TABLE III.
Comparison of the interaction models.

Factors/metrics		DP	FCFS	Significant difference
Trust		Yes	Yes	—
Usefulness	Task focus	High	High	No
	Decision time	Low	High	**Yes**
	Consensus	High	Low	**Yes**
Pleasure		High	Low	**Yes**
Comfort		Moderate	Moderate	No

In our experiment, we compared the DP model to the FCFS model that is mostly common for conflict resolution. The comparison overcomes an issue of determining a threshold value for applying each metrics, because such determination is usually dependent on users' tasks. The comparison was supported by two well-examined considerations. One consideration was two-pillared prerequisites. The observations of both prerequisites built a solid foundation to apply the homogenous setup of interaction devices and haptic cues for the comparison. Another consideration was scenarios of multi-user collaboration. The two-pillared prerequisites used a well-controlled scenario, which inflicts conflicts at each interaction opportunity. In contrast, the comparison employed a quasi-practical scenario, which mimics a collaborative meeting of industrial experts. Depending on the tasks of each expert, the scenario did not always inflict conflicts at some interaction opportunities as validated in our work on usability of collaborative VEs [20]. Together, the two considerations not only confirmed conflict resolution of the DP model for providing the perception of EII but also verified the quasi-practical scenario for multi-user collaboration — a novel aspect of the experiment.

As summarized in Table III, both DP and FCFS models ensure the fulfillment of *trust* and realize the same level of *comfort* for a multi-user collaborative VE. The DP model underlies the fulfillment of the framework of multi-user satisfaction significantly better compared to the FCFS model. Although both models have an indifference of high *TF*, the DP model results in lower *DT* and higher *CS* than the FCFS model.

These observations indicate that the DP model elevates the *usefulness* of the VE by promoting genuine collaboration. The elevation concurs with higher *pleasure* under the DP model compared to the FCFS model. Consequently, the DP model fosters higher multi-user satisfaction than the FCFS model. This outcome results directly from the distinct principles of conflict resolution used by both DP and FCFS models, as described in Section III. In multi-user collaboration, the DP model meets cognitive needs of the users to defy the "winner-takes-all" mentality of the FCFS model. That is, the perception of EII plays a crucial role in fulfilling multi-user satisfaction. Evidently, the DP model persuades collaboration by offering EII for all users. Hence, how an interaction model resolves conflicts to meet cognitive needs has a great impact on achieving a common goal in a multi-user collaborative VE.

Interestingly, the DP model was independent of the types of interaction devices. The *comfort* under both DP and FCFS models were similar, as observed in the two-pillared prerequisites and the experiment. That is, the utilization of different types of interaction devices, such as a cheap mouse, a reasonably costed Omni device, and an expensive crafted tool with a range camera, did not impair the workload of multiple users in collaborative work. This opens a possibility of using cheaper interaction devices for a multi-user collaborative VE if developmental costs of the VE are a concern. The benefit of cost reduction certainly strengthens the DP model for use in industrial settings of collaboration.

In both well-controlled and quasi-practical scenarios, we imposed a constraint of no verbal communication among the participants of each group. This constraint enabled us to examine the effect of the DP model on multi-user satisfaction. However, verbal and visual communication among multiple co-located or distributed users might also contribute to their perceived collaboration and thus to multi-user satisfaction. Further work remains to examine the robustness of the DP model to fulfill the framework of multi-user satisfaction under verbal and visual communication.

VII. Conclusion

We presented a framework of multi-user satisfaction to incorporate both the ISO/IEC 25010:2011 standard and the cognitive needs of multiple users. Based on the framework, we conducted an experiment on assessing the DP and FCFS models for multi-user interaction within a collaborative VE. The experiment was supported by two well-examined considerations. One consideration is a prerequisite verification. The verification ensured that the perception of EII offered by the DP model is independent of the types of interaction devices. Another consideration is the use of a validated quasi-practical scenario to mimic a collaborative meeting of industrial experts. The results of the experiment revealed that, compared to the FCFS model, the DP model induces significantly higher degrees of *usefulness* and *pleasure* while sustaining intended system behaviors by *trust* and maintaining a similar level of *comfort* for multi-user satisfaction. The experiment sheds a light on how to design interaction models to promote genuine collaboration of multiple users within VEs without considering verbal communication. Future work will thus validate the metrics of the framework and verify the robustness of the DP model under verbal communication in a practical scenario of multi-user collaboration.

References

[1] A. P. Schouten, B. van den Hooff, and F. Feldberg, "Virtual team work: Group decision making in 3D virtual environments," *Commun. Res.*, pp. 1–31, Nov., 2013.

[2] H. Patel, M. Pettitt, and J. R. Wilson, "Factors of collaborative working: A framework for a collaboration model," *Appl. Ergonom.*, vol. 43, no. 1, pp. 1–26, Jan., 2012.

[3] W. Chen, N. Ladeveze, C. Clavel, D. Mestre, and P. Bourdot, "User cohabitation in multi-stereoscopic immersive virtual environment for individual navigation tasks," in *Proc. IEEE Virtual Reality Conf.*, Arles, France, 2015, pp. 47–54.

[4] N. Yuill and Y. Rogers, "Mechanisms for collaboration: a design and evaluation framework for multi-user interfaces," *ACM Trans. Comput. Hum. Interact.*, vol. 19, no. 1, pp. 1–25, Mar., 2012.

[5] M. A. Nacenta, D. Pinelle, C. Gutwin, and R. Mandryk, "Individual and group support in tabletop interaction techniques," in *Tabletops—Horizontal Interactive Displays*, Springer, London, pp. 303–333, Mar., 2012.

[6] E. Tse, S. Greenberg, C. Shen, C. Forlines, and R. Kodama, "Exploring true multi-user multimodal interaction over a digital table," in *Proc. 7th ACM Conf. Des. Interact. Syst.*, Cape Town, South Africa, 2008, pp. 109–118.

[7] D. Pinelle, M. Barjawi, M. Nacenta, and R. Mandryk, "An evaluation of coordination techniques for protecting objects and territories in tabletop groupware," in *Proc. Conf. Hum. Factors Comput. Syst.*, Boston, MA, 2009, pp. 2129–2138.

[8] G. Eckes, *Six Sigma Team Dynamics: The Elusive Key to Project Success*. John Wiley & Sons, 2002.

[9] Y. Lou, W. Wu, and H. Zhang, "Magic input: A multi-user interaction system for SAGE based large tiled-display environment," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, 2012, pp. 157–162.

[10] A. Simon and C. Stern, "Coordination policies for co-located collaborative travel," in *Proc. IEEE Virtual Reality Conf.,* Reno, NV, USA, 2008, pp. 285–286.

[11] D. Sibbet, *Visual Meetings: How Graphics, Sticky Notes and Idea Mapping Can Transform Group Productivity*. John Wiley & Sons, 2010.

[12] A. Erfanian, T. Zeng, and Y. Hu, "Dynamic strategies of conflict resolution on human perception of equality within multi-user collaborative virtual environments," in *Proc. 9th Int. IEEE Conf. Collab. Comput.: Netw. Appl. Worksharing*, Austin, TX, USA, Oct., 2013, pp. 363–370.

[13] Y. Rogers, Y. k. Lim, W. R. Hazlewood, and P. Marshall, "Equal opportunities: do shareable interfaces promote more group participation than single user displays?," *Hum. Comput. Interact.*, vol. 24, no. 1–2, pp. 79–116, Apr., 2009.

[14] L. Razmerita and K. Kirchner, "Social media collaboration in the classroom: a study of group collaboration," in *Collab. Technol. Lect. Notes Comput. Sci.*, vol. 8658, Springer International Publishing, pp. 279–286, Sept., 2014.

[15] K. Hornbæk, "Current practice in measuring usability: challenges to usability studies and research," *Int. J. Hum. Comput. Stud.*, vol. 64, no. 2, pp. 79–102, Feb., 2006.

[16] G. Brajnik and C. Giachin, "Using sketches and storyboards to assess impact of age difference in user experience," *Int. J. Hum. Comput. Stud.*, vol. 72, no. 6, pp. 552–566, Jun., 2014.

[17] S. P. Parikh, J. M. Esposito, and J. Searock, "The role of verbal and nonverbal communication in a two-person, cooperative manipulation task," *Adv. Hum. Comput. Interact.*, vol. 2014, pp. 1–10, Aug., 2014.

[18] K. Sinmai and P. Andras, "Mapping on surfaces: supporting collaborative work using interactive tabletop," in *Collab. Technol. Lect. Notes Comput. Sci.,* vol. 8658, Springer International Publishing, pp. 319–334, Sep., 2014.

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

[20] A. Erfanian and Y. Hu, "The effect of interaction models on multi-user usability of collaborative virtual environments," in *Proc. Int. Conf. Collab. Comput.: Netw., Appl. Worksharing*, Miami, FL., USA, Oct., 2014, pp. 187–194.

[21] P. Cairns, "Experimental methods in human-computer interaction," in *Encyclopedia of Human-Computer Interaction*, M. Soegaard, and R. F. Dam, 2nd ed., Interaction Design Foundation, 2016, chap. 34, [online]. Available at: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/experimental-methods-in-human-computer-interaction#chapter_start.

[22] R. S. Pressman, *Software Engineering: A Practitioner's Approach*, Palgrave Macmillan, 2005.

[23] G. Mark, S. Voida, and A. Cardello, "A pace not dictated by electrons: an empirical study of work without email," in *Proc. Conf. Hum. Factors Comput. Syst.*, Austin, TX, USA, 2012, pp. 555–564.

[24] R. Huang, S. Kahai, and R. Jestice, "The contingent effects of leadership on team collaboration in virtual teams," *Comput. Hum. Behav.*, vol. 26, no. 5, pp. 1098–1110, Sep. 2010.

[25] P. L. McLeod"An assessment of the experimental literature on electronic support of group work: Results of a meta-analysis," *Hum. Comput. Interact.*, vol. 7, pp. 257–280, 1992.

[26] E. Herrera-Viedma, F. Herrera, and F. Chiclana, "A consensus model for multiperson decision making with different preference structures," *IEEE Trans. Syst. Man Cybernet. Pt. A Syst. Hum.*, vol. 32, no. 3, pp. 394-402, May 2002.

[27] I. E. Allen, and C. A. Seaman. "Likert scales and data analyses," *Qual. Prog.*, vol. 40, no. 7, pp. 64–65, Jul. 2007.

[28] J. P. Formby, W.J. Smith, and B. Zheng, "The coefficient of variation, stochastic dominance and inequality: A new interpretation", *Econ. Lett.*, vol. 62, no. 3, pp. 319–-323, Mar. 1999.

[29] G. Adamides, G. Christou, C. Katsanos, M. Xenos, and T. Hadzilacos, "Usability guidelines for the design of robot teleoperation: a taxonomy," *IEEE Trans. Hum. Mach. Syst.*, vol. 45, no. 2, pp. 256–262, 2015.

[30] S. G. Hart, "NASA-Task Load Index (NASA-TLX); 20 years later," in *Proc. Hum. Factors Ergonom. Soc. 50th Annu. Meet.*, vol. 50, no. 9, pp. 904–908, Oct., 2006.

[31] R. Nelson, *Probability, Stochastic Processes, and Queueing Theory: The Mathematics of Computer Performance Modelling*. New York: Springer-Verlag, 1995.

[32] M. G. Kendall, *Advanced Theory of Statistics*, vol. 6. London; New York: Arnold; Distributed in the USA by Oxford University Press, 1994.

[33] A. T. Welford, "Choice reaction time: Basic concepts," *React. Times*, pp. 73–128, 1980.

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

[35] D. C. Montgomery, *Design and Analysis of Experiments*. John Wiley & Sons, 2008.

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

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

[38] P. Schmutz, S. Heinz, Y. Métrailler, and K. Opwis, "Cognitive load in ecommerce applications: measurement and effects on user satisfaction," *Adv. Hum. Comput. Interact.*, vol. 2009, pp. 1–9, 2009.

[39] J. Goodman, S. Brewster, and P. Gray, "Using field experiments to evaluate mobile guides," in *Proc. HCI Mobile Guides, Workshop at Mobile HCI*, 2004, Springer, Glaskow, 2004, pp. 38–44.

[40] A. Riccio, F. Leotta, L. Bianchi, F. Aloise, C. Zickler, E.J. Hoogerwerf, A. Kübler, D. Mattia, and F. Cincotti, "Workload measurement in a communication application operated through a P300-based brain-computer interface," *J. Neural Eng.*, vol. 8, no. 2, pp. 1–6, Apr., 2011.

[41] R. H. Bokhari, "The relationship between system usage and user satisfaction: a meta-analysis," *J. Enterprise Inf. Manag.*, vol. 18, pp. 211–234, 2005.

[42] E. L.-C. Law and S. Abrahão, "Interplay between user experience (UX) evaluation and system development," *Int. J Hum. Comput. Stud.*, vol. 72, no. 6, pp. 523–525, 2014.