Questioning the use of virtual reality in the assessment of the physical impacts of real-task gestures and tasks

Conference Paper · October 2017

DOI: 10.1109/VSMM.2017.8346271

3 authors:

Sobhi Ahmed
Paris8 University

Laure Leroy
Université de Vincennes - Paris 8

Ari Bouaniche
Université de Vincennes - Paris 8

Questioning the Use of Virtual Reality in the Assessment of the Physical Impacts of Real-Task Gestures and Tasks

Sobhi Ahmed
Computer science department
Birzeit University
Palestine
sahmed@birzeit.edu

Laure Leroy* and Ari Bouaniche†
Paragraphe Laboratory
University Paris 8
France
*laure.leroy02@univ-paris8.fr
†abouaniche@yahoo.com

***Abstract*—In the context of growing preoccupation about ergonomics and the minimizing of negative impacts of body movements performed during gestural interaction with a VR system, we implemented a gestural fatigue assessment system based on various standards and specifications. While our approach focused on VR, similar systems have been commercialized and propose to assess real-world gestural fatigue using a virtual reality environment. We wonder whether this is a valid approach, and use data collected by the system we have devised to appreciate fatigue due to a task completed both in a real-world environment as well as a VR environment. Participants were also asked to appreciate the fatigue they felt subjectively, thanks to questionnaires. We found that fatigue levels were higher in the virtual environment for almost all joints. Task duration was always higher in virtual reality environment. Furthermore, there seems to be no correlation between the automated objective results of fatigue assessment in the real-world and those obtained assessing the same tasks in a VR environment, thus questioning the validity of the transferability advocated by such systems. Correlation for subjective results is, however, high.**

Keywords—gestural interaction, gesture assessment, musculoskeletal disorders, gestural fatigue, ergonomics, virtual reality.

I. Introduction

Gestural interactions aim to provide intuitive, more natural, and easier ways to interact with systems, through using gestures. [1][2][3]. A gesture, according to [4], is "a physical motion that has information". This type of interaction is used in virtual reality environments and, for the sake of immersion, the gestures used to interact with such systems often aim to resemble to gestures used in daily life. These interactions are supposed to entail less cognitive and physical effort than interactions traditionally associated with computer systems: the use of a mouse, for example, demands an undesirable physical effort because of the mouse's distance from the user, which calls for the their arm to be outstretched while requiring a very accurate gesture when pointing [5].

The gestures used in gestural interactions may however provoke physical fatigue and musculoskeletal disorders as much as real gestures do. This fatigue can be caused by movements requiring substantial physical effort [6][7][8][9]. What is more, the extended and/or frequent use of such systems can result in an overuse of the muscles in charge of performing said gestures [10]. Such movements could occasion injuries called "Repetitive Strain Injuries" (RSIs). Several diseases have been associated with RSIs such as tendinitis, bursite, tenosynovitis, carpal tunnel syndrome, etc. [11].

Few studies have been conducted on how to reduce the physical impact of gestural interactions on the human body, and the lack thereof has sometimes resulted in the creation of non-ergonomic, stressful gestures that are difficult to use [6]. Interaction with such systems can, in turn, lead to various musculoskeletal injuries. This has led us to propose an approach for gestural fatigue assessment [12].

Similar preoccupations about gestural fatigue assessment have also been addressed in the workplace, first by creating and implementing worker safety standards and specifications, but also —more recently— with automated gestural fatigue assessment systems similar to ours (ergoWide [13], ergoAudit [14]). What is striking, however, is that such systems invariably offer to assess the operator's real-world fatigue while submitting them to a task conducted within a VR environment. Such a discrepancy deserves a proper investigation, and we propose to study, in this paper, the difference in fatigue levels after the completion of tasks with virtual objects as well as real ones, under very similar conditions. The level of fatigue was assessed and recorded in each environment using our gestural fatigue assessment system. Subjective evaluations of fatigue levels were also obtained. We analyzed and compared the data for each body part.

In this paper, we study this issue by implementing an experiment whose goal is to verify whether virtual and real interactions are indeed similar, and whether the systems positing such a thing can be used to asses real-world

978-1-5386-4494-2/17/$31.00 ©2017 IEEE

workplace tasks. In our experiment, subjects performed the same tasks under closely similar conditions, in real and virtual environments.

II. Related Work

A. Virtual assessment of real-life gestures

It is first worthy of notice that this this issue has not been studied extensively: References [15][16] and [17] have proposed experiments addressing VR-based system simulating a real workplace task in a virtual environment, and then studied the difference in fatigue while performing the task both in virtual and real environments. Reference [15]'s goal was to check whether virtual reality technologies can be used to assess real-world workplace ergonomics, focusing particularly on physical fatigue. Their results show that physical fatigue was higher in real environments according to most objective measures, including the RULA technique (cf. Related work). They found, on the contrary, that fatigue was higher in a virtual environment according to subjective appreciation. One of the reasons for such findings could be the difference in conditions between the virtual and the real environments. In the virtual environment, subjects were instructed to use a flystick to manipulate the objects, while in the real-world environment, subjects manipulated objects using only their hands. Other factors are also mentioned in possibly influencing the difference in fatigue, such as motion range, missing haptic feedback, technical constraints, etc. In the experiment, only upper body parts were assessed. Time spent in performing the task in a virtual environment was higher than that in a real-world setting. The study furthermore evaluated global felt fatigue and did not offer detailed results per jointt. Reference [17] studied the difference in torso motion while performing a task of lifting boxes in virtual as well as real environments. The study only focuses on the difference in lumbar angle motion, as well as the duration and acceleration. There seems to be no significant difference in angle range between the two environments, but in contrast, significant differences were found in acceleration and task duration: this was probably caused by the use of an HMD device. Reference [16] compared the fatigue levels between virtual and real environments in a drilling task. The study uses 5 measurements: elbow angle, maximum force capacity, time, Body part discomfort (BPD) and Rated Perceived Exertion (RPE). The only tested joint was the elbow, because of technical limitations related to the sensors used in the experiment. The aforementioned indicators were higher in the virtual environment; the difference was significant for almost all of them (the difference in elbow angle was not significantly different according to ergonomic methods such as RULA). As to fatigue, it was higher in the virtual environment according to most indicators. The time spent completing the virtual task was greater than that of the real-world setting. The main drawbacks identified in the study were the level of presence, and the technical limitation which prevented studying more joints.

B. Gesture assessment: existing methods and standards

Many factors may influence fatigue caused by gestures used during interactions and physical movements used in daily life and in workplace amongst these factors: joint angles, repetition, duration, force exerted, as well as other factors such as the use of tools occasioning pressure on some body parts [18][19][6][20].

There exist some methods and norms that deal with gestural fatigue. Most evaluation systems are based on some of these methods. Some of them present specifications of the levels of fatigue caused by physical movements depending on some risk factors, and present principles to evaluate those may cause fatigue (such as RULA). These methods and standards can be classified into two categories:

1) Subjective methods

Most studies on the assessment of the negative impact of gestures and physical movements in general resort to subjective methods [19][21]. Amongst those, one can find:

- The Body Discomfort Diagram method (BDD), which assesses the level of discomfort in different parts of the body using a body diagram and an assessment scale. The diagram allows identifying and assessing the places and sources of discomfort by marking the affected areas [22].
- Scoring methods, where a number of points is assigned to each single movement and criterion, resulting in a final score which determines the gesture's level of comfort. Each single score is decided either by the users [19] or by experts (ergonomists, etc.) [23].
- Other methods are used, such as questionnaires [24], interviews, open-ended questions [21].

2) Assessment Methods and Standards

There exist methods and standards which allow the assessment of physical movements in a more objective way:

- Electromyogram. The electromyogram is a tool which measures muscle activity through the detection and recording of electric signals sent by muscle motor cells used during activity. The electric signal is amplified and processed to determine the level of muscle force exerted. [25][26]. This technique is used by [21] to measure muscle activity pertaining to the gestures and effort when interacting with touch-enabled devices. It is also use by [15] in their evaluation process.
- RULA (Rapid Upper Limb Assessment). RULA is a risk-factor assessment technique for upper limbs, geared towards individuals subjected to postures, forces and muscle loads potentially leading to MSDs [23]. The assessed factors are: number of movements, static work, force, work posture and working time. RULA was used by many systems that deal with the evaluation of physical movements [13][14][15].

RULA allows the attribution of a final assessment score for each posture ranging from 1 to 7. This score indicates the level of discomfort for the posture: the higher the score, the higher the risk. It follows diagrams specifying the ranges of joint angles for various body parts. In these diagrams, a score is given to each movement depending on its angle (the farther the angle from a neutral position, the higher the score). This numbering system is also used to specify the level of force exerted as well as static and repetitive muscular activity. To calculate the scores, three score charts —defined by ergonomists— are used [23].

- The ISO 11226 standard. The ISO 11226 standard [27] aims to assess health hazards for workers involved in manual labor. The assessment process involves specifying and classifying posture conditions for each body part as acceptable or not. These conditions comprise joint angle, time-related aspects and movement repetition. The classification is based on experimental studies as well as the current knowledge in ergonomics.

  The assessment procedure is a one- or two-step process. The first step measures joint angles. If said angles do not exceed a given limit, the posture is deemed 'acceptable'. If not, the second step focuses on the time span for which the posture is sustained. Extreme angles are never recommended. There exist several methods to recognize postures, such as observation, video, etc. Other factors are considered while assessing static postures, such as support (or its absence), sitting or standing position, etc.

- The AFNOR NF EN 1005-4 standard (Safety of machinery – Human physical performance). NF EN 1005-4 is an AFNOR standard [28] aiming to improve machine design in order to decrease health risks by avoiding postures and stressful movements leading to MSDs. This is done through the specification of various recommendations as well as a posture- and movement-related risks assessment method.

  It defines a posture and movement assessment procedure related to working with machinery. The assessment can either be 'acceptable', 'acceptable under conditions' or 'unacceptable'. The assessed risk factors are: movement angle, gesture time, frequency, etc. In situations determined as 'acceptable under conditions', other risk factors must be considered, such as duration, repetition, period of recovery, the presence of a support to the body, etc.

III. Our Approach to Assessing Gestural Fatigue

We developed an assessment method for gestures used during interaction so as to minimize their negative physical impacts. It allows detecting the conditions and variables of users' freeform gestures, assess them, and determine their level of fatigue automatically according to various pre-existing methods and standards. Variables include joint angles, duration, frequency, supports for the body, movement and posture style (weight distribution on both feet, rotation, etc.). This application can assess the fatigue levels on almost all body parts (shoulder, elbow, neck, trunk, wrist, etc.). It could be used in the design phase of gestural interactions to decide which gestures are the least fatiguing to users.

The variables are assessed based on specifications for acceptable and unacceptable movements in various studies and standards such as, RULA, ISO 11226 and AFNOR 1005-4. [23][27][28][29]. (cf. Related work).

As preparation, we organized the data related to each joint in tables specifying all possible movement types for said joint, and sifted through standards to detail acceptable or unacceptable values for various movement criteria.

Our setup uses a Microsoft Kinect for Xbox® sensor [30] to detect motion angle, duration, and repetition. Other variables not detected by Kinect are entered manually by the user, for example, if the weight is distributed equally on both feet, the presence of supports for the body, etc.

We validated our approach by evaluating its results through a comparison between the system's evaluations and subjects' evaluations of their fatigue levels after performing some tasks using gestural interactions. We found that the system was able to faithfully assess physical gestural fatigue during interaction. We therefore propose to use our evaluation system in our experiment to assess fatigue levels while performing tasks in virtual and real environments. [31]

IV. Experiment

Our experiment aimed to study the difference in fatigue levels performing tasks in virtual and real environments, recording said levels in both environments. To do that, we have designed extremely similar tasks to be performed in a virtual as well as a real environment. The conditions of the virtual environment task were reproduced in a real environment using real objects that mimic their virtual counterparts as much as possible. In this section, we introduce our experimental setup in more detail.

A. Participants

Twenty-six participants (aged 29 ± 10 years old), 17 males and 9 females, all had beginner level with gestural interactions) were tested, because we were aiming to test a gestural interface destined to the general public. We therefore focused on potential new users of virtual and augmented reality environments.

B. Tasks and procedure

Participants were asked to perform —in a random order — two tasks in a virtual as well as a real environment. The tasks were about arranging objects: the subject would pick an object from a stock box, and then move and drop it into another designated box. One task was deemed difficult when the other was deemed easy (cf. Fig. 1).

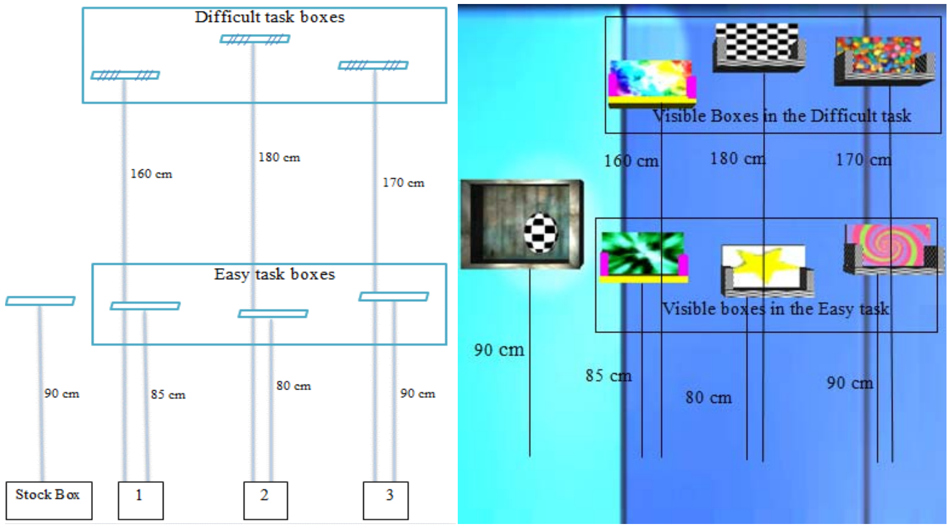

Fig. 1. : Boxes locations in real environment (left) and in virtual environment (right).

- Difficult task: the task was designed to be tiring, as boxes were positioned above the level of the center of the body; their heights were respectively 160, 180, and 170 cm. In this case, the subject had to raise their hand above shoulder level in order to move the object and drop it into the appropriate box. In addition, we made sure that this task requested more precision than the second.
- Easy task: the task was supposed to be easier; boxes were positioned around body center level, their heights being respectively 85, 80, and 90 cm.

Subjects completed the task in three steps, using gestural interaction in the virtual environment and actual movements in the real environment:

1) Picking the object from the stock box (at a height of 90 cm), by pointing at it with their right hand and then closing it (virtual environment) / grabbing an object with fingers (real environment).

2) Moving the object by moving their right hand towards the illuminated box (virtual environment) /the box number uttered by the experimenter (real environment).

3) Dropping the object in the appropriate box by opening their hand.

In each environment, there was a total of six boxes. In the virtual environment, only three boxes were visible during each task. In each task, a light indicated to subjects where to drop their object. However, in the real environment, a randomly selected box number was indicated orally to the subject by the experimenter.

In each condition, the task was repeated 30 times in order to move 30 objects from the stock box to the other boxes. The number of times was chosen after pre-tests. The subjects were asked to return to a resting position between each task. The order in which tasks were performed was random. During task completion, our system collected data to perform a real-time assessment.

Each subject performed the task in both conditions and in both environments sequentially, this means that subjects performed both tasks in random order in the first environment which was selected randomly, and then performed both tasks also in random order in the second, remaining environment (cf. Fig. 2)s.

We began by introducing our work and the experiment steps orally. As shown in Fig. 2, subjects started their first task in the first randomly chosen environment, then, during a 10-minutes break between the two conditions, filled a questionnaire about the fatigue levels felt. Next, the subject performed the task in the second condition within the same environment, which was followed by a second questionnaire about their levels of fatigue. In a second place, subjects repeated the same procedure done in the first environment, in the second one.

Experiment duration was about 53 minutes, including 13 minutes on average for performing tasks, and 40 minutes for breaks.

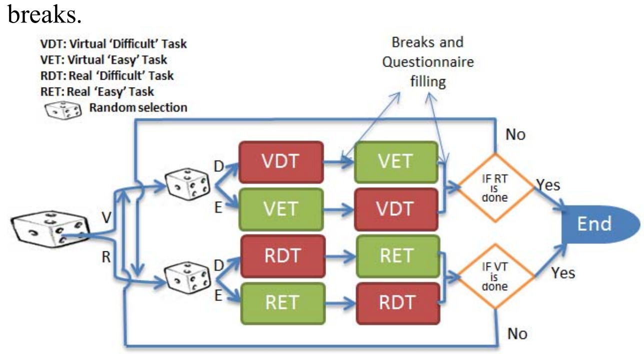

Fig. 2. Experiment Progress

C. Apparatus

We used three PCs running Windows 7. Each PC was combined to a Microsoft Kinect for Xbox® motion sensors. For each environment, one PC connected with a Kinect sensor was used to track user movements, process data and display assessment results. This computer ran C# code, developed in Microsoft Visual Studio 2013. In the virtual environment, the Kinect sensor (Kinect 1) was placed in the middle of the active zone, facing the active zone directly. Its height was about 70 cm (cf. Fig. 3). In the real environment the Kinect was behind the boxes facing the middle of the active zone, its height was about 75 cm (cf. Fig. 4).

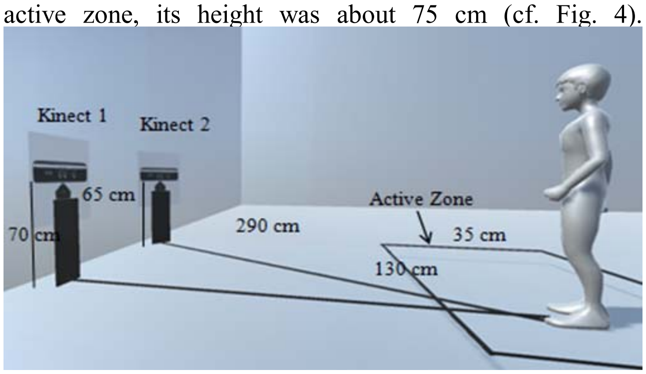

Fig. 3. Kinect locations and active zone in the virtual environment

In the virtual environment, a second computer was used to run the test application comprising the gestural interface; a video projector was attached to this computer. Unity3D was used to develop our test interface. The computer was also connected to a Kinect sensor which allowed users to manipulate the gestural interface. The Kinect was placed 290 cm from the subject to the left of the middle of the active zone (Kinect 2). Its height was about 65 cm, its rotation angle was about 30 (cf. Fig. 3).

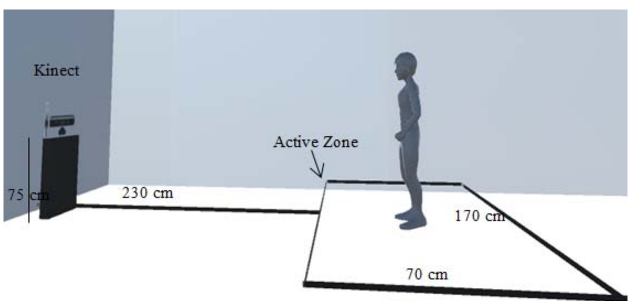

Fig. 4. Kinect locations and active zone in the real environment

To reproduce the same task conditions in the real environment we materialized the outlines of the destination boxes with wire, so as not to disrupt the Kinect sensor's detection. The objects to be grabbed were small Styrofoam shapes of negligible weight, in order to require of subjects no additional physical effort other than the movements they were to perform.

The active zone in both environments was the zone in which the subject was allowed to move to manipulate the gestural interface. Within this area, the subjects could be tracked by the Kinect sensor, which was located in front of the middle box (in the boxes' center). We had indicated the limits of the active zone with markers. The Kinect sensor was used to allow users to interact with the system in conditions as natural as possible. (cf. Fig. 3 and Fig. 4).

In the virtual environment, we used a Wizard-of-Oz technique [32] to simulate picking and dropping objects. When the subject tried to pick an object (by closing their hand), the experimenter pressed a button, and when they tried to drop it (by opening the hand), the experimenter pressed another button. Subjects were unaware, only knowing that the opening / closing movements were responsible for picking and dropping objects. We used this technique to overcome the limitations of Kinect in detecting accurate movement of the wrist.

For the virtual tasks, we chose to use a non-fully immersive environment (no S3D visualization) to avoid additional fatigue that might have been provoked by the environment conditions and tools. Some devices used in these setups may influence fatigue level: for example, the use of an HMD or a flytstick involve additional weight and stress to muscles and joints, which in turn may influence physical fatigue. In addition, the relationship between stereoscopic visualization and physical fatigue is still unclear, which is why we avoided using it, even though some studies show that there is no clear relationship between stereoscopic vision and physical body fatigue (neck / back / shoulder pain). [33][34][35].

D. Data Collection

The comparison of fatigue levels is done through collecting and analyzing fatigue assessment results for each condition, per joint, in both environments. Two categories of data were collected:

- Objective data collected by the system (cf. 3. Our approach to assessing gestural fatigue): the system detected gesture variables, and analyzed them to evaluate the level of fatigue associated with the gestures, determining whether they were acceptable. Evaluation results for each joint according to each standard were logged every 0.5 second, thus yielding results for ISO (combined with AFNOR) and RULA. ISO and AFNOR were combined into one result because AFNOR is based on ISO specifications, and merely adds some additional factors like repetition of movement. Additional data was also collected by the software, such as information about the subject (name, date of birth, etc.), and information about the task (condition, duration, order, etc.)
- Subjective data: Subjects filled in a questionnaire about their level of fatigue for each body part, the technical and cognitive difficulties they experienced, as well as their physical exercise capabilities. They could also add comments and remarks about the experiment. We used a six-point Likert-type scale for subjective evaluation (from 0 for absence of fatigue to 5 for extreme fatigue) [36].

V. Results and Discussion

A. Statistical analysis

We aimed to study the difference, if any, of the means in fatigue for each joint in both environments. For each joint, we compared three means: ISO and RULA means, given by the assessment systems, as well as a subjective fatigue level mean calculated from the questionnaire data. To compare the pairs of means, we used a Wilcoxon test whose null hypothesis was that there was no difference between the levels of fatigue in both environments. We considered the null hypothesis to be refuted if the test yielded a p-value lesser than 0.05. A Wilcoxon test was used because results did not follow a normal distribution.

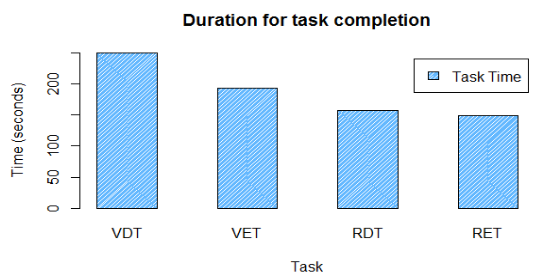

Fig. 5. Average duration for each performed tasks; VDT: virtual difficult task, VET: virtual easy task, RDT: real difficult task, RET: real easy task.

To study the relationship between evaluation results obtained in virtual and real environments, we used a Spearman linear correlation test. The correlation coefficient ρ was computed for each indicator (ISO, RULA and subjective) in both environments, with a level of confidence set to $p < 0.05$.

B. Duration of task completion

The duration of tasks performed in the virtual environment was higher than that in the real environment (cf. Fig. 5). This difference may be related to the perceptive differences between both environments: the distance of the display from subjects display location and the haptic feedback were not the same in both environments. This increased the aiming time for the destination box in the virtual environment, where only visual feedback was available. However, in the real environment, both visual and haptic feedbacks were available.

On the other hand, the use of technology itself may have disturbed the subjects, because of some necessary learning time, or because subjects may have slowed down their movements to be sure that they were correctly tracked. In addition, the familiarity with real-world tasks and objects allowed subjects to learn and perform quickly, which in turn decreased the completion time in the real environment compared to the virtual condition.

Fig. 5 shows the average duration for each condition. For the difficult task, in the virtual environment the average was 250 seconds, but in the real environment it was only 158 seconds. The difference was statistically significant ($p < 0.01$). For the easy task, in the virtual environment, the duration was 194 seconds, however in the real environment it was only 150 seconds. Again, this difference was statistically significant ($p < 0.01$).

TABLE I. Average levels of fatigue for right shoulder, right elbow, right wrist and neck in both real and virtual environment for "difficult" tasks, given by subjects, as well as system data (ISO and RULA).

Difficult real task – Difficult virtual task								
Body Part		***ISO***		***RULA***		***Subjective***	
		RDT	***VDT***	***RDT***	***VDT***	***RDT***	***VDT***
Right Shoulder	Mean (SD)	0.648 (0.08)	0.771 (0.12)	0.322 (0.04)	0.398 (0.08)	0.331 (0.2)	0.454 (0.24)
	P-Value	*0.00088*		*0.00088*		*0.00796*	
Right Elbow	Mean (SD)	0.768 (0.10)	0.885 (0.13)	0.595 (0.02)	0.629 (0.04)	0.154 (0.21)	0.169 (0.21)
	P-Value	*0.00251*		*0.00251*		0.7786	
Right Wrist	Mean (SD)	NA	NA	0.517 (0.08)	0.54()	0.108 (0.13)	0.154 (0.17)
	P-Value	NA		0.2326		0.09534	
Neck	Mean (SD)	0.032 (0.06)	0.419 (0.3)	0.227 (0.04)	0.334 (0.16)	0.1 (0.15)	0.107 (0.14)
	P-Value	*0.00009*		*0.00096*		0.803	
Trunk	Mean (SD)	0.438 (0.13)	0.292 (0.32)	0.389 (0.03)	0.362 (0.08)	0.069 (0.11)	0.062 (0.09)
	P-Value	*0.02679*		*0.00236*		1	

C. Assessment of fatigue

Table 1 and Table 2 show average levels of fatigue for some joints, for difficult and easy tasks in both the real and virtual environments. The studied joints are the right shoulder, the right elbow, the right wrist, the neck, and the trunk; we chose these body parts because we think that they were the

TABLE II. Average levels of fatigue for right shoulder, right elbow, right wrist, neck and trunk in both real and virtual environment for "easy" tasks, given by subjects, as well as system data (ISO and RULA).

Easy Real task – Easy virtual task							
Body Part		ISO		RULA		Subjective	
		ERT	EVT	ERT	EVT	ERT	EVT
Right Shoulder	Mean (SD)	0.436 (0.08)	0.348 (0.18)	0.212 (0.02)	0.209 (0.03)	0.154 (0.14)	0.192 (0.12)
	P-Value	0.8		0.4352		0.212	
Right Elbow	Mean (SD)	0.847 (0.07)	0.888 (0.14)	0.619 (0.02)	0.651 (0.05)	0.069 (0.14)	0.176 (0.12)
	P-Value	*0.0462*		*0.00017*		*0.00315*	
Right Wrist	Mean (SD)	0	0	0.489 (0.05)	0.512 (0.09)	0.138 (0.4)	0.146 (0.17)
	P-Value	NA		0.1874		0.2925	
Neck	Mean (SD)	0.019 (0.03)	0.202 (0.28)	0.243 (0.05)	0.262 (0.13)	0.038 (0.08)	0.038 (0.08)
	P-Value	*0.00015*		0.5539		1	
Trunk	Mean (SD)	0.502 (0.13)	0.138 (0.16)	0.408 (0.04)	0.342 (0.02)	0.031 (0.07)	0.023 (0.07)
	P-Value	*0.00002*		*0.00001*		0.7728	

most active body parts during task performance, considering what was requested of subjects. The studied averages are the assessment results provided by our system (RULA and ISO) and the assessment results given by subjects. As those tables show, most results (except for the Assessment results should be discussed for each joint.

1) Shoulder fatigue

For the right shoulder, in the difficult task, it is clear that the average level of fatigue is significantly higher in the virtual environment than the real-world setting (cf. Fig. 6). We think that this difference is due to the fact that the duration of the virtual task was greater than that of the real-world task (cf. 5.2. Duration of task completion). subjects spent more time performing the virtual task. According to the standards and methods on which our system is based, the more joints are involved, the higher the level of fatigue. Subjective results confirm this hypothesis: subjects are not familiar with manipulating virtual objects, where the only feedback is visual, and this, in turn, demands more time and concentration from subjects. The shoulder is the most active joint during the task, that's why we think that it is most affected by an increase in duration.

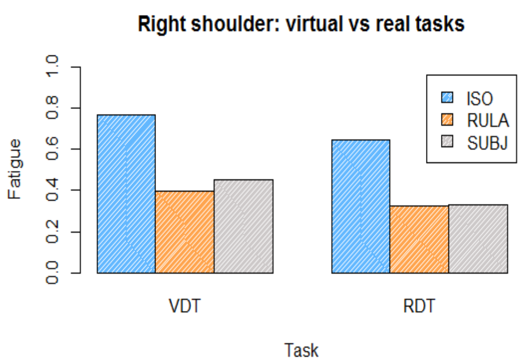

Fig. 6. Fatigue levels for the right shoulder in the difficult tasks in both virtual (VDT) and real environments (RDT) according to ISO, RULA and subjects (SUBJ)

On the other hand, in the easy task, the difference in fatigue levels between the virtual and real environments was not statistically significant. We think that the shoulder was less active in the easy task and the difference in fatigue levels was not clear-cut. In addition, the elbow is more involved and tired in the easy task, so subjects may have focused more on their elbow fatigue than their shoulder's.

2) Elbow fatigue

We notice that the average levels of fatigue are always greater in the virtual environment than the real-world setting. In the difficult task, we see that the difference was significant according to system measures (ISO and RULA) and not significant according to subject evaluation (cf. Fig. 7). We think that this difference is due to the longer duration of the task in the virtual environment compared to that of the real environment, which affects the period of time for which the elbow is used. We also think that subjects may not feel this difference in fatigue despite its existence (thus indicating that the system's evaluation is more reliable than that of subjects'). It is also possible that subjects do not detect the difference because the shoulder was more involved (and therefore more tired) than the elbow in said task, which made subjects concentrate more on the former joint. In the easy task, the difference is clearer; the differences in fatigue levels were significant according to all measures (ISO, RULA and subjects). In this task the elbow is more involved than in the difficult setting, which makes it easier for subjects to feel the difference in fatigue between the two environments. Another factor which may make the fatigue difference clearer is the very low fatigue level in the easy task in the real environment, as well as a task which was very simple and easy to learn, compounded with the longer duration in the virtual environment. In the virtual environment, subjects had a tendency to limit their movements as much as possible: they were trying to reach objects by extending their arms instead of moving closer to the box. This involved, in turn, more extensive use of their elbow in the virtual environment and, as a consequence, increased its level of fatigue.

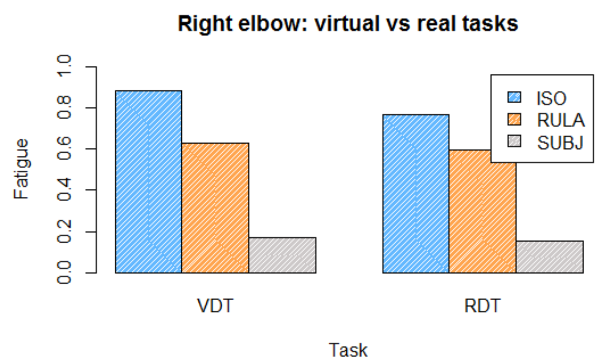

Fig. 7. Fatigue levels for the right elbow in the difficult task in both virtual (VDT) and real environments (RDT) according to ISO, RULA and subjects (SUBJ).

3) Neck fatigue

For the neck, the fatigue levels in the virtual environment were always higher than those in the real environment. The difference was also clearer in the difficult task were the fatigue level could be felt easily. In the difficult task, the differences in fatigue levels between virtual and real environments were significant in the system measures (ISO and RULA) but not in subjects' evaluation (cf. Fig. 8). In this task, the neck joint angles often exceeded the acceptable angular levels according to the specifications on which our system is based. The longer task duration in the virtual environment made this all the more tiring. In addition, we have noticed that the fast and easy learning process of a task done in a real environment may minimize its fatigue level: indeed, some subjects quickly learned how to perform the task in a real environment and sometimes continued without even looking at the destination boxes. This may have reduced the involvement of the neck joint and thus decreased its fatigue level. In the easy task, we think that the level of fatigue was low and by consequence, the difference was not clear.

4) Fatigue in other joints

For other joints, such as for the wrist, we notice that there is no significant difference in fatigue levels between the virtual and real environments. We think this is because of the nature of the wrist movements in the task. The wrist is used quickly (closing the hand for picking up the object and opening the hand for dropping it), so the time difference

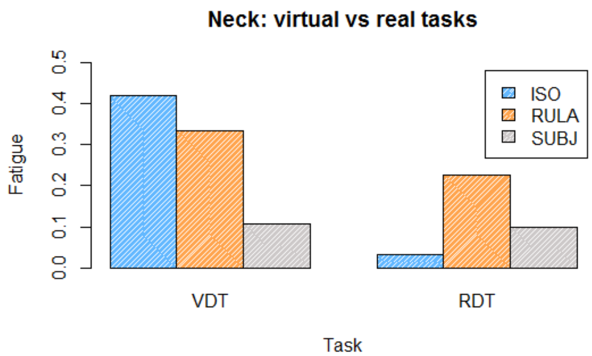

Fig. 8. Fatigue levels for the neck in the difficult task in both virtual (VDT) and real environments (RDT) according to ISO, RULA and subjects (SUBJ).

5) Fatigue in other joints

For other joints, such as for the wrist, we notice that there is no significant difference in fatigue levels between the virtual and real environments. We think this is because of the nature of the wrist movements in the task. The wrist is used quickly (closing the hand for picking up the object and opening the hand for dropping it), so the time difference did not affect the fatigue level for this body part. In addition, this movement is very simple and is often used in daily life; our experiment does not apply any additional technical constraints because we use the Wizard-of-Oz technique (see IV.C. Apparatus).

We also found that the fatigue level in the real environment was higher than that in the virtual environment for the trunk (cf. Fig. 10). The trunk was not really involved in the movements required by the tasks, and it was used in the same way in both environments. We think that this difference was due to the experimental setup, especially in the real-world objects installation. In the real environment, boxes were closer to subjects, which made some subjects arch their trunk backwards in order to look at the destination boxes and avoid colliding with them. In addition, some subjects were performing the real-world task without moving, which required that they bend their back to reach the boxes, cf. Fig. 9.

6) Analysis

To summarize our previous findings, we have noticed that fatigue levels were generally higher in a virtual environment for most joints (except for the trunk, cf. 5.3.4. Fatigue in other joints). This difference was sometimes evident and clear to subjects, sometimes not. This was probably because the low level of fatigue in some tasks made it difficult for subjects to notice a difference. This would indicate that our system is able to detect fatigue levels that cannot be detected by subjects. The specifications on which our assessment system is based consider any movement that may potentially cause a musculoskeletal problem as stressful. It is therefore possible that this assessment is more sensitive than subjects'.

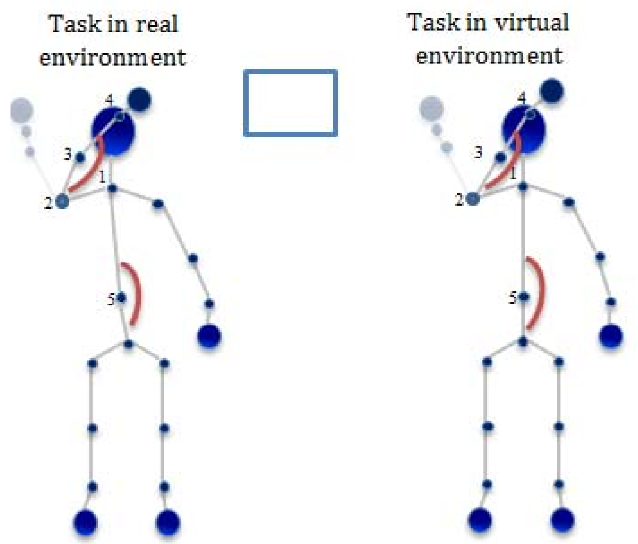

Fig. 9. Differences in back positions in a real and virtual environment

In that same perspective, we think that subjects may sometimes underestimate their fatigue probably because of their psychological situation (pleasure, amusement, familiarity with tasks, stress, etc.) or because they were focused on another, more stressed joint. Such factors are not taken into consideration during the assessment process, even if they are indirectly included in the subjects' evaluation. We hope to integrate them in the assessment process in the future.

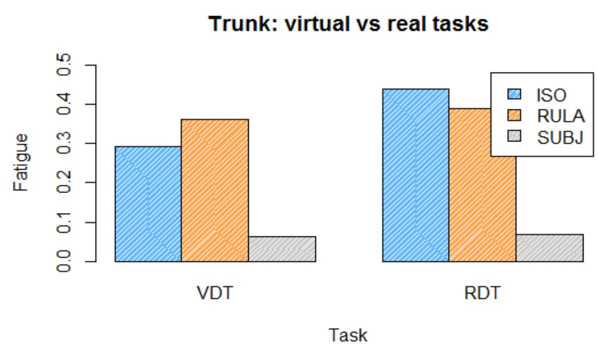

Fig. 10. Fatigue levels for the trunk in the difficult task in both virtual (VDT) and real environments (RDT) according to ISO, RULA and subjects (SUBJ)

We think that a clear difference in task durations for the difficult tasks makes it easier to detect and feel a difference in fatigue levels. In the easy task, on the other hand, the difference was small.

We studied the linear correlation of our ISO, RULA, and subjective evaluation results between virtual and real environments. For all studied joints, we found a good to very good correlation between subjective appreciations of fatigue in virtual and real environments (cf. Table 3). However, little to no correlation was found in the other assessment results (ISO and RULA). Objective assessment results in virtual tasks are therefore not correlated to those in real tasks, thus potentially indicating one more reason why the use of virtual reality to assess real tasks must seriously be put to question.

TABLE III. CORRELATION RESULTS BETWEEN FATIGUE LEVELS IN VIRTUAL AND REAL ENVIRONMENTS

Body Part	Task	ISO		RULA		Subjective	
		Rho	P-value	Rho	P-value	Rho	P-value
Shoulder	difficult	0.31	0.130	0.08	0.692	0.54	0.004
	easy	0.34	0.088	0.41	0.040	0.40	0.046
Elbow	difficult	0.02	0.935	0.14	0.507	0.50	0.011
	easy	0.33	0.094	0.22	0.274	0.47	0.015
Neck	difficult	0.34	0.094	0.11	0.584	0.57	0.002
	easy	0.28	0.159	0.19	0.360	0.75	<0.001
Trunk	difficult	0.14	0.504	0.16	0.432	0.77	<0.001
	easy	0.25	0.210	0.10	0.617	0.51	0.007
Wrist	difficult			0.36	0.068	0.66	0.003
	easy			0.31	0.120	0.27	0.018

As mentioned before (cf. 2.1. Virtual assessment of real-life gestures), the issue of fatigue differences between virtual and real environments have been studied by [15], among others. Their results show that, according to RULA, gestural fatigue in a real environment was greater than that in a virtual environment. However, subjects' results show that fatigue in the virtual environment was greater than that of the real environment.

We think that the major reason for the divergence in our findings has to do with our experimental setup: we tried to reproduce the same conditions in both virtual and real environments. Subjects used the same gestures and movements to perform the task; despite this, there were still some differences in gestures. Other studies have partially studied the issue by evaluating only one joint [16][17], when we found that differences may differ from one joint to another, hence the importance of studying and analyzing all involved joints.

VI. CONCLUSION AND FUTURE WORK

Gestural interactions promise well in many domains, especially in virtual reality where easier, more intuitive and more natural interactions with extended and unlimited vocabulary are needed. Since this type of interactions may involve the whole of the human body in the interaction process (as is the case in real-world interaction), it is important to study the potential fatigue that may affect all joints in this type of interaction. One of the major understudied issues of gestural interaction is its physical fatigue. That is why we chose to study how to assess said fatigue in order to avoid it and help designing more comfortable, restful gestures.

We studied some physical fatigue assessment methods that are used in other domains, such as the workplace, and then proposed an assessment approach that is based on a set of specifications of standards and existent methods that deal with fatiguing movements. In addition, we studied the potential difference between the fatigue provoked by gestures in a virtual environment compared to those performed in a real-world setting. We designed an experiment where subjects performed the same tasks in virtual and real environments. We found that the fatigue level in a virtual environment was very often higher than it was in a real-world context. We think that this is because of the important variation of some conditions related to performing tasks in virtual or real environments, such as duration, feedback (or lack thereof), presence (or lack thereof), etc. These findings make the issue of physical fatigue and its assessment in virtual interactions more important, because this fatigue not only exists, but it is more severe than in real life. This can also help us to know how we can use virtual interactions applications better, such as the use of virtual environments for assessing real task ergonomics.

As future work, we think about improving our assessment method by integrating other specifications, using a more accurate detection technique, and evaluating more complex gestures. That being said, we also envision integrating other factors in the assessment process such as psychological and cognitive factors and study their effects on the assessment process.

REFERENCES

[1] M. Rauterberg, "From gesture to action: natural user interfaces," in *Mens-machine interactie : Diesrede 1999*, Technical University of Eindhoven, Ed. Technical University of Eindhoven, 1999, pp. 15–25.

[2] T. Baudel and M. Beaudouin-Lafon, "Charade: remote control of objects using free-hand gestures," *Commun. ACM*, vol. 36, no. 7, pp. 28–35, 1993.

[3] D. Saffer, *Designing gestural interfaces: Touchscreens and interactive devices*. O'Reilly Media, Inc., 2008.

[4] G. Kurtenbach and E. A. Hulteen, "Gestures in human-computer communication," *art human-computer interface Des.*, pp. 309–317, 1990.

[5] A. Lalumière and C. Collinge, *Revue de littérature et avis d'experts sur les troubles musculo-squelettiques associés à la souris d'ordinateur*. Québec: Institut de recherche en santé et en sécurité du travail du Québec, 1999.

[6] X. Yan and N. Aimaiti, "Gesture-based interaction implication for the future and," master thesis, Department of Computer Science, Umea University, Umea, Sweden, 2011.

[7] D. A. Bowman *et al.*, "New directions in 3D user interfaces.," *IJVR*, vol. 5, no. 2, pp. 3–14, 2006.

[8] W. T. Freeman and C. Weissman, "Television control by hand gestures," in *Proc. of Intl. Workshop on Automatic Face and Gesture Recognition*, 1995, pp. 179–183.

[9] T. Isenberg and M. Hancock, "Gestures vs. Postures: 'Gestural' Touch Interaction in 3D Environments," in *Proceedings of the CHI Workshop on''The 3rd Dimension of CHI: Touching and Designing 3D User Interfaces''(3DCHI 2012, May 5, 2012, Austin, TX, USA)*, 2012, pp. 53–61.

[10] D. A. Sparks, L. M. Coughlin, and D. M. Chase, “Did too much Wii cause your patient’s injury?,” *J. Fam. Pract.*, vol. 60, no. 7, pp. 404–410, 2011.

[11] C. Jones and B. Hammig, “Case report: injuries associated with interactive game consoles: preliminary data,” *Phys. Sportsmed.*, vol. 37, no. 1, pp. 138–140, 2009.

[12] S. Ahmed, L. Leroy, and A. Bouaniche, “Study of Stressful Gestural Interactions an Approach for Assessing their Negative Physical Impacts,” in *Proc. 11 th Intl Conf. Disability, Virtual Reality & Associated Technologies*, 2016.

[13] CLARTE, “Ergowide : Conception ergonomique des postes de production.” [Online]. Available: http://www.clarte.asso.fr/realite-virtuelle.php/ERGOWIDE3/. [Accessed: 20-Jul-2016].

[14] [HRV, “ERGAUDIT : Evaluer l’ergonomie de vos postes de travail dans votre environnement de production.” [Online]. Available: http://www.hrv-simulation.com/ergaudit.html. [Accessed: 22-Jul-2016].

[15] C. Pontonnier, A. Samani, M. Badawi, P. Madeleine, and G. Dumont, “Assessing the Ability of a VR-Based Assembly Task Simulation to Evaluate Physical Risk Factors,” *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 5, pp. 664–674, 2014.

[16] B. Hu, L. Ma, W. Zhang, G. Salvendy, D. Chablat, and F. Bennis, “Predicting real-world ergonomic measurements by simulation in a virtual environment,” *Int. J. Ind. Ergon.*, vol. 41, no. 1, pp. 64–71, 2011.

[17] L. E. Whitman, M. Jorgensen, K. Hathiyari, and D. Malzahn, “Virtual reality: its usefulness for ergonomic analysis,” in *Proceedings of the 36th conference on Winter simulation*, 2004, pp. 1740–1745.

[18] M. Aptel, F. Cail, and A. Aublet-Cuvelier, “Les troubles musculosquelettiques du membre supérieur (TMS-MS)- Guide pour les préventeurs,” 2011.

[19] M. Nielsen, M. Störring, T. B. Moeslund, and E. Granum, “A procedure for developing intuitive and ergonomic gesture interfaces for man-machine interaction,” in *Proceedings of the 5th International Gesture Workshop*, 2003, pp. 1–12.

[20] S. Simoneau, M. St-Vincent, and D. Chicoine, “Les LATR, mieux les comprendre pour mieux les prévenir,” St-Léonard, Québec: Association paritaire pour la santé et la sécurité du travail Secteur Fabrication de produits en métal et de produits électriques, Québec, 1996.

[21] L. Muse and S. C. Peres, “The Ergonomic Implications of Gesturing: Examining Single and Mixed Use with Appropriate Placement,” in *Texas Regional Human Factors & Ergonomics Society Symposium*, 2011.

[22] J. A. Cameron, “Assessing work-related body-part discomfort: current strategies and a behaviorally oriented assessment tool,” *Int. J. Ind. Ergon.*, vol. 18, no. 5, pp. 389–398, 1996.

[23] L. McAtamney and E. N. Corlett, “RULA: a survey method for the investigation of work-related upper limb disorders,” *Appl. Ergon.*, vol. 24, no. 2, pp. 91–99, 1993.

[24] V. Ha, K. M. Inkpen, R. L. Mandryk, and T. Whalen, “Direct intentions: the effects of input devices on collaboration around a tabletop display,” in *First IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP’06)*, 2006, p. 8--pp.

[25] C. Long, P. W. Conrad, E. A. Hall, and S. L. Furler, “Intrinsic-extrinsic muscle control of the hand in power grip and precision handling,” *J Bone Jt. Surg Am*, vol. 52, no. 5, pp. 853–867, 1970.

[26] A. Freivalds, *Biomechanics of the upper limbs: mechanics, modeling and musculoskeletal injuries*. CRC press, 2011.

[27] ISO International Organization for Standardization, *Ergonomics-Evaluation of Static Working Postures*. International Organization for Standardization, 2000.

[28] CEN (Comité Européen de Normalisation), “Evaluation of Working Posture in Relation to Machinery - prEN 1005-4,” *Saf. Mach. – Hum. Phys. Perform. – Part 4 Eval. Work. posture Relat. to Mach.*, 1998.

[29] M. Aptel, A. Gerling, and F. Cail, “Méthode de prévention des troubles musculosquelettiques du membre supérieur et outils simples,” Lorraine, 2000.

[30] C. Zerpa, C. Lees, P. Patel, E. Pryzsucha, and P. Patel, “The Use of Microsoft Kinect for Human Movement Analysis,” *Int. J. Sport. Sci.*, vol. 5, no. 4, pp. 120–127, 2015.

[31] S. Ahmed, L. Leroy, and A. Bouaniche, “Reducing Gestural Interaction Physical Stress in Virtual an Experiment,” in *13th International Conference on Human Computer Interaction, RoCHI 2016*, 2016.

[32] N. Beringer, “Evoking gestures in SmartKom-Design of the graphical user interface,” in *International Gesture Workshop*, 2001, pp. 228–240.

[33] S. Yang *et al.*, “Stereoscopic viewing and reported perceived immersion and symptoms,” *Optom. Vis. Sci.*, vol. 89, no. 7, pp. 1068–1080, 2012.

[34] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, “Vergence--accommodation conflicts hinder visual performance and cause visual fatigue,” *J. Vis.*, vol. 8, no. 3, p. 33, 2008.

[35] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks, “The zone of comfort: Predicting visual discomfort with stereo displays,” *J. Vis.*, vol. 11, no. 8, p. 11, 2011.

[36] R. Likert, “A technique for the measurement of attitudes.,” *Arch. Psychol.*, vol. 22, no. 140, pp. 5–55, 1932.