Navigation and interaction in a real-scale digital mock-up using natural language and user gesture

M. A. Mirzaei, J.-R. Chardonnet, F. Mérienne
Arts et Métiers ParisTech – CNRS Le2i
Institut Image
2 rue Thomas Dumorey
71100 Chalon-sur-Saône, France
{ali.mirzaei, jean-remy.chardonnet, frederic.merienne}@ensam.eu

A. Genty
Coordinator of the CALLISTO-SARI project
Bouygues Bâtiment International
1 rue Eugene Freyssinet
78061 Saint Quentin en Yvelines, France
ar.genty@bouygues-construction.com

ABSTRACT
This paper tries to demonstrate a very new real-scale 3D system and sum up some firsthand and cutting edge results concerning multi-modal navigation and interaction interfaces. This work is part of the CALLISTO-SARI collaborative project. It aims at constructing an immersive room, developing a set of software tools and some navigation/interaction interfaces. Two sets of interfaces will be introduced here: 1) interaction devices, 2) natural language (speech processing) and user gesture. The survey on this system using subjective observation (Simulator Sickness Questionnaire, SSQ) and objective measurements (Center of Gravity, COG) shows that using natural languages and gesture-based interfaces induced less cyber-sickness comparing to device-based interfaces. Therefore, gesture-based is more efficient than device-based interfaces.

Categories and Subject Descriptors
H.1.2 [**Models and principles**]: User/Machine Systems – *human factors, human information processing.*
H.5.2 [**Information interfaces and presentation**]: User Interfaces – *evaluation/methodology, input devices and strategies, interaction styles, voice I/O*.

General Terms
Algorithms, Measurement, Design, Experimentation, Performance, Verification, Human Factors

Keywords
Gesture perception, speech processing, Interaction/Navigation interfaces, real-time signal processing, and full-scale immersive 3D system.

1. INTRODUCTION
The construction field is currently undergoing many environmental (new regulations, energy constraints, etc.) and industrial (better processes) changes. The changes require moving from the use of 2D plans to CAD (Computer Aided Design) 3D Building Information Model (BIM). The BIM includes semantics into the construction process (e.g., structure, air conditioning/ventilation, mechanical, electrical, plumbing, etc.) and data for simulation (materials/structure resistance, energy consumption, thermal calculations, lighting, acoustic simulations, etc.). These issues must be addressed throughout the construction project but mainly, at the beginning during the design phase to fulfill the customer's requirements, during the construction work to anticipate technical constraints on site, and during the maintenance phase to control the building. Introducing 3D models in the construction process is a main way to (1) test virtually and correct a construction project before the realization, (2) reduce costs due to the construction of a real mock-up, (3) avoid mistakes on site that generates material wastes. The CALLISTO-SARI project, whose consortium is composed of: one large group, Bouygues Bâtiment International (leader), two small companies, Immersion S.A., Art Graphique et Patrimoine, two semi-public institutions, Universcience (Cité des Sciences et de l'Industrie), le Centre des Sciences et Techniques du Bâtiment, and three research laboratories, Arts et Métiers ParisTech, Ecole Centrale Paris, le Laboratoire des Usages en Technologies d'Information Numérique – Paris 8, came about for two main reasons: the need to communicate or share efficiently 3D visualizations and the need to help construction actors and public to use 3D CAD model. The final immersive room, shown in Figure 1, was installed in a science and technology museum in Paris. Different challenges related to VR technologies have been discussed in the literature such as usability, precision, delay between command and visual feedback, cyber sickness, and so on. Interaction and navigation in a 3D immersive environment bring out technical challenges such as: how to implement comfortable interactions systems, how to overcome well-known problems of cyber-sickness and general discomfort? How to clearly identify the origin of discomfort or sickness and eliminate them or minimize their effects by proper design? This paper shows how different navigation interfaces can address these issues.

Laval Virtual VRIC'14, April 9-11, 2014, Laval, France.

Figure 1. Callisto immersive room in operation.

1.1 Related work
Different navigation and interaction interfaces have been developed and evaluated since the first virtual environment was

established. We make a short review over some of the recently-developed interfaces and their evaluation methods.

Navigation/Interaction HMI: the very first navigation interfaces were all based on electro-mechanical devices [28]. These devices provided a set of rich hardware resources for the developers however most of the time they were not handy and portable for Virtual Reality (VR) end-users. The recently developed devices are even more sophisticated [10], however they provide better ways of navigation and interaction with virtual environments and allow easier immersion of the user in VR. Despite all advantages of these devices, they make the physical movement of the users strictly limited. For instance, omnidirectional treadmills [1,4] were proposed to allow cyber walking inside a real-scale virtual environment (VE) feasible. But too many physical constraints are needed to be imposed to avoid damages and respect safety regulations. In parallel with cyber walk by omnidirectional treadmills (CWOT), walking in place (WIP) [26] proposed to make navigation in virtual environments by human natural locomotion feasible. Later, some concepts like Walking in Place by Cyber Carpet (WIPCC) [3] were proposed to perform the walking concept in VEs. Chris Hand [5] made a fairly complete surveillance on different interaction interfaces that had been developed until late 1997. The interaction interfaces that have been developed after this date were mostly gesture-based. Interfaces proposed by [20] and [21] are two good examples for this type. In addition to gesture, speech processing has been the topic of research in Human Machine Interaction (HMI) since the development of the first sound processing electronics peripherals [25]. Since a couple of years back, speech processing stepped practically in VR little by little to make interaction in virtual environments even more natural. We will talk about this technique more in this paper later.

Evaluation method: totally in computational machinery, the first aspect is to develop a computer based application or a system, and the second aspect is to show this system has no cyber effects and is usable by the end users (evaluation). The evaluation relays on human factors, psychological, cognitive research methodologies and experimental data analysis. These methodologies have two distinctive parts: 1) self-report questionnaires (psychological measurements) 2) psychophysiological measurements. Different standard questionnaires have been proposed by psychologists and neuroscientists for subjective studies in different applications (see Kennedy SSQ [8], Slater's questionnaire [23], NASA questionnaire [16]). Usually, verifying the hypothesis by the results from psychological data analysis is not very reliable. For that reason, psychophysiological measurements are recorded, analyzed and interpreted to find a correlation between two data sets and to discover subject independent factors for future studies. These factors make evaluation methods more parameter dependent. Blood pressure (BP), heart rate (HR) [6], skin resistance (SR) [20], center of gravity (COG) [22], eye movement (EM) [27], pupil dilation (PD) [19], electromyography (EMG) [7], electrogastrography (EGG) [17], electrooculography (EOG) [15], video-oculography (VOG) [13], electroencephalograph (EEG) [9] are some of the most used psychophysiological measurements.

1.2 Contribution

This paper has two main contributions: (**1**) it introduces speech processing (natural HMI) as an efficient interaction interface for object manipulation. (**2**) It explains two navigation interfaces based on user gesture (natural navigation interface: NNI) and electro-mechanical devices (device-based navigation interface: DBNI). Our hypothesis is: "***NNI is better than DBNI in terms of cyber-sickness***".

The paper is organized as follows: the navigation/interaction interfaces will be introduced in Section 2. Walking gesture (NNI) and fly-stick (DBNI) will be selected to establish the test-bench in Section 3. Section 4 will present and discuss the results to show how one interface is better than the other.

2. NAVIGATION AND INTERACTION

We will introduce two navigation/interaction interfaces in this section. The navigation interfaces provide a way to perform natural walking (similar to WIP) using the XBOX Kinect (NNI candidate) and move artificially using a Flystick (DBNI candidate). Interaction includes sound menu manipulation using speech processing, natural language interface (NLI) and a Flystick as device-based interaction (DBI). We alternatively will call this interface Speak-to-VR interface in this paper.

2.1 Gesture-based interface (NNI)

FAAST VRPN server [24] is used to define, generate and stream walking gesture in the VR system. The walking gesture was defined in a similar way as in [12]. In belief, if the left/right foot moves up 5 cm above the ground, this gesture will be considered as walking one step forward and the scene is moved backward 2 m/s to give the sensation of forward movement to the user. When the feet are on the ground, the scene remains constant (Figure 2.a).

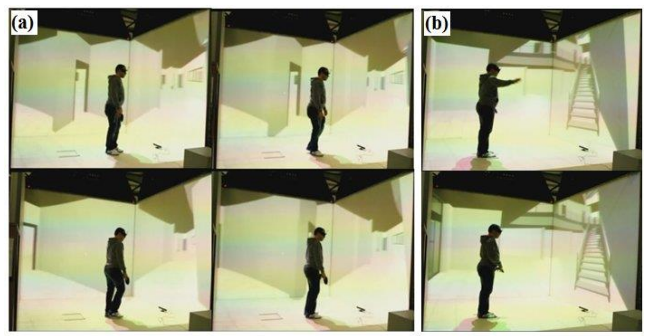

Figure 2. Forward movement using walking gesture (a), counter-clockwise rotation by hand gesture (b).

When the user brings his right hand 10 cm above his right shoulder, the scene starts rotating clockwise around the z-axis of the user (ego-centric rotation) with the speed of 10 deg/s (Figure 2.b). When the hand comes under the shoulder, the rotation is terminated. We use the same threshold and the same gesture for the left hand to rotate counter-clockwise.

2.2 Device-based navigation (DBNI)

The second navigation interface was designed based on the so-called Flystick. The joy-stick on the Flystick is used for both translation and rotation. If the joy-stick is pushed forward, the scene will be moved backward and if it is pushed to the left and right, the scene rotates clockwise and counter-clockwise. The velocity of rotation and translation is set the same as for the first navigation interface.

2.3 Interaction with virtual objects using speech processing (NLI) and Flystick (DBI)

Speech processing technology has progressed recently both in terms of processing speed and accuracy. Microsoft Speech SDK 5.1 [2] improved the previous version of Speech SDK. Now, it also includes freely distributable text-to-speech (TTS) and speech recognition (SR) engines. Here, the Speak-to-VR VRPN server based on Microsoft Speech SDK will be explained in summary.

Microsoft Speech SDK is used with Minimum Variance Distortionless Response (MVDR) algorithm [14] for microphone array to improve its performance. A media player displayed in a virtual environment is developed for interaction purpose (object manipulation interface). In brief, speech processing application processes the speech signal (Figure 3.c) by extracting features in the time and the frequency domains and then matching the closest word in a dictionary. A unique code is being assigned to the word after the recognition. This code is streamed in the VR network by a VRPN server. A function is assigned to each unique word in the client application. The function is executed by the arrival of the code. In this application, we used media play functions such as "Play", "Stop", and so on. The music is played/ stopped by saying "Play", "Stop", for instance (Figure 3.a). The same media player application is developed using a Flystick as an interaction device (Figure 3.b).

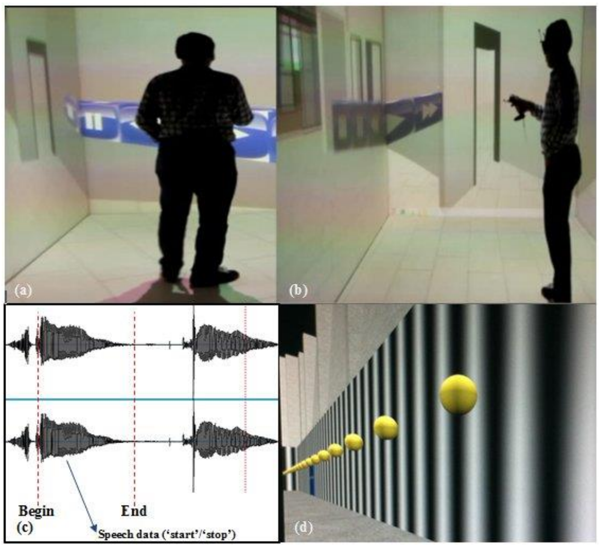

Figure 3. Interaction with a virtual media player using Speak-to-VR interface (a), media player manipulation using a Flystick (b), speech signal detail (c), and test setup and path planning for the navigation evaluation (d).

3. TEST SETUP

The previous section introduced both navigation and interaction interfaces. However, to evaluate interfaces by a subjective study we will consider only two different navigation interfaces: NNI and DBNI. 17 healthy subjects (5 females, 12 males) were selected to navigate along a straight line while they were being exposed to a periodic black and white perturbation stimulus. The frequency of the pattern was set in a value which did not annoy the participants. A model of the interior space of a building was selected to set up a navigation path. The path was arranged in parallel and one meter away from the virtual walls and it was marked up by a set of yellow balls. The test was performed in a 4-sided CAVE system for primary study (Figure 3.d). TECHNO CONCEPT CG sensor [18] used to record the instantaneous center of gravity (COG) of the participants.

4. RESULTS AND DISCUSSIONS

We collected pre- and post-exposure SSQ questionnaires and recorded COG signal for 30 s for each participant for NNI and DBNI. The final sickness score was calculated by the correction factor proposed in Kennedy SSQ [8] for three sub-scores; "Nausea", "Disorientation" and "Oculomotor". Figure 4 shows the difference between NNI and DBNI. Nausea in average has the highest score. The final score is calculated using the sub-score for each participant and the SSQ correction factor. Final scores show a significant difference between NNI and DBNI. We found the associated score of DBNI ($\mu = 350$, $\sigma = 50.63$) is significantly higher than NNI ($\mu = 150$, $\sigma = 25.63$), $p = 0.0032$ ($p<0.005$) and $t(16) = 4.35$.

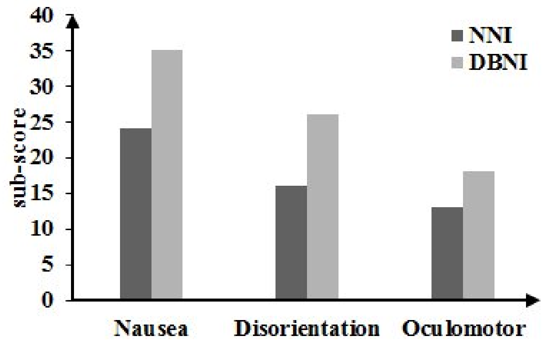

Figure 4. Sub-score calculated from SSQ.

Before this study, we established another experiment [11] to study the effect of different parameters of a navigation mechanism on cyber-sickness. In brief, we used the same test-bench but different subjects to evaluate the navigation mechanisms. The measurement (COG) and psychological (Kennedy SSQ) criteria were the same, however psychophysiological measurements were analyzed in the frequency-domain rather than time-domain. The result shows that the difference between low (LF) and high frequency (HF) components of COG signal is highly correlated ($r = 95.45\%$, $p = 0.00083$) with the level of sickness. Relying on previous results, we can say "*the higher the distance between the LF and HF the worse is the navigation mechanism*". By applying t-test we found $\Delta f = 2.3$Hz for NNI and $\Delta f = 4.56$Hz for DBNI which is significantly higher, $p<0.01$, ($t(16) = 5.67$, $p = 0.0054$) for DBNI than NNI. In turn, it means NNI is a better navigation interface than DBNI. Because it uses more sensory information to generate movements as a result it is more consistent with proprioception and creates less sensory conflicts.

5. CONCLUSION

In this paper, the immersive room, its objective and challenges were briefly introduced. The concept behind Speak-to-VR interface and its application in VEs were explained. Different navigation/interaction interfaces, subjective and objective methods of evaluation were reviewed. COG and Kennedy SSQ were selected as psychophysiological measurement and psychological criteria. NNI and DBNI were taken as two examples of gesture-based and device-based navigation interfaces. We found less sickness in NNI than DBNI ($p<0.005$). Besides, we found less difference between LF and HF components in NNI than DBNI ($p< 0.01$). These two criteria show that gesture-based navigation interface (NNI) works better than the device-based (DBNI).

These techniques are to be ported to the Callisto immersive room. Future work include the effect of the perception of the scale on user's performance during a navigation task, which is another great issue.

6. ACKNOWLEDGMENTS

This work has been done under the FUI CALLISTO-SARI project grant. We would like to be kindly grateful from Bouygues R&D director Trino Beltran for his support in this project.

7. REFERENCES

[1] David EE Carmein. Omni-directional treadmill. Patent, November 28 2000. US Patent 6,152,854.

[2] Eric Chang, Yu Shi, Jian-Lai Zhou, and Chao Huang. Speech lab in a box: a mandarin speech toolbox to jumpstart speech related research. In *INTERSPEECH*, pages 2799–2802, 2001.

[3] Alessandro De Luca, Raffaella Mattone, and Paolo Robuffo Giordano. Acceleration-level control of the cybercarpet. In *IEEE Robotics and Automation (ICRA)*, 2330–2335, 2007.

[4] Alessandro De Luca, Raffaella Mattone, Paolo Robuffo Giordano, and HH Bulthoff. Control design and experimental evaluation of the 2d cyberwalk platform. In *IEEE/RSJ Intelligent Robots and Systems (IROS)*, 5051–5058, 2009.

[5] Chris Hand. A survey of 3d interaction techniques. In *Computer graphics forum*, volume 16, pages 269–281. Wiley Online Library, 1997.

[6] Nis Hjortskov, Dag Rissen, Anne Katrine Blangsted, Nils Fallentin, Ulf Lundberg, and Karen Søgaard. The effect of mental stress on heart rate variability and blood pressure during computer work. *European journal of applied physiology*, 92(1-2):84–89, 2004.

[7] Paul W Hodges and Bang H Bui. A comparison of computer-based methods for the determination of onset of muscle contraction using electromyography. *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control*, 101(6):511–519, 1996.

[8] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.

[9] Chin-Teng Lin, Shang-Wen Chuang, Yu-Chieh Chen, Li-Wei Ko, Sheng-Fu Liang, and Tzyy-Ping Jung. Eeg effects of motion sickness induced in a dynamic virtual reality environment. In *IEEE Engineering in Medicine and Biology Society (EMBS)*, 3872–3875, 2007.

[10] Maud Marchal, Julien Pettré, and Anatole Lécuyer. Joyman: A human-scale joystick for navigating in virtual worlds. In *IEEE 3D User Interfaces (3DUI)*, 19–26, 2011.

[11] M. A. Mirzaei, "Study the effect of the navigation parameters on the level of visually induced motion sickness", Annual technical report, Arts et Métiers ParisTech, May 2013, France.

[12] M Ali Mirzaei, Jean-Rémy Chardonnet, Frédéric Merienne, and Christian Pere. Improvement of the real-time gesture analysis by a new mother wavelet and the application for the navigation inside a scale-one 3d system. In *IEEE Advanced Video and Signal Based Surveillance*, 270–275, 2013.

[13] Carlos H Morimoto and Marcio RM Mimica. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*, 98(1):4–24, 2005.

[14] Manohar N Murthi and Bhaskar D Rao. Minimum variance distortionless response (mvdr) modeling of voiced speech. In *IEEE Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 1687–1690, 1997.

[15] Anselmo Frizera Neto, Wanderley Cardoso Celeste, Vinicius Ruiz Martins, et al. Human-machine interface based on electro-biological signals for mobile vehicles. In *IEEE Industrial Electronics*, volume 4, pages 2954–2959, 2006.

[16] Jan M Noyes and Daniel PJ Bruneau. A self-analysis of the nasa-tlx workload measure. *Ergonomics*, 50(4):514–519, 2007.

[17] M Oinuma, K Hirayanagi, K Yajima, M Igarashi, and Y Arakawa. Changes in cardio-respiratory function, heart rate variability, and electrogastrogram preceding motion sickness-like symptoms induced by virtual reality stimulus. *Japanese Journal of Aerospace and Environmental Medicine*, 41(3):99–109, 2004.

[18] Thierry Paillard, Frederic Noe, Terence Riviere, Vincent Marion, Richard Montoya, and Philippe Dupui. Postural performance and strategy in the unipedal stance of soccer players at different levels of competition. *Journal of athletic training*, 41(2):172, 2006.

[19] Marco Pedrotti, Mohammad Ali Mirzaei, Adrien Tedesco, Jean-Rémy Chardonnet, Frédéric Mérienne, Simone Benedetto, and Thierry Baccino. Automatic stress classification with pupil diameter analysis. *International Journal of Human-Computer Interaction*, 30(3):220-236, 2014.

[20] Gang Ren, Chuan Li, Eamonn O'Neill, and Philip Willis. 3d freehand gestural navigation for interactive public displays. *IEEE Computer Graphics and Applications*, 33(2):47–55, 2013.

[21] Mattias Roupé, Petra Bosch-Sijtsema, and Mikael Johansson. Interactive navigation interface for virtual reality using the human body. *Computers, Environment and Urban Systems*, 43:42–50, 2014.

[22] Takeshi Shimba. An estimation of center of gravity from force platform data. *Journal of Biomechanics*, 17(1):53–60, 1984.

[23] Mel Slater. Measuring presence: A response to the witmer and singer presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 8(5):560–565, 1999.

[24] Evan A Suma, Belinda Lange, Albert Rizzo, David M Krum, and Mark Bolas. Faast: The flexible action and articulated skeleton toolkit. In *IEEE Virtual Reality Conference (VR)*, 247–248, 2011.

[25] Ivan Tashev. Kinect development kit: A toolkit for gesture- and speech-based human-machine interaction [best of the web]. *Signal Processing Magazine, IEEE*, 30(5):129–131, 2013.

[26] Martin Usoh, Kevin Arthur, Mary C Whitton, Rui Bastos, Anthony Steed, Mel Slater, and Frederick P Brooks Jr. Walking> walking-in-place> flying, in virtual environments. In *Siggraph*, volume 99, pages 359–364, 1999.

[27] Gerald Westheimer. Eye movement responses to a horizontally moving visual stimulus. *Archives of Ophthalmology*, 52(6):932, 1954.

[28] J Zhang, Chon Wai Lip, Soh-Khim Ong, and Andrew YC Nee. Development of a shoe-mounted assistive user interface for navigation. *International Journal of Sensor Networks*, 9(1):3–12, 2011.