



HAL
open science

A study on differences in human perception between a real and an AR scene viewed in an OST-HMD

Yuan Gao, Yue Liu, Jean-Marie Normand, Guillaume Moreau, Xue Gao,
Yongtian Wang

► **To cite this version:**

Yuan Gao, Yue Liu, Jean-Marie Normand, Guillaume Moreau, Xue Gao, et al.. A study on differences in human perception between a real and an AR scene viewed in an OST-HMD. *Journal of the Society for Information Display*, 2019, pp.1-17. 10.1002/jsid.752 . hal-01987255

HAL Id: hal-01987255

<https://inria.hal.science/hal-01987255v1>

Submitted on 21 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A study on differences in human perception between a real and an AR scene viewed in an OST-HMD

Yuan Gao^{1,2} | Yue Liu^{1,3} | Jean-Marie Normand²  | Guillaume Moreau²  | Xue Gao¹ | Yongtian Wang^{1,3}

¹Beijing Engineering Research Center of Mixed Reality and Advanced Display, School of Optics and Photonics, Beijing Institute of Technology, Beijing, China

²Ecole Centrale de Nantes, AAU UMR CNRS1563-INRIA Hybrid

³AICFVE of Beijing Film Academy, Beijing, China

Correspondence

Yue Liu, Beijing Engineering Research Center of Mixed Reality and Advanced Display, School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China; AICFVE of Beijing Film Academy, 4, Xitucheng Rd, Haidian, Beijing, China.
Email: liuyue@bit.edu.cn

Yuan Gao, Beijing Engineering Research Center of Mixed Reality and Advanced Display, School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China; Ecole Centrale de Nantes, AAU UMR CNRS1563-INRIA Hybrid.
Email: gaoyuan1988.ok@163.com

Funding information

National Key Research and Development Program of China, Grant/Award Number: 2016YFB0401202; National Natural Science Foundation of China, Grant/Award Number: 61661146002

Abstract

With the recent growth in the development of augmented reality (AR) technologies, it is becoming important to study human perception of AR scenes. In order to detect whether users will suffer more from visual and operator fatigue when watching virtual objects through optical see-through head-mounted displays (OST-HMDs), compared with watching real objects in the real world, we propose a comparative experiment including a virtual magic cube task and a real magic cube task. The scores of the subjective questionnaires (SQ) and the values of the critical flicker frequency (CFF) were obtained from 18 participants. In our study, we use several electrooculogram (EOG) and heart rate variability (HRV) measures as objective indicators of visual and operator fatigue. Statistical analyses were performed to deal with the subjective and objective indicators in the two tasks. Our results suggest that participants were very likely to suffer more from visual and operator fatigue when watching virtual objects presented by the OST-HMD. In addition, the present study provides hints that HRV and EOG measures could be used to explore how visual and operator fatigue are induced by AR content. Finally, three novel HRV measures are proposed to be used as potential indicators of operator fatigue.

KEYWORDS

augmented reality, ECG, EOG, human perception, optical see-through HMD, visual fatigue

1 | INTRODUCTION

Based on the development of stereoscopic technologies, augmented reality (AR) is evolving and attracting more and more attention. AR refers to the set of technologies that can superimpose digital information onto real environments and present them to observers.¹ As AR can enhance humans' perception of the real world, it can be

used widely in different areas, such as education, the industry, medical sciences, and the military.

Among the numerous display options for AR systems, head-mounted displays (HMDs) have the advantages of leaving the users' hands free and of adding virtual information to real environments from the viewer's point of view. There are two main types of HMDs for AR applications: (1) *video see-through* head-mounted displays

(VST-HMDs) and (2) *optical see-through* head-mounted displays (OST-HMDs).² VST-HMDs compose the virtual objects into a video stream that is captured by cameras attached to the HMD. In contrast, OST-HMDs receive the information from an optical element, which can combine the virtual objects with an optical view of the real world. However, there are also many limitations regarding the AR-HMD technology. Since most of the HMDs have a fixed focal length, a focal rivalry will arise when the depth of the virtual objects is different from the focal length, which will influence users' perception in AR. Indeed, some researchers have come up with ideas to deal with the problem, such as using free-form design.³ However, there is still a need for measuring methods for perception issues in AR. In this work, we focus on perception judgment in OST-HMDs.

In the present study, a comparative experiment with two tasks was designed to investigate differences in human perception when watching virtual objects presented in an OST-HMD, compared with watching real objects through the OST-HMD. In order to measure visual or operator fatigue, we used the scores of subjective questionnaires before and after the tasks, as well as other objective measures. Our results show that some subjects suffered more from visual and operator fatigue after watching virtual objects presented by the OST-HMD than watching the real objects. Moreover, we found that some physiological measures could serve as effective indicators of operator fatigue.

2 | RELATED WORK

Over the past decade, many researchers focused on perception issues in AR systems, most of them being dedicated to depth perception. Some researchers conducted various experiments to detect human accuracy of depth perception under different conditions in AR.⁴⁻⁶ Studies also focused on finding proper methods to conduct depth assessments.⁷⁻⁹ There are also quantitative depth perception studies for stereoscopic disparity.^{10,11} However, in those studies, the detected human perception depended on the results of task performances. In contrast, the present research studies human operator fatigue by analyzing users' physiological signals and subjective scores. It should be noted that visual fatigue is a feeling of weariness resulting from a visual task. It can be of ocular, muscular, or psychic causes.¹² Moreover, there is no uniform definition for the operator fatigue. In the present research, operator fatigue can be explained as a complex state that affects a person's ability to be attentive and to remain alert.¹³

Traditionally, visual fatigue caused by stereoscopic contents presented by visual display terminals can be measured using (1) subjective questionnaires,¹⁴ (2) critical flicker frequency (CFF), which is defined as the threshold at which light from an intermittent source is seen half the time as flickering and half the time as fused or continuous,^{15,16} or (3) accommodative response or visual function index of human eyes.¹⁷

In addition, some physiological signals have been shown to be good indicators of fatigue. Heart rate variability (HRV) is one of these typical indicators. HRV is defined as the physiological phenomenon of variation in the time interval between heartbeats¹⁸ and is usually recorded by electrocardiography (ECG). ECG records the electrical activity of a participant's heart using electrodes placed on the skin over a period of time.¹⁹ McCraty et al²⁰ proposed that humans' neurological and cognitive functions can influence their heart rate and rhythm. Conversely, the heart's sensory input can also influence cognitive functions, performance, and working memory.²¹ Park et al²² conducted a 2D-3D comparative experiment to investigate whether the visual fatigue from 3D contents can affect humans' autonomic regulations and heart rhythms. The results show that subjects of the 3D group have a significant increase of heart rate compared with those from the 2D group. Heinze et al²³ studied the relationship between HRV and numerous operator fatigue measures and found that several measures can be strong indicators of operator fatigue. Electrooculography (EOG) can also serve as indicators of visual fatigue and operator fatigue. EOG measures the corneo-retinal standing potential that exists between the front and the back of the human eye.²⁴ Stern et al²⁵ proposed that a higher blink rate was related to higher eyestrain. Yu et al²⁶ investigated the visual fatigue caused by 2D and 3D displays through measuring the eye movement based on EOG. Inoue et al²⁷ conducted an experiment to confirm that the video-oculography (VOG) and EOG can be used as indicators of visual fatigue when the users are watching 3D contents. Kaneko et al²⁸ used the EOG waveform as an objective indicator of the visual fatigue assessment of a visual display terminal.

Interestingly, few researchers have studied users' fatigue in AR systems. Tümler et al²⁹ designed an experiment to compare users' strain between an AR-supported and a non-AR-supported work task. They analyzed HRV measures and found that there is no difference between the two tasks. Later, Grubert et al³⁰ designed a new experiment based on Tümler's one. They extended the duration of the tasks, but, once again, there was no significant difference between the two tasks. However, these studies used a few HRV measures, and no powerful evidence was mentioned in their paper to prove that these measures are strong enough to indicate human strain.

3 | MATERIALS AND METHODS

Before the formal experiment, all the participants have performed visual function tests, including the visual acuity test, the stereo acuity test, and the Ishihara³¹ test to ensure that they have normal or correcting normal vision functions. As a result, 18 participants (nine males and nine females) have passed the tests and could participate in the experiment. The average age of the participants is 25.11 years, their ages ranging from 23 to 29, with an SD = 1.45.

Since our comparative experiment consisted of two tasks, all the participants needed to perform both of them. The subjects were randomly divided into two groups: One group performed the virtual magic cube task (viewing AR content) first, and the other group performed the real magic cube task (viewing real objects) first. In order to eliminate learning effects, the users had to wait at least 1 week before performing the second task. Moreover, to guarantee that the participants have similar physiological states in the two tasks, they were asked to perform the two tasks at the same time of two different days.

3.1 | Experiment environment and stimuli

Figure 1 gives an overview of the experimental environment, as well as of the equipment used for the research. The experiment consists of two tasks, which are the virtual magic cube task (Figure 1A) and the real magic cube task (Figure 1B). A stereo OST-HMD with the resolution of 800×600 pixels per eye, designed by the Beijing Institute of Technology, was used to present the virtual 3D images in the experiment.³² The OST-HMD has a binocular visual system with free-form surface optics. The effective focal length of the optics is 15 mm. The field of view is 45H32V. The size of the LCD screens is 0.61 in. The device named BIOPAC MP150 and the software

AcqKnowledge 4.01 were used to record the ECG and EOG signals of the subjects during the experiment. A CFF measurement device was used to gain the subjects' CFF values.

Before the formal experiment, the subjects sat in front of the desk and held the wooden support of the OST-HMD. In the virtual magic cube task, the subjects watched virtual magic cubes that were presented on the platform in the real world through the OST-HMD (cf Figure 1A). The AR scene was produced by the software Unity 3D. The depth and size of the virtual magic cube were set in Unity directly. The virtual cameras' distance was set as an adjustable variable in the program to match different people's interpupillary distance (IPD). An interpupillary ruler was used to measure the participants' IPD before the experiment. Correspondingly, the OST-HMD was turned off in the real magic cube task, and a pair of real magic cubes was located at the same place as the virtual magic cubes. In both tasks, the magic cubes were located at a distance of 0.9 m from the viewer and of the same size. We selected 0.9 m as the watching distance due to that it exceeds the smallest distance of distinct vision of 25 cm, and considering the angular resolution, the viewing distance, and the field of view, 0.9 m was selected to guarantee that the virtual magic cubes look the same size with the real magic cubes. Moreover, the HMD's brightness at the exit pupil position was 150 nits, and the transmittance of the HMD is 60%. In addition, the ambient brightness was kept at 300 lx. The subjects could watch the real magic cubes through the optical lens of the OST-HMD (cf Figure 1B).

In both tasks, seven pairs of magic cubes with fixed patterns served as stimuli. Each pair of magic cubes was presented for 3 minutes. The presentation order of the seven pairs of magic cubes was identical in the two tasks and was the same for all subjects. The OST-HMD was fixed on the holder in order to make the experimental

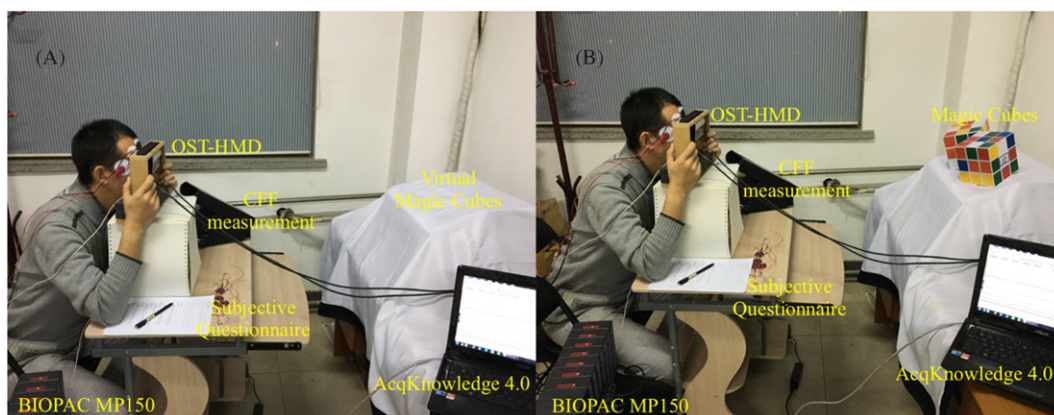


FIGURE 1 The experimental environment: A, the virtual magic cube task; B, the real magic cube task. CFF, critical flicker frequency; OST-HMD, see-through head-mounted display

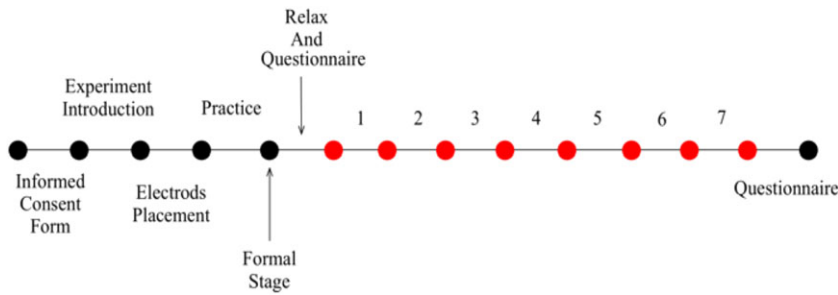


FIGURE 2 The experimental procedure

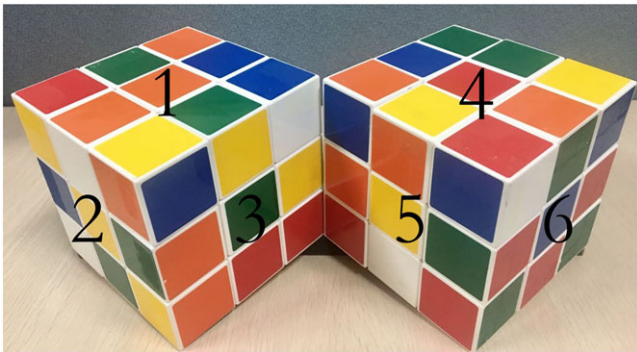


FIGURE 3 Example of one pair of real magic cubes

settings more consistent between trials, as well as to eliminate the influence of the fatigue caused by the weight of the HMD.

3.2 | Experiment procedure

The two tasks of the experiment have the same procedure, which is depicted in Figure 2. Before the formal experiment, written informed consent was obtained and the experiment procedure was explained to each participant. Then the electrodes for ECG and EOG were connected to the subjects. The software began to record the biological data after the connection, and the recording was performed for all the experiment. In addition, an extra experiment was designed for the subjects to practice the tasks before the formal ones.

Then, the subjects began the formal experiment. They were asked to relax for 3 minutes and to answer the subjective questionnaires on visual fatigue. Next, the CFF values of the subjects were tested and recorded by the experimenter. Subsequently, the subjects began to perform the task.

In the real magic cube task, the subjects could see a pair of real magic cubes presented on the platform. In fact, subjects could see six faces of these two magic cubes, as depicted in Figure 3. We have prepared seven pairs of magic cubes with random patterns. Therefore, the patterns on the seven pairs of magic cubes are different from each other and have not been changed during the

experiment. The faces of the cubes were of six different colors: red, orange, yellow, green, blue, and white. The subjects' task was to count the number of each color that appeared on the six faces of the pair of magic cubes. The time allocated for this was 3 minutes. The counting task was in fact irrelevant but served as a guarantee that the subjects always kept their eyes on the magic cubes during the experiment. The subjects were then asked to write down their answers after each session. After that, their CFF values were tested. The experimenter replaced the magic cubes manually. After all these seven sessions, the subjects were asked to answer the subjective questionnaire once again. Moreover, the experimenter marked the time of the beginning and ending of each session in the software.

The procedure of the virtual magic cube task is the same as the real one. The patterns of the virtual magic cubes are the same with the real ones. The only difference is that the virtual magic cubes were presented by the OST-HMD.

4 | RESULTS OF THE EXPERIMENT

4.1 | CFF value

Many previous studies used CFF as an indicator of visual fatigue caused by visual display terminals. The decrease of CFF values is characteristic of a diminution of participants' sensory perception functions.³³

The present research measured participants' CFF values before and after every session for each task; there is thus a total of eight CFF measurements for each subject. Figure 4 represents the mean values of CFF for all the subjects after every session of the tasks. The blue continuous line represents the mean CFF value in the real magic cube task, while the red dotted line is used for the virtual magic cube task. Error bars are graphical representations of the variability of data; it is used on graphs to indicate the error or uncertainty in a reported measurement.* In the present study, the error bars were

*https://en.wikipedia.org/wiki/Error_bar.

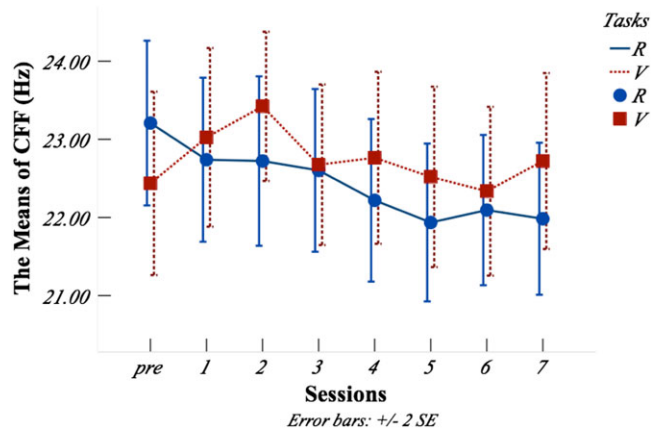


FIGURE 4 Mean value of critical flicker frequency (CFF) during the experimental process

represented by $\pm 2SE$ (SE: standard error). This method shows a 95% confidence interval. In general, the blue line decreases with the task process, whereas the red line reaches its apex in session 2, before showing a downtrend from sessions 2 to 6. Finally, a slight increase can be seen in session 7.

A paired sample t test was used to compare the CFF values between the two tasks. The results show that there is no significant difference between the two tasks. In addition, a paired sample t test was used to compare different sessions within the real magic cube task. The result shows that the participants' CFF values of previewing, sessions 1 and 2 were significantly higher than those of sessions 4 to 7. The CFF values of session 3 are significantly higher than those of sessions 5 to 7. Table 1 gives the significant results of paired sample t test for the real magic cube task.

In summary, we can see that CFF values decreased significantly after the experiment in the real magic cube task. The decreasing trend is not as noticeable in the virtual magic cube task as in the real one. However, we found a phenomenon in Figure 3, where the mean CFF value began to decrease after session 2 in the virtual magic cube task; a paired sample t test was thus applied to compare session 2 with all the other sessions. The results show that the significant differences arise between session 2 and all the following sessions (session 2 vs session 3: $t_{17} = 3.619$, $P = 0.002$; session 2 vs session 4: $t_{17} = 2.462$, $P = 0.025$; session 2 vs session 5: $t_{17} = 3.179$, $P = 0.005$; session 2 vs

session 6: $t_{17} = 3.654$, $P = 0.002$; session 2 vs session 7: $t_{17} = 3.042$, $P = 0.007$). Although the mean value increased during the first two sessions in the virtual magic cube task, the increase is not statistically significant. In other words, the participants' sensory perception functions declined over time in both tasks.

4.2 | Subjective questionnaire

The subjective questionnaire in the present research selected nine symptoms that have a relationship with visual fatigue, including the symptoms of burning, ache, strain, irritation, tearing, blur, double vision, dryness, and headache. The selection of these nine symptoms is based on Sheedy's theory, in which Sheedy et al³⁴ divided visual fatigue into two constellations, labeled as external and internal symptom factors (ESF and ISF). ESF can induce dry eye symptoms, such as burning, irritation, tearing, and dryness, while ISF are related to accommodative and vergence stress, such as ache, strain, and headache. In addition, we added another two symptoms blur and double vision in the present subjective questionnaire. Subjects need to give score for each symptom on a 0 to 100 scale, where the higher the score, the more serious the symptom is. Scoring was achieved by making a cross on a 10-cm-long line on a piece of paper. The experimenters calculated the scores based on the cross's location on the line.

The mean scores of the subjective questionnaires have been calculated by averaging all the subjects' subjective scores before and after the tasks, and are shown in Figure 5, where each bar represents a condition, namely, real-pre, real-post, virtual-pre, and virtual-post. The mean subjective scores before and after the same task were compared. Results show that the mean subjective scores have increased after both tasks.

The paired sample t test was used to compare the subjects' previewing and postviewing scores of each subjective symptom. Results show that significant increases were found on all the symptoms' subjective scores, except for tearing in the real magic cube task, and headache in both of the tasks. Table 2 gives the results of the paired sample t test. In other words, participants suffered from more serious visual fatigue in ESF and ISF after both

TABLE 1 The paired sample t test results for the real magic cube task between different sessions

Results of Paired Sample t Test	Session 4	Session 5	Session 6	Session 7
Previewing	$t_{17} = 3.510$, $P = 0.003$	$t_{17} = 4.374$, $P = 0.001$	$t_{17} = 3.316$, $P = 0.004$	$t_{17} = 4.223$, $P = 0.001$
Session 1	$t_{17} = 2.112$, $P = 0.050$	$t_{17} = 3.646$, $P = 0.002$	$t_{17} = 3.071$, $P = 0.007$	$t_{17} = 3.783$, $P = 0.001$
Session 2	$t_{17} = 2.108$, $P = 0.050$	$t_{17} = 3.753$, $P = 0.002$	$t_{17} = 2.716$, $P = 0.015$	$t_{17} = 2.774$, $P = 0.013$
Session 3	No significant result	$t_{17} = 3.2$, $P = 0.005$	$t_{17} = 2.322$, $P = 0.033$	$t_{17} = 2.498$, $P = 0.023$

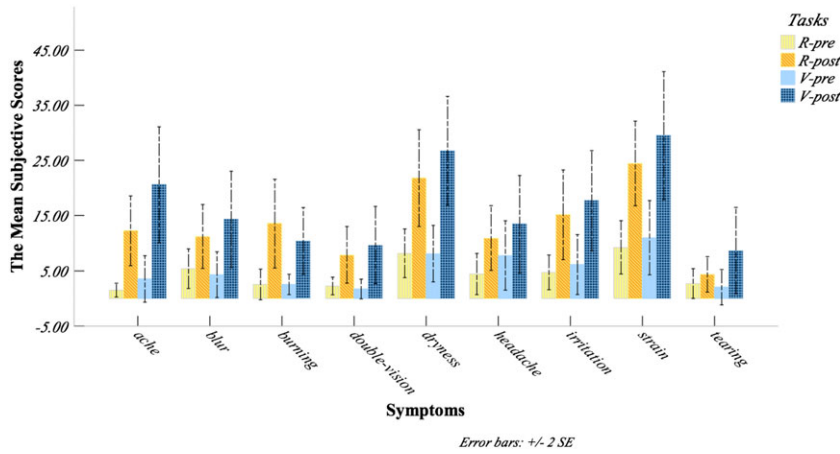


FIGURE 5 Mean scores of the subjective questionnaires

TABLE 2 The paired sample t test results for all the symptoms between previewing and postviewing

	Symptoms								
	Burning	Ache	Strain	Irritation	Tearing	Blur	Double Vision	Dryness	Headache
R- t_{17}	-2.870	-3.463	-3.964	-2.748	-1.390	-2.420	-2.783	-3.645	-1.957
R-P	0.011	0.003	0.001	0.014	0.182	0.027	0.013	0.002	0.067
V- t_{17}	-2.808	-3.903	-3.304	-2.978	-2.134	-2.450	-2.477	-3.662	-1.116
V-P	0.012	0.001	0.004	0.008	0.048	0.025	0.024	0.002	0.280

tasks. After that, a paired sample t test was used to compare the mean values of all the symptoms between the two tasks. The results show that the mean subjective score after the virtual magic cube task (mean = 17.14, SD = 7.24) is significantly higher ($t_8 = 5.69$, $P = 0.000$) than that after the real magic cube task (mean = 13.22, SD = 6.35). Finally, there is no significant difference between the two tasks before the experiment ($t_8 = 1.37$, $P = 0.208$).

4.3 | ECG data

As mentioned above, the subjects' ECG signals were recorded, and we removed artifacts in those signals by applying a correction threshold in AcqKnowledge 4.0. However, we had to discard ECG signals from three of the subjects due to significant noise. This was probably caused by a poor connection between the electrodes' and the subjects' skin (the ECG electrodes were stuck on the legs and arms of participants). As a conclusion, 15 ECG signals were used in the analysis.

4.3.1 | HRV measures used in the study

In order to analyze the participants' time-domain HRV measures, we used the Kubios HRV 2.22 software, which

is also able to compute several nonlinear interesting HRV measures.

Table 3 lists the HRV measures used in our research.³⁵ The abbreviation RR intervals in Table 3 represents the time intervals (in milliseconds) between two successive heart beats. Therefore, STD-RR is the standard deviation of RR intervals, and Mean-RR is the mean of

TABLE 3 Description of the heart rate variability (HRV) measures³⁵

	Measure	Units	Description
Time domain	STD-RR	ms	Standard deviation of RR intervals
	Mean-RR	ms	The mean of RR intervals
	STD-HR	1/min	Standard deviation of instantaneous heart rate values
	NN50		Interval pairs that differ more than 50 ms
	pNN50	%	NN50 divided by the total number of RR intervals
Nonlinear	SD1,SD2	ms	The standard deviation of the Poincaré plot perpendicular to (SD1) and along (SD2) the line of identity
	DET	%	Determinism
	ShanEn		Shannon entropy

RR intervals. In addition, STD-HR is the standard deviation of instantaneous heart rate values. NN50 represents the number of the successive RR intervals that differ more than 50 milliseconds, and pNN50 is the corresponding relative amount:

$$\text{pNN50} = \frac{\text{NN50}}{N - 1} \times 100\%.$$

N is the number of the successive RR intervals.

As for the nonlinear HRV measures SD1 and SD2 in Table 2, a diagram as shown in Figure 5 helps explaining the meanings. The Poincaré plot is a commonly nonlinear method that is a graphical representation of the correlation between successive RR intervals, as shown in Figure 6. An ellipse was used to parameterize the shape of the plot, and it is oriented according to the line of identity ($\text{RR}_j = \text{RR}_{j+1}$). The standard deviation of the points perpendicular to the line of identity denoted by SD1

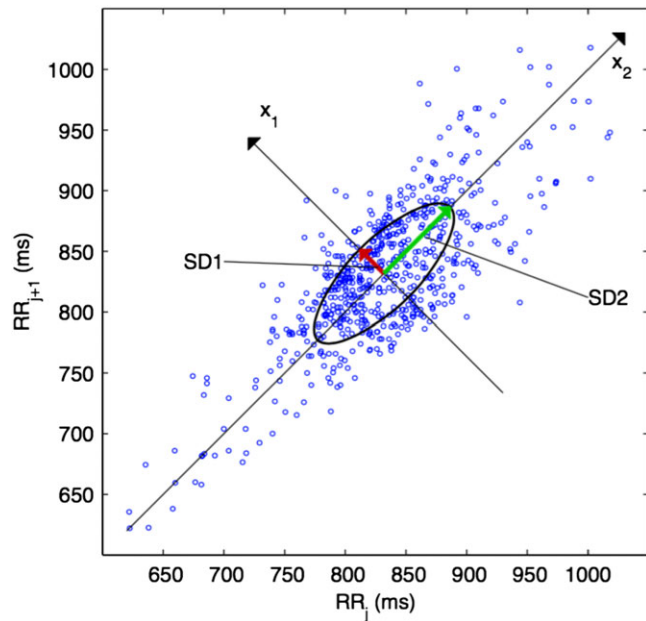


FIGURE 6 Poincaré plot analysis with the ellipse fitting procedure. SD1 and SD2 are the standard deviations in the directions x_1 and x_2 , where x_2 is the line of identity for which $\text{RR}_j = \text{RR}_{j+1}$ ³⁵

describes short-term variability, whereas the standard deviation of the points along the line of identity denoted by SD2 describes long-term variability.

As for DET% and ShanEn, they are both measures using another nonlinear method, named recurrence plot analysis: DET% is the measure of the determinism of RR intervals time series; ShanEn represents Shannon information entropy of the diagonal line length distribution.

Heinze et al²³ calculated the correlations between HRV measures and multiple fatigue measures under three different conditions. Their results show that HRV measures are, in general, good indicators of operator fatigue. SD2 and STD-RR can be considered as strong indicators of operator fatigue, as they have positive correlations with almost all the fatigue measures. They also found that Mean-RR, pNN50, and SD1 have nonsignificant correlations with all fatigue measures in all three conditions of their experiments. Comparatively, some previous studies found that a decreasing Mean-RR has a significant correlation with chronic fatigue syndrome³⁶ and that a decreasing pNN50 value represents physical fatigue.³⁷

As a consequence, we chose to use the reliable fatigue indicators SD2 and STD-RR to detect the operator fatigue induced by the two tasks. Nevertheless, we also analyzed the Mean-RR, pNN50, and SD1 HRV measures to confirm results obtained by Heinze et al.²³ In addition, some previous studies proposed that the subjects' heart rates increase after watching 3D contents.^{20,32} Therefore, heart rate-related measure STD-HR was selected to be analyzed. Furthermore, some nonlinear HRV measures (DET% and ShanEn) have also been selected to detect whether they could be used as potential reliable indicators of operator fatigue.

4.3.2 | Statistical analysis and results

Figure 7 gives the trends of the above three measures. The mean values and standard deviation of Mean-RR, pNN50, and SD1 are presented in Table 4. In order to analyze the

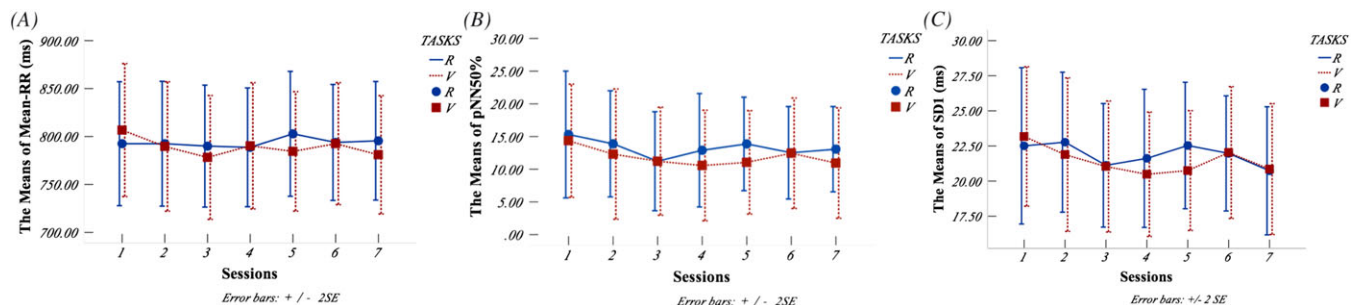


FIGURE 7 A, The mean values of Mean-RR. B, The mean values of pNN50. C, The mean values of SD1

TABLE 4 The mean values and standard deviation of Mean-RR, pNN50, and SD1

		Sessions						
		1	2	3	4	5	6	7
Mean-RR (mean)	R	792.52	792.50	789.99	788.72	802.83	793.81	795.52
	V	806.71	789.63	778.19	790.30	784.49	792.53	780.94
Mean-RR (SD)	R	116.76	117.51	114.86	111.82	117.64	109.18	111.66
	V	125.25	121.86	116.61	118.84	112.47	114.90	111.43
pNN50 (mean)	R	15.31	13.89	11.223	12.91	13.87	12.52	13.07
	V	14.36	12.31	11.224	10.57	11.05	12.45	10.95
pNN50 (SD)	R	16.79	14.06	13.11	15.02	12.38	12.24	11.29
	V	14.99	17.25	14.31	14.64	13.69	14.63	14.63
SD1 (mean)	R	22.51	22.77	21.12	21.62	22.53	21.97	20.73
	V	23.18	22.03	21.04	20.49	20.75	22.03	20.85
SD1 (SD)	R	10.78	9.65	8.52	9.52	8.72	7.94	8.83
	V	9.61	10.58	9.02	8.57	8.26	9.09	9.02

time effect, a paired sample *t* test was conducted on different sessions within the same task. Results show that compared with session 1, all the following sessions (except session 6) show a significant decrease in Mean-RR in the virtual magic cube task (session 1 vs session 2: $t_{14} = 2.497$, $P = 0.026$; session 1 vs session 3: $t_{14} = 4.043$, $P = 0.001$; session 1 vs session 4: $t_{14} = 2.304$, $P = 0.037$; session 1 vs session 5: $t_{14} = 2.684$, $P = 0.018$; session 1 vs session 7: $t_{14} = 2.473$, $P = 0.027$). No significant difference was found during the real magic cube task.

Furthermore, significant decreases in pNN50 values were found in sessions 3 and 4 compared with session 1 in the virtual magic cube task (session 1 vs session 3: $t_{14} = 2.266$, $P = 0.04$; session 1 vs session 4: $t_{14} = 2.906$, $P = 0.011$). As for pNN50 in the real magic cube task, although the mean value decreased in the first three sessions, no significant difference was found.

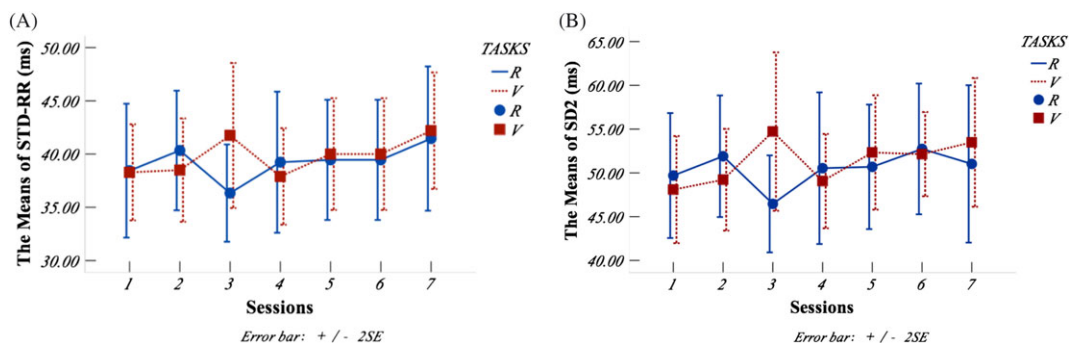
The results of the paired sample *t* test also show that compared with session 1, all the following sessions (except sessions 2 and 6) show a significant decrease in SD1 in the virtual magic cube task (session 1 vs session 3: $t_{14} = 2.815$, $P = 0.014$; session 1 vs session 4: $t_{14} = 3.449$, $P = 0.004$; session 1 vs session 5:

$t_{14} = 2.419$, $P = 0.03$; session 1 vs session 7: $t_{14} = 2.464$, $P = 0.027$). No significant difference was found during the real magic cube task.

The paired sample *t* test also showed no significant difference between the same sessions of the two tasks for these three HRV measures (pNN50, Mean-RR, and SD1).

Moreover, a paired sample *t* test analysis was performed to compare all the HRV measures of corresponding sessions in the two tasks. As previously, a 95% confidence interval was used during statistical analysis. Results show that the reliable fatigue indicators STD-RR and SD2 both have significant differences between the two tasks in session 3 (STD-RR: $t_{14} = -2.306$, $P = 0.037$; SD2: $t_{14} = -2.387$, $P = 0.032$), but no significant difference was noticed in the other sessions.

The trends of the STD-RR and SD2 indicators during the two tasks are depicted in Figure 8. It is interesting to note that they exhibit similar trends and that their mean values are both higher in the virtual magic cube task than in the real magic cube tasks in session 3. No abnormal data were found in session 3. After they fluctuate during the first three sessions, the mean values of

**FIGURE 8** A, The mean values of STD-RR. B, The mean values of SD2

STD-RR and SD2 increased stably in the following sessions. The mean values and standard deviation of these two measures are given in Table 5. In order to detect changes as the time goes by, a paired sample *t* test was applied on the sessions within each task. Results found a significant increase in session 7 compared with session 4 in the virtual magic cube task for the STD-RR indicator ($t_{14} = -2.351, P = 0.034$). Similarly, a significant increase also appeared in session 7 compared with session 4 in the real magic cube task for SD2 ($t_{14} = 3.436, P = 0.004$).

It is worth noting that the following three HRV measures also have a significant difference between the two tasks in session 3: STD-HR ($t_{14} = -2.417, P = 0.030$), DET% ($t_{14} = -2.273, P = 0.039$), and ShanEn ($t_{14} = -2.213, P = 0.044$). They also have similar trends with STD-RR and SD2 (see Figure 9) that were proved to be reliable fatigue indicators in the literature.²³ Eleven out of 15 subjects were found to have a higher STD-HR value in the virtual magic cube task than in the real one in session 3. Regarding DET% and ShanEn, 10 out of 15 subjects have a higher value in session 3 of the virtual magic cube task than in the same session of the real magic cube task. Once again, no abnormal data were found in these subjects. The mean values and standard deviation of STD-HR, DET%, and ShanEn are given in Table 6.

4.4 | EOG data

EOG data were obtained through electrodes placed around the subjects' eyes. A software named Eogui3³⁸ was used to record and extract the EOG data. Eogui3 is a MATLAB software that has a graphical user interface (GUI) to analyze EOG data. It can detect saccades and blinks in the EOG. A correction threshold in the software was used to remove artifacts in the EOG signals. However, for some of our subjects, the extraction of EOG measures remained impossible after the artifacts' removal. In order to guarantee the accuracy of results, EOG signals that have too much noise were removed from the final data. Considering the need to compare results obtained from both tasks, we only kept data from subjects whose EOG signals were good enough in both tasks. As a consequence, six subjects out of 18 were used in the statistical analysis.

Three measures (“Blink duration,” “Blink delay,” and “Blink rate”) were selected as visual fatigue indicators in the present research. Several previous studies investigated the relationship between blink duration and human fatigue. Caffier et al³⁹ found that the blink duration was significantly longer in a drowsy condition than in an alert condition (50 ms longer). This difference is noticeable since it is about one-quarter of the original blink duration

TABLE 5 The mean values and standard deviation of STD-RR and SD2

		Sessions						
		1	2	3	4	5	6	7
STD-RR (Mean)	R	38.44	40.34	36.33	39.24	39.46	40.68	41.46
	V	38.28	38.49	41.74	37.89	39.99	40.40	42.19
STD-RR (SD)	R	11.35	10.14	8.25	11.96	10.19	10.05	12.23
	V	8.15	8.77	12.29	8.18	9.48	7.25	9.88
SD2 (Mean)	R	48.30	50.39	45.23	49.05	49.39	50.19	52.74
	V	46.71	47.46	52.99	48.26	50.76	51.77	52.14
SD2 (SD)	R	13.64	13.53	10.84	16.23	13.47	16.03	13.48
	V	12.00	12.32	17.23	9.93	13.02	14.55	8.69

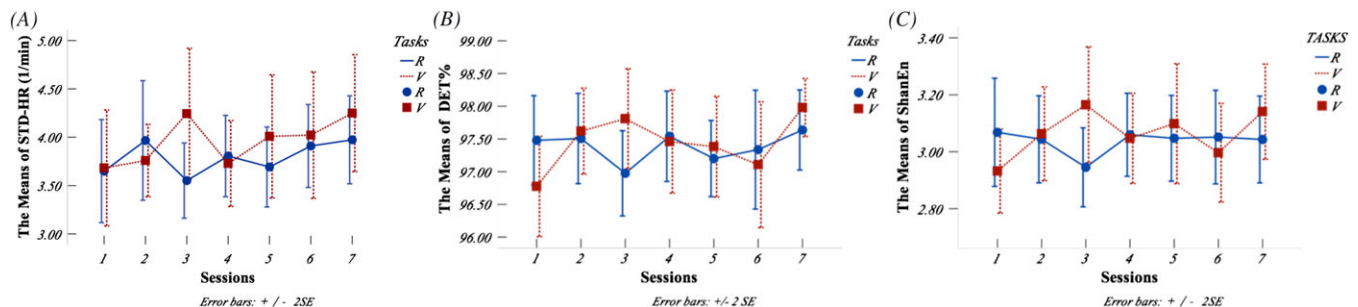


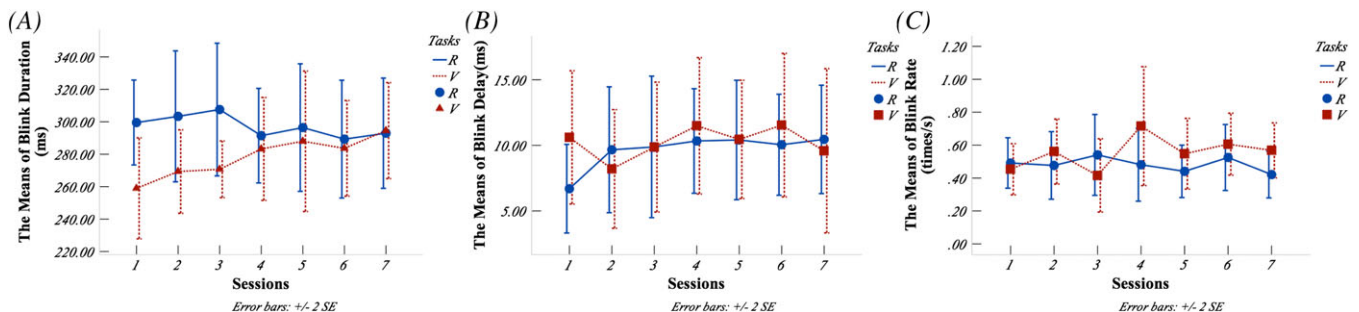
FIGURE 9 A, The mean values of STD-HR (1/min). B, The mean values of DET%. C, The mean values of ShanEn

TABLE 6 The mean values and standard deviation of STD-HR, DET%, and ShanEn

		Sessions						
		1	2	3	4	5	6	7
STD-HR (mean)	R	3.65	3.97	3.55	3.81	3.69	3.91	3.97
	V	3.68	3.76	4.24	3.73	4.00	4.02	4.24
STD-HR (SD)	R	0.96	1.12	0.70	0.76	0.75	0.77	0.82
	V	1.08	0.68	1.22	0.80	1.15	1.18	1.09
DET% (mean)	R	97.48	97.51	96.98	97.54	97.20	97.34	97.64
	V	96.77	97.62	97.81	97.46	97.38	97.11	97.98
DET% (SD)	R	1.32	1.34	1.26	1.34	1.13	1.76	1.19
	V	1.49	1.27	1.48	1.53	1.49	1.86	0.86
ShanEn (mean)	R	3.07	3.04	2.95	3.06	3.05	3.05	3.04
	V	2.93	3.06	3.16	3.05	3.09	2.99	3.14
ShanEn (SD)	R	0.34	0.28	0.25	0.26	0.27	0.29	0.26
	V	0.27	0.29	0.37	0.29	0.38	0.31	0.30

(200 ms). Hsieh et al⁴⁰ proposed a system that utilizes eye blink durations as one of the main schemes to warn drivers about fatigue. Schleicher et al⁴¹ stated that blink duration is by far the most important variable to indicate subjective, as well as video-rated fatigue, and the second most important is the delay of reopening the eye. Other researchers pointed out that the delay of reopening is an indicator of sleepiness, and the longer the delay of reopening, the sleepier the subjects.^{39,42} In addition, many previous studies investigated the influence of visual fatigue on blink rate. Kim et al⁴³ analyzed the individual blink rate and found that eye blinks increased with moderate visual fatigue compared with low visual fatigue for five out of seven subjects. Li et al⁴⁴ focused on the relationship between visual fatigue, blink rate, and 3D motion characteristics. They found that blink rates increased along with subjective scores of visual fatigue when the subjects watched static 3D contents.

Figure 10 shows blink duration, blink delay, and blink rate as measured in our experiment. Figure 10A shows that the mean blink duration increases with the progress of the virtual magic cube task, whereas it generally remains stable during the real magic cube task. Although the mean blink duration in the real magic cube task is higher than that in the virtual magic cube task, they reach almost the same level in session 7 given the increase of blink duration in the virtual magic cube task. Figure 10B shows that mean blink delay fluctuates while remaining at similar values during the experiment in the two tasks. Moreover, the mean blink delay value in the real magic cube task has an increasing trend with the progress of the task. Figure 10C gives the trends of blink rate during the two tasks. At the beginning, the mean blink rates have similar values in the two tasks. After session 3, the mean blink rate in the virtual magic cube task was always higher than that in the real one.

**FIGURE 10** From left to right. A, The mean values of blink duration (ms). B, The mean values of blink delay (ms). C, The mean values of blink rate (times/s)

A paired sample t test was used to compare blink duration, blink delay, and blink rate of all sessions between the two tasks. Results show that no significant difference was found in “Blink duration” and “Blink delay” between the two tasks for all the sessions.

As for the parameter “Blink rate,” a significant difference was found between the two tasks in session 2 ($t_5 = -3.683$, $P = 0.014$) and session 7 ($t_5 = -4.167$, $P = 0.009$). Through observing each subject's blink rate data, it can be found that all the subjects have a higher blink rate in sessions 2 and 7 of the virtual magic cube task than the corresponding sessions in the real magic cube task. It is not very clear to us why these significant results only appear in sessions 2 and 7. Results reveal that participants may be more affected by visual fatigue during the virtual magic cube task than during the real

magic cube task. Nevertheless, additional experiments need to be conducted to confirm this tendency.

Detailed data of mean values and standard deviation for the EOG measures, as well as their t and P values of the paired sample t test, are shown in Tables 7, 8, and 9.

With Table 6, it can be found that the mean value of the subjects' blink duration increased of 35.6 milliseconds after the virtual magic cube task, which is about 13.7% of the original blink duration (258.96 ms). On the contrary, the mean value of the blink duration did not change as much during the real magic cube task. In order to detect the time effect on the EOG measures, a paired sample t test was also used in the comparison between the first session with each other session of the two tasks. Results show a significant difference for blink duration between the first and the last sessions in the virtual magic cube

TABLE 7 Blink duration: mean values, standard deviation, and t values and P values of the paired sample t test between two tasks

		Blink Duration						
		1	2	3	4	5	6	7
Mean	R	299.54	303.35	307.56	291.45	296.43	289.28	292.96
	V	258.96	269.31	270.68	283.25	287.98	283.64	294.56
SD	R	32.06	49.44	50.11	35.70	48.18	44.45	41.61
	V	38.13	31.49	21.43	38.90	53.04	36.09	36.25
t_5		2.381	1.470	1.624	0.578	0.307	0.512	-0.076
P		0.063	0.201	0.165	0.588	0.771	0.63	0.943

TABLE 8 Blink delay: mean values, standard deviation, and t values and P values of the paired sample t test between two tasks

		Blink Delay						
		1	2	3	4	5	6	7
Mean	R	6.69	9.66	9.88	10.33	10.41	10.05	10.45
	V	10.61	8.20	9.86	11.49	10.45	11.54	9.58
SD	R	4.14	5.87	6.60	4.88	5.57	4.71	5.06
	V	6.22	5.54	6.60	6.36	5.52	6.69	7.67
t_5		-2.019	0.554	0.008	-0.593	-0.04	-0.582	0.263
P		0.099	0.603	0.994	0.579	0.97	0.586	0.803

TABLE 9 Blink rate: mean values, standard deviation, and t values and P values of the paired sample t test between two tasks

		Blink Rate						
		1	2	3	4	5	6	7
Mean	R	0.49	0.48	0.54	0.48	0.44	0.52	0.42
	V	0.45	0.56	0.42	0.72	0.55	0.60	0.56
SD	R	0.18	0.25	0.3	0.27	0.19	0.24	0.17
	V	0.19	0.24	0.27	0.44	0.26	0.23	0.20
t_5		1.013	-3.68	0.987	-2.022	-1.917	-2.061	-4.167
P		0.358	0.014	0.369	0.099	0.113	0.094	0.009

task ($t_5 = -4.978, P = 0.04$). As for blink delay, a significant difference arises between the first and the fourth sessions ($t_5 = -2.99, P = 0.03$) and between the first and the last sessions ($t_5 = -2.876, P = 0.035$) in the real magic cube task. The blink rates' results show that significant differences emerge in the virtual magic cube task: session 1 vs session 2 ($t_5 = -2.603, P = 0.048$), session 1 vs session 6 ($t_5 = -3.241, P = 0.023$), and session 1 vs session 7 ($t_5 = -3.421, P = 0.019$).

Considering that physiological signals of individuals may greatly differ, all the subjects' blink data were observed. Results show that there is no outlier in the data.

As for blink duration, the increasing magnitude was computed after the two tasks individually. Results show that while all subjects exhibit a longer blink duration after the virtual magic cube task, only half of them increased their blink duration after the real magic cube task. The increasing magnitudes are listed in Table 10. Note that the average increase in blink duration magnitude is 35.6 milliseconds with a base average value of 258.96 milliseconds.

From the blink delay's raw data, we can see that four out of six subjects have a longer blink delay in the virtual magic cube task than in the real magic cube task during the first two sessions. In the following sessions, this is still true for three out of six subjects. All the subjects had a longer blink delay in the last session of the real magic cube task, whereas only two out of six subjects had a longer blink delay in the last session of the virtual magic cube task.

Table 11 presents the number of subjects having a higher blink rate in the virtual magic cube task than the real magic cube task for each session. Note that in sessions 2 and 7, all the subjects have a higher blink rate in the virtual magic cube task. We also found that the difference between the two tasks is bigger in session 7 than in session 2.

TABLE 10 Subjects' increasing magnitude after the two tasks

	Subject					
	1	2	3	4	5	6
R	-1.83	-10.07	11.79	-81.78	2.42	40.02
V	39.89	46.01	39.55	52.85	2.65	32.66

TABLE 11 The number of the subjects who have higher blink rate in the virtual magic cube task compared with the real cube task

	Session						
	1	2	3	4	5	6	7
Number	2	6	3	4	5	5	6

5 | ANALYSIS AND DISCUSSION

5.1 | CFF and subjective scores

Results of CFF show that (1) the participants' sensory perception functions declined over time in both tasks and that (2) no significant difference was found between the tasks. Moreover, results of the subjective scores show that participants suffered from visual fatigue after the two tasks in both ESF and ISF and that visual fatigue was more important after the virtual magic cube task than after the real magic cube task.

5.2 | ECG data

5.2.1 | Mean-RR, pNN50, and SD1

We know that a decrease of Mean-RR and pNN50 is significantly correlated with human fatigue (see, eg, Tarvainen et al³⁵ and Yoshiharu et al³⁶). Previous research also shows that Mean-RR has a consistent decrease under mental workload,⁴⁵ whereas pNN50 increased under heavier mental workload condition.^{46,47} Some previous studies also show that there is a decrease of SD1 when human do physical exercise.^{48,49} In our experiment, it can be seen that subjects suffer from operator fatigue and mental workload during the virtual magic cube task given the significant decrease of the Mean-RR indicator in nearly all the sessions following session 1 (with the exception of session 6). As for results of the pNN50 indicator, they highlight that subjects got operator fatigue in sessions 3 and 4 of the virtual magic cube task. This phenomenon may be an accumulation of mental workload strong enough to make the pNN50 indicator increase sufficiently to counterbalance the effects of operator fatigue after session 4. The result also shows the significant decrease of the SD1 indicator in nearly all the sessions following session 1 (with the exception of sessions 2 and 6) in the virtual magic cube task. However, no significant result was found in the real magic cube task. Moreover, no significant difference was found between the two tasks.

It should be noted that previous research found different relationships between Mean-RR, pNN50, operator fatigue, and mental workload. Heinze et al²³ found that Mean-RR, pNN50, and SD1 have no significant relationship with operator fatigue, while Cinaz et al found that pNN50 decreases significantly under mental workload.⁵⁰ In our study, we found results that are consistent with experiments that link a decrease in Mean-RR with both operator fatigue and mental workload^{35,45} and a decrease in SD1 with operator fatigue^{48,49} and that correlate a

pNN50 decrease with operator fatigue and a pNN50 increase with mental workload.^{36,46,47}

5.2.2 | STD-RR and SD2

Based on results of a previous study,²³ these two indicators show positive correlations with fatigue. As a consequence, it can be inferred that most of the subjects suffered from more operator fatigue in session 3 of the virtual magic cube task than in the same session of the real magic cube task. No significant differences were noted for the other sessions.

Some questions naturally arise: Why did significant differences only appear in session 3? Based on previous studies,²³ HRV measures can only reliably indicate operator fatigue under strict experimental conditions in order to eliminate the influence of mental workload. They also mentioned that some HRV measures such as HR and HRV showed contradicting sensitivities to workload compared with operator fatigue. In addition, we also found that STD-RR and SD2 have similar contradicting sensitivities to workload and fatigue: STD-RR and SD2 increased with the increasing of operator fatigue,²³ whereas STD-RR decreased with the increasing of mental workload^{50,51} and SD2 was significant lower in the high mental workload condition than in the low mental workload condition.⁵²

It is quite obvious that the tasks used in our study (counting the number of cubes with different colors) can induce a mental workload after a number of repetitions. From our results, it can be shown that operator fatigue mainly influences HRV measures in session 3. In fact, both operator fatigue and mental workload only appear after a certain amount of time. Therefore, we can hypothesize that the subjects did not experience neither operator fatigue nor mental workload during the first two sessions.

However, operator fatigue had been accumulated to an extent level in session 3 in the virtual magic cube task; the subjects' operator fatigue mainly influenced values of STD-RR and SD2 and makes the mean value of STD-RR and SD2 increase after session 2. However, we noticed an increase of the mean values of STD-RR and SD2 indicators starting in session 3 of the virtual magic cube task. This could mean that subjects suffered from operator fatigue in session 3 of the virtual magic cube task.

In order to explain why we did not notice such an increase in the subsequent sessions, we provide the following explanation. We know that STD-RR and SD2 showed contradicting sensitivities to workload compared with operator fatigue: They increased with operator fatigue and decreased with mental workload.^{22,47,51,52} However, mental workload only appears after a certain amount of time. As a

consequence, it is possible that STD-RR and SD2 started to be impacted by the mental workload after session 3 (which represents roughly 10 min).

Since operator fatigue and mental workload have contradicting effects on STD-RR and SD2 (operator fatigue corresponds to an increase, while mental workload is characterized by a decrease of the same indicators), it is highly possible that the mental workload counterbalances effects of operator fatigue in sessions 4 to 7 of the virtual magic cube task. It is possible that for the real magic cube task, operator fatigue only showed up in session 4 and was therefore directly counterbalanced by mental workload. This hypothesis needs however to be confirmed by further experiments.

5.2.3 | STD-HR, DET%, and ShanEn

These three measures show similar trends with STD-RR and SD2. It can be thus concluded that they may have the potential to be used as reliable indicators of operator fatigue. However, further experiments need to be carried out to confirm these results.

5.3 | EOG data

5.3.1 | Blink duration

From the results of the paired sample *t* test, blink duration shows a significant increase after the virtual magic cube task with an average increase in magnitude of 35.6 milliseconds with a base average value of 258.96 milliseconds. As mentioned above (see Wei et al⁵¹), an increase in blink duration of 50 milliseconds (given a base value of 200 ms) arises when subjects experience drowsiness. Even if our results have not reached the extent of a drowsy condition, we still believe that the significant increase of blink duration reflects that subjects suffered from visual fatigue, to a certain extent, after the virtual magic cube task. Some previous studies have proved that time influences human's visual fatigue significantly.⁵³ Besides, many studies on the difference of the visual fatigue caused by the 2D and 3D display terminals show that the significant difference was found after a long continuous watching time of about 15 minutes.⁵⁴ However, the task time for each session in the present experiment was only 3 minutes. Although blink duration mean values in the real magic cube task are higher than those in the virtual cube task, the trends of those mean values lead us to believe that they may have become higher in the virtual magic cube task than in the real magic cube task, if the experiment was longer. However, a further study is needed to verify the idea.

5.3.2 | Blink delay

According to our statistical analysis, there is no significant difference between the two tasks. Nevertheless, we noticed a significant increase of blink delay in session 7 compared with the blink delay in session 1 in the real magic cube task (individual observations show that all subjects had an increased blink delay after this task).

It is interesting to notice that blink duration and blink delay only proved that subjects suffered from visual fatigue in one of the two tasks (the virtual magic cube task and the real magic cube one). However, the results of these two EOG measures did not show any significant difference between the two tasks.

5.3.3 | Blink rate

A question arises as to why significant differences regarding blink rate between the two tasks appeared only between session 2 and session 7. In order to answer this question, individual blink rates were observed. Table 10 shows that all subjects have a higher blink rate in the virtual magic cube task in sessions 2 and 7. Moreover, the difference in session 7 is larger than that in session 2. In other words, the difference between the two tasks increased as the experiment progressed. Combining these trends with the results of the statistical analysis, it can be noted that (1) the virtual magic cube task increased the subjects blink rate significantly and (2) the subjects suffered more from visual fatigue after the virtual magic cube task than the real magic cube task.

6 | CONCLUSION

This study proposes a comparative experiment to investigate differences in human perception when either watching virtual objects displayed in an OST-HMD or watching real objects while wearing an OST-HMD. This study also provides hints that HRV and EOG measures could be used to explore visual and operator fatigue induced by AR content. In addition, CFF measures and subjective questionnaires were used to evaluate subjects' visual fatigue in the experiment.

Among all the measures, two of them show a significant difference between the two tasks: the subjective scores and blink rate. Those indicators reveal that (1) participants got operator and visual fatigue after both tasks and (2) visual and operator fatigue were significantly more important in the virtual magic cube task.

The present research also gives some limited findings that can serve as reference data for future research. First, among the HRV measures, STD-RR and SD2 show that

participants suffered more from operator fatigue in the virtual magic cube task than in the real magic cube task only in session 3. This might be explained by mental workload accumulation. Indeed, mental workload can influence operator fatigue after session 3 (about 10 min). Mental workload accumulation can also partly explain results of the Mean-RR and pNN50 indicators:

- Mean-RR shows significant decrease in the virtual magic cube task due to the combined action of operator fatigue and mental workload.
- pNN50 decreased significantly during sessions 3 and 4 of the virtual magic cube task due to operator fatigue. After session 4, we believe that mental workload counterbalances effects of operator fatigue.

As for the real magic cube task, we hypothesize that it may not induce significant operator fatigue before the end of session 4 and that after this session, it is likely that the limited operator fatigue is also counterbalanced by the effects of mental workload. In addition, the SD1 indicator shows significant decrease due to operator fatigue in the virtual magic cube task. However, it is unclear as to why there is no difference in the real magic cube task or between the two tasks for the Mean-RR, pNN50, and SD1 indicators.

Secondly, some of the measures reveal a time effect during the task: The CFF values indicate that participants' sensory perception functions declined with the tasks progress. Moreover, the significant increase of blink duration after the virtual magic cube task indicates that participants got visual fatigue after this task. Blink delay results show that participants got visual fatigue after the real magic cube task. Thirdly, we also found out that several other HRV measures have the potential to be used as indicators of operator fatigue, namely, STD-HR, DET%, and ShanEn. However, further experiments need to be conducted to confirm these findings.

Although not all measures show significant differences between the two tasks, the present study still found some evidences showing that the virtual magic cube task induces more visual and operator fatigue than the real magic cube task. This difference can be explained by the following factors: (1) The virtual magic cubes presented in the OST-HMD are self-luminous, but the real magic cubes are diffuse objects, and (2) the contradiction between accommodation and convergence when using OST-HMDs could also induce more virtual fatigue.

Nevertheless, our current research also has some limitations: Simpler and more boring tasks should be used in future studies to eliminate the mental workload effect. Indeed, as mentioned above, mental workload may influence the detection of operator fatigue. Regarding the

experiment, it should probably be made longer in order to induce larger differences between the two tasks. Finally, the brightness of virtual objects should also be taken into consideration in future research.

ACKNOWLEDGMENTS

This work is supported by the National Key Research and Development Program of China (No.2016YFB0401202) and the National Natural Science Foundation of China (Grant No.61661146002).

ORCID

Jean-Marie Normand  <https://orcid.org/0000-0003-0557-4356>

Guillaume Moreau  <https://orcid.org/0000-0003-2215-1865>

REFERENCES

- Gabbard JL, Swan JE, Hix D, Schulman RS, Lucas J, Gupta D. An empirical user-based study of text drawing styles and outdoor background textures for augmented reality, in *Virtual Reality, 2005. Proceedings. VR 2005. IEEE. IEEE*. 2005: 11–18. <https://doi.org/10.1109/VR.2005.1492748>
- Gao Y, Liu Y, Cheng D, Wang Y. A review on development of head mounted display. *J Comput Aided Des Comput Graph*. 2016;6(1):896–904. <https://doi.org/10.3969/j.issn.1003-9775.2016.06.004>
- Hu X, Hua H. High-resolution optical see-through multi-focal-plane head-mounted display using freeform optics. *Opt Express*. 2014;22(11):13 896–13 903. <https://doi.org/10.1364/OE.22.013896>
- Livingston MA, Ai Z, Swan JE, Smallman HS. Indoor vs. outdoor depth perception for mobile augmented reality, in *2009 IEEE Virtual Reality Conference*. 2009: 55–62. <https://doi.org/10.1109/VR.2009.4810999>
- Swan JEI, Livingston MA, Smallman HS, et al. A perceptual matching technique for depth judgments in optical, see-through augmented reality, in *IEEE Virtual Reality Conference (VR 2006)*. 2006:19–26. <https://doi.org/10.1109/VR.2006.13>, ISSN=1087–8270.
- Swan JE, Jones A, Kolstad E, Livingston MA, Smallman HS. Egocentric depth judgments in optical, see-through augmented reality. *IEEE Trans Vis Comput Graph*. May 2007;13(3): 429–442. <https://doi.org/10.1109/TVCG.2007.1035>
- Singh G, Ii JES, Jones JA. Depth judgments by reaching and matching in near-field augmented reality, in *2012 IEEE Virtual Reality Workshops (VRW)*. 2012:165–166. <https://doi.org/10.1109/VR.2012.6180933>
- Swan JE, Singh G, Ellis S. Matching and reaching depth judgments with real and augmented reality targets. *IEEE Trans Vis Comput Graph*. November 2015;21(11):1289–1298. <https://doi.org/10.1109/TVCG.2015.2459895>
- Koheishima, Moser KR, Rompapas DC, et al. SharpView: improved clarity of defocused content on optical see-through head-mounted displays, in *2016 IEEE Symposium on 3D User Interfaces (3DUI)*. 2016:173–181. <https://doi.org/10.1109/3DUI.2016.7460049>
- Eisemann E. A perceptual model for disparity. *ACM Transactions on Graphics (Proceedings SIGGRAPH 2011, Vancouver)*. 2011;30(4). <https://doi.org/10.1145/1964921.1964991>
- Du SP, Masia B, Hu SM, Gutierrez D. A metric of visual comfort for stereoscopic motion. *ACM Trans Graph*. 2013;32(6): 222:1–222:9. <https://doi.org/10.1145/2508363.2508387>
- Millodot M. *Dictionary of optometry and visual science*. 7th ed. UK: Butterworth Heinemann—Elsevier; 2009:465. <https://doi.org/10.1111/j.1444-0938.2009.00388.x>
- Scott R. Calhoun LT, USCG. Human factors in ship design: preventing and reducing shipboard operator fatigue. University of Michigan. 2006.
- Lambooji M, Ijsselsteijn W, Fortuin M, Heynderickx I. Visual discomfort and visual fatigue of stereoscopic displays: a review. *J Imaging Sci Technol*. 2009;53(3):30 201–1–30 201–14. <https://doi.org/10.2352/J.ImagingSci.Technol.2009.53.3.030201>
- Flicker fusion threshold. https://en.wikipedia.org/wiki/Flicker_fusion_threshold, retrieved: 2018–01–30.
- Murata K, Araki S, Yokoyama K, Yamashita K, Okumatsu T, Sakou S. Accumulation of VDT work-related visual fatigue assessed by visual evoked potential, near point distance and critical flicker fusion. *Ind Health*. 1996;34(2):61–69. <https://doi.org/10.2486/indhealth.34.61>
- Emoto M, Niida T, Okano F. Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television. *J Disp Technol*. December 2005;1(2):328–340. <https://doi.org/10.1109/JDT.2005.858938>
- Heart rate variability. https://en.wikipedia.org/wiki/Heart_rate_variability, retrieved: 2018-01-30.
- Electrocardiography. <https://en.wikipedia.org/wiki/Electrocardiography>, retrieved: 2018-01-30.
- McCraty R, Atkinson M, Tomasino D, Bradley RT. The coherent heart: heart-brain interactions, psychophysiological coherence, and the emergence of system-wise order. *Int For Rev*. 2009;5(2):10–115.
- Hansen AL, Johnsen BH, Thayer JF. Vagal influence on working memory and attention. *Int J Psychophysiol*. 2003;48(3):263–274. [https://doi.org/10.1016/S0167-8760\(03\)00073-4](https://doi.org/10.1016/S0167-8760(03)00073-4)
- Park S, Won MJ, Mun S, Lee EC, Whang M. Does visual fatigue from 3D displays affect autonomic regulation and heart rhythm? *Int J Psychophysiol*. 2014;92(1):42–48. <https://doi.org/10.1016/j.ijpsycho.2014.02.003>
- Heinze C, Trutschel U, Schnupp T, et al. Operator fatigue estimation using heart rate measures. In: Dössel O, Schlegel WC, editors *World Congress on Medical Physics and Biomedical Engineering WC*: Springer; 2009. pp. 930–933 https://doi.org/10.1007/978-3-642-03882-2_248.
- Electrooculography. <https://en.wikipedia.org/wiki/Electrooculography>, retrieved: 2018–01–30.
- Stern JA, Boyer D, Schroeder D. Blink rate: a possible measure of fatigue. *Hum Factors*. 1994;36(2):285–297. <https://doi.org/10.1177/001872089403600209>

26. Yu JH, Lee BH, Kim DH. EOG based eye movement measure of visual fatigue caused by 2D and 3D displays, in *Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics*. 2012:305–308. <https://doi.org/10.1109/BHI.2012.6211573>
27. Inoue T, Kato YO, Ozawa J. Evaluating visual fatigue by sensing eye movement during viewing of 3D images, in *The 1st IEEE Global Conference on Consumer Electronics 2012*. 2012:486–490. <https://doi.org/10.1109/GCCE.2012.6379666>
28. Kaneko K, Sakamoto K. Spontaneous blinks as a criterion of visual fatigue during prolonged work on visual display terminals. *Percept Mot Skills*. 2001;92(1):234–250. <https://doi.org/10.2466/pms.2001.92.1.234>
29. Tümler J, Mecke R, Schenk M, et al. Mobile augmented reality in industrial applications: approaches for solution of user-related issues, in 2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality. 2008:87–90. <https://doi.org/10.1109/ISMAR.2008.4637330>
30. Grubert J, Hamacher D, Mecke R, et al. Extended investigations of user-related issues in mobile industrial AR. *IEEE International Symposium on Mixed and Augmented Reality IEEE*, 2010:229–230. <https://doi.org/10.1109/ismar.2010.5643581>
31. Birch J, Mckeever LM. Survey of the accuracy of new pseudoisochromatic plates. *Ophthalmic Physiol Opt*. 1993;13(1):35–40. <https://doi.org/10.1111/j.1475-1313.1993.tb00423.x>
32. Cheng D, Hua H, Wang Y. Optical see-through free-form head-mounted display. 2016: US Patent 9,239,453.
33. M. Kumashiro, “Practical measurement of psychophysiological functions for determining workloads,” in *Evaluation of human work*, 3rd ed., J. R. Wilson, N. Corlett, Ed. Taylor & Francis: Boca Raton, 2005, pp. 605–627.
34. Sheedy JE, Hayes JN, Jon E. Is all asthenopia the same? *Optom Vis Sci*. November 2003;80(11):732–739. <https://doi.org/10.1097/00006324-200311000-00008>
35. Tarvainen MP. Kubios HRV users guide. http://www.kubios.com/downloads/Kubios_HRV_Users_Guide.pdf, retrieved: 2018-01-30.
36. Yoshiharu Y, Lamanca JJ, Natelson BH. A measure of heart rate variability is sensitive to orthostatic challenge in women with chronic fatigue syndrome. *Exp Biol Med*. 2003;228(2):167–174. <https://doi.org/10.1177/153537020322800206>
37. Pichot V, Bourin E, Roche F, et al. Quantification of cumulated physical fatigue at the workplace. *Pflugers Arch*. 2002;445(2). <https://doi.org/10.1007/s00424-002-0917-7>
38. Schleicher R. Eogui—a software to analyze electro-oculogram (EOG) recordings. <https://www.mathworks.com/matlabcentral/fileexchange/32493>, retrieved: 2018-01-30.
39. Caffier PP, Erdmann U, Ullsperger P. Experimental evaluation of eye-blink parameters as a drowsiness measure. *Eur J Appl Physiol*. May 2003;89(3):319–325. <https://doi.org/10.1007/s00421-003-0807-5>
40. Hsieh CS, Tai CC. An improved and portable eye-blink duration detection system to warn of driver fatigue. *Instrum Sci Technol*. 2013;41(5):429–444. <https://doi.org/10.1080/10739149.2013.796560>
41. Schleicher R, Galley N, Briest S, Galley L. Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? *Ergonomics*. 2008;51(7):982–1010. <https://doi.org/10.1080/00140130701817062>
42. Tucker AJ, Johns MW. The duration of eyelid movements during blinks: changes with drowsiness. *Sleep*. 2005;28:A122.
43. Kim D, Choi S, Park S, Sohn K. Stereoscopic visual fatigue measurement based on fusional response curve and eye-blinks, in *2011 17th International Conference on Digital Signal Processing (DSP)*. 2011:1–6. <https://doi.org/10.1109/ICDSP.2011.6004999>
44. Li J, Barkowsky M, Callet PL. Visual discomfort of stereoscopic 3D videos: influence of 3D motion. *Displays*. 2014;35(1):49–57. <https://doi.org/10.1016/j.displa.2014.01.002>
45. Lo JC, Prorail ES, Meijer SA. Measuring mental workload with low-cost and wearable sensors: insights into the accuracy, obtrusiveness, and research usability of three instruments. *J Cogn Eng Decision Mak*. 2017;11(4):323–336. <https://doi.org/10.1177/1555343417716040>
46. Guo W, Tian X, Tan J, Wang L. Change in heart rate variability indexes due to high driving workload in turning left at the intersection in real road environment, in *Proceedings of the 2016 3rd International Conference on Materials Engineering, Manufacturing Technology and Control (2016)*. 2016:804–807. <https://doi.org/10.2991/icmemtc-16.2016.158>
47. Guo W, Tian X, Tan J, Li Z, Li L. Driver's mental workload estimation based on empirical physiological indicators, in *Chinese Association of Automation (YAC), Youth Academic Annual Conference of IEEE, 2016*, no. 61503007. 2016. <https://doi.org/10.1109/YAC.2016.7804916>
48. Machado HG, Simões RP, Mendes RG, et al. Cardiac autonomic modulation during progressive upper limb exercise by patients with coronary artery disease. *Braz J Med Biol Res*. 2011;44(12):1276–1284. <https://doi.org/10.1590/S0100-879X2011007500134>
49. Schaal K, Le Meur Y, Bieuzen F, et al. Effect of recovery mode on postexercise vagal reactivation in elite synchronized swimmers. *Appl Physiol Nutr Metab*. 2013;38(2):126–133. <https://doi.org/10.1139/apnm-2012-0155>
50. Cinaz B, Marca RL, Tröster G. Monitoring of mental workload levels during an everyday life office-work scenario. *Pers Ubiquit Comput*. 2011;17(2):229–239. <https://doi.org/10.1007/s00779-011-0466-1>
51. Wei Z, Zhuang D, Wanyan X, Zhang H, Chen L. A theoretical model of mental workload in pilots based on multiple experimental measurements, in *International Conference on Engineering Psychology and Cognitive Ergonomics, Springer, Cham*. 2008:104–113. <https://doi.org/10.1007/978-3-319-07515-011>
52. Krüger A, Cilliers PJ. A systems approach to the assessment of mental workload in a safety-critical environment. *Doctoral Dissertation, University of Pretoria*. 2008. <https://ieeexplore.ieee.org/abstract/document/7804916/metrics>.
53. Lin YH, Chen CY, Lu SY, Lin YC. Visual fatigue during VDT work: effects of time-based and environment-based conditions. *Displays*. 2008;29(5):487–492. <https://doi.org/10.1016/j.displa.2008.04.003>
54. Kim YJ, Lee EC. EEG based comparative measurement of visual fatigue caused by 2D and 3D displays. *Commun Comput Inform Sci*. 2011;174:289–292. https://doi.org/10.1007/978-3-642-22095-1_59

AUTHOR BIOGRAPHIES



Yuan Gao is a PhD student of Optical Engineering at the School of Optoelectronics, Beijing Institute of Technology from the year of 2014. From November 2016 to May 2018, she studied as a joint doctoral student in the Computer Science and Mathematics Department, École Centrale de Nantes, France. Her current research interests include the human visual fatigue in virtual reality and augmented reality, human's depth perception in head-mounted displays, and the plausibility of virtual reality and augmented reality.



Yue Liu received his PhD degree in Telecommunication and Information System from Jilin University, Jilin Province, China, in 2000. He is currently a Professor of optical engineering at the School of Optoelectronics, Beijing Institute of Technology, Beijing. His research interests include human computer interaction, virtual and augmented reality, accurate tracking of the pose of camera, 3D display system, and camera calibration. He has published more than 100 technical papers. Dr Liu is a member of council of China Society of Image and Graphics, a member of China Society of Optics, and he also serves on the editorial board of the Optical Technique.



Jean-Marie Normand received his PhD degree in Computer Science, University of Nantes, France, in 2008. He is currently associate professor in the Computer Science and Mathematics Department, École Centrale de Nantes, France. His research interests include embodiment and interaction in immersive Virtual Reality as well as plausibility of virtual, augmented, and real environments.



Guillaume Moreau received his PhD degree in Computer Science, University of Rennes I, France, in 1998. He is currently a Professor in the Computer Science and Mathematics, Ecole Centrale de Nantes, France, where he also serves as the Dean of Studies. His research interests deal with perception in virtual and augmented worlds. He has been studying visual fatigue in AR/VR as well as plausibility and fidelity of virtual and real environments.



Xue Gao received her bachelor's degree in Optical Information Science and Technology from Beijing Jiaotong University, Beijing, China, in 2015. She is currently a student of Optical Engineering for her master's degree at the School of Optoelectronics, Beijing Institute of Technology, Beijing. Her research interests in projection based augmented reality and museum exhibition design based on virtual reality.



Yongtian Wang received the BSc degree in Precision Instrumentation from Tianjin University, China, in 1982, and the PhD degree in Optics from the University of Reading, UK, in 1986. He is currently a Yangtze River Scholar of the Chinese Ministry of Education and a Professor and the Director of the Center for Research on Optoelectronics and Information Technology in Beijing Institute of Technology. Dr Wang is a Fellow of the Optical Society of America and the International Society for Optical Engineers. His research interests include optical design and CAD, optical instrumentation, image processing, and virtual reality (VR) and augmented reality (AR) technologies and applications.

How to cite this article: Gao Y, Liu Y, Normand J-M, Moreau G, Gao X, Wang Y. A study on differences in human perception between a real and an AR scene viewed in an OST-HMD. *J Soc Inf Display*. 2019;1–17. <https://doi.org/10.1002/jsid.752>