Journal of Environmental Psychology 65 (2019) 101338

Contents lists available at ScienceDirect

Journal of Environmental Psychology

journal homepage: www.elsevier.com/locate/jep

Virtual reality: A new method to investigate cognitive load during navigation

A. Armougum[a,c,*], E. Orriols[a], A. Gaston-Bellegarde[a], C. Joie-La Marle[a,c], P. Piolino[a,b]

[a] Université de Paris, MC2Lab, F-92100 Boulogne-Billancourt, France
[b] Institut Universitaire de France (IUF), Paris, France
[c] SNCF Mobilités, Transilien, SID, Lab Mass Transit, Paris, France

ARTICLE INFO

Handling Editor: Sander van der Linden

Keywords:
Cognitive load
Virtual reality
Memory
Electrodermal activity
Train travelers
Expertise

ABSTRACT

Cognitive load has for long been studied in relation with learning processes. In our study, we investigated the impact of cognitive load in real-life situations taking the example of train travelers looking for relevant information in a train station. For this purpose, we created a virtual reality model of the tested train station from which we conducted a real-life study. Our aim was to compare travelers' cognitive load impact in real-life environmental condition versus virtual reality simulation of the same environment. Regular and occasional travelers were recruited and were assumed as experts and novices, respectively. Investigation of cognitive load was based on physiological, subjective and behavioral aspects. These were measured using electrodermal activity, NASA-Task Load Index and recognition of relevant factual and contextual information seen by travelers in the train station, respectively. These three indicators used in both real-life and virtual reality were expected to follow the same trend, irrespective of the environmental condition, but were expected to vary with respect to expertise level. A higher cognitive load was forecasted for novice travelers than for expert travelers. The findings revealed no difference on the three indicators between virtual and real-life conditions. However, novice travelers showed higher cognitive load responses than expert travelers, in both environmental conditions. Our results suggest virtual reality as a promising technique for cognitive load analysis during navigation, and an effective method for neurocognitive assessments in daily life situations.

1. Introduction

"*I move, therefore I am*" said the Japanese writer Haruki Murakami (2009). Human beings are in constant movement for living, working and discovering the world. We rely on information all around us to understand and control our environment. This is the case, whenever we are in a train station seeking for the right information to reach our destination on time. During this process of information acquisition, cognitive load is generated. This phenomenon corresponds to the total amount of resources invested in learning process (Sweller, 1988). From this definition, cognitive load can be hypothetically represented as a reservoir of cognitive resources available to perform multiple tasks. It is thus necessary to optimize its level, while avoiding saturation or overload of our cognitive resources (Kirsh, 2000). This is usually the case, for instance, by the end of a big day of work, when we are sometimes unable to process a cognitively low-demanding task. To be able to use strategically our limited cognitive resources, we first need to understand the underlying theory: Cognitive Load Theory (CLT) (Sweller, 1988; Sweller, Van Merrienboer, & Paas, 1998).

CLT does not have its own model *per se*, but instead is constructed from another validated model, working memory model (Baddeley, Logie, Bressi, Sala, & Spinnler, 1986). CLT is completed with studies conducted on attentional resources to account for limitations in cognitive resources and individual differences (Chanquoy, Tricot & Sweller, 2007). Working memory resources are available in three inter depending pools: *intrinsic cognitive load* (Sweller, 1994) which depends on the structure of the task itself, *extraneous cognitive load* (Sweller & Chandler, 1991) which arises whenever cognitive resources are allocated to working memory for operations that are irrelevant to the task, and *Germane cognitive load* (Sweller et al., 1998) characterized as the learning process itself. It is related to schema creation, automatization and procedural thinking and action (Chanquoy et al., 2007) (). Therefore, to avoid cognitive overload, we should optimize intrinsic cognitive load, reduce extraneous cognitive load and maximize Germane cognitive load. Several authors have worked on cognitive load optimization, especially in the field of educational psychology (for examples in visual

* Corresponding author. Université de Paris, MC2Lab, F-92100 Boulogne-Billancourt, France
E-mail address: allan.armougum@parisdescartes.fr (A. Armougum).

https://doi.org/10.1016/j.jenvp.2019.101338
Received 19 March 2019; Received in revised form 9 July 2019; Accepted 19 August 2019
Available online 22 August 2019

representations in science education: Cook, 2006; in statistics: Paas, 1992; in problem solving: Sweller, 1988; in health education: Van Merriënboer & Sweller, 2010). These studies have put in light other non-cognitive factors affecting our cognitive resources that need to be considered when conducting studies on cognitive load. A few examples of these factors are: emotional state, fatigue level, vigilance level, expertise level, to name a few (see Chanquoy et al., 2007 for review). These non-cognitive factors together with cognitive factors, such as memory and attention, form the mental load (Jex, 1998). These phenomena suggest the use of different types of measure for quantifying cognitive load. The combination of three main types of cognitive load measurement techniques are mostly used in cognitive load investigation (Cain, 2007; Cegarra & Chevalier, 2008; Kramer, 1991). The first one corresponds to *physiological measures* (Paas, Renkl, & Sweller, 2003; Sweller, Ayres, & Kalyuga, 2011) such as electroencephalography (Gevins et al., 1998; Ke et al., 2014), Heart rate: Paas & Van Merriënboer, 1994), and more recently electrodermal activity (EDA) (Shi, Ruiz, Taib, Choi, & Chen, 2007, pp. 2651–2656). It measures skin conductance response (former *galvanic skin response*), which is known to rise during cognitive effort, providing an indirect measure of cognitive load (Chen et al., 2016, pp. 87–99). EDA has also been investigated in relation to other physiological factors that may interfere with cognitive load, such as stress level (Setz et al., 2009). EDA reflects generalized changes in the state of arousal. These can be caused by emotional, cognitive or physical stimulation (Picard, 2015). EDA levels have also been shown to vary with different cognitive load levels (e.g. in real-world driving tasks: Healey & Picard, 2005). A second type of cognitive load measure is related to *subjective measures* (Vidulich & Tsang, 2012). Two main subjective scales used in the literature are the NASA-Task Load Index (Hart & Staveland, 1988) and the Paas scale (Paas, 1992). A third type of measure recommended in experiments conducted in cognitive load theory is *behavioral measures* through performance analysis (Brunken, Plass, & Leutner, 2003; Paas & Van Merriënboer, 1994). The measuring techniques are usually chosen according to the experimental context. So far, most of these studies have investigated cognitive load in a static context, such as performing a cognitive task on a computer. This static configuration, however, does not reflect our everyday life activities, where we learn about our environment by moving around and exploring it. This raises the question on how our cognitive load varies during navigation in a non-static ecological context.

We are obviously limited in available techniques to investigate cognitive load in a non-static condition. This is due to the fact that we cannot create the same scenario with same external real-life factors, on and on again for a statistically reliable sample size. One promising method, which has proven its efficacy for reproducing real-life environments, is virtual reality (VR) technology (Koutsoudis, Arnaoutoglou, & Chamzas, 2007; Van Veen, Distler & Braun, 1998; Whyte, Bouchlaghem & Thorpe, 2000).

VR technology has bloomed since the first appearance of 3D representations in 1838 with the use of stereoscope. Today numerous virtual headsets capable of displaying high 3D quality images have been developed (e.g. HTC vive, Oculus rift, Tobii) and are used in neuroscience and psychology research, to investigate healthy adults (e.g. virtual visit of museum: Carrozzino & Bergamasco, 2010; evolution of episodic memory in normal ageing: Plancher, Gyselinck, Nicolas, & Piolino, 2010; influence of action on episodic memory: Plancher, Barra, Orriols, & Piolino, 2013; enactment and episodic memory among adults: Jebara, Orriols, Zaoui, Berthoz, & Piolino, 2014; working memory and episodic memory: Plancher, Gyselinck, & Piolino, 2018) as well as patients (e.g. mild cognitive impaired patients and Alzheimer patients: Plancher, Tirard, Gyselinck, Nicolas, & Piolino, 2012; psychotherapy: Riva, 2005). In these fields of research, VR is commonly used to help researchers design new fictional and ecological scenarios that allow them to control random variables that are usually difficult or impossible to manipulate in real-life. VR also helps to create complex real-life environments or structures that represent an effective method for neurocognitive assessments of cognitive load in daily life situations. This technology is also compatible with other interfaces, for instance with a crowd simulator, to observe how an individual's reaction depends on different crowd densities, if immersed in a virtual environment (VE) (Musse & Thalmann, 2001; Ulicny & Thalmann, 2001).

One important aspect of using VR technology is to generate self-experience in individuals, bodily immersion and sensation, which were believed to arise solely in real-life condition (Murray & Gordon, 2001). Studies that succeeded to recreate a VE close to real-life condition, managed to generate same self-experience in individuals. For instance, in a study conducted in VR in the subway of London which aimed to show that healthy participants experienced unfounded paranoid thoughts, had put in light that the same degree of paranoid reactions were observed in VR environment as in real-life situations. This proved VR to be a potent tool for the understanding of paranoid reactions (Freeman et al., 2008). In another study, social phobia therapy had been investigated in VR. The efficacy of this technology was shown to be significantly close to classical cognitive behavior therapy (Klinger et al., 2005). More recently, in a study on mindfulness practice and meditation, psychological and physiological responses to stressful situations had been studied in an immersive VR environment. An increase in mindfulness and anxiety levels had been shown through the use of realistic life situations designed in a virtual reality context (Crescentini, Chittaro, Capurso, Sioni, & Fabbro, 2016). This immersive level is closely related to the optimization of the VE and is measured through subjective sense of presence of the participant in the VE (Schubert, Friedmann, & Regenbrecht, 2001). If the VE is sufficiently vivid, extensive and inclusive (Slater & Wilbur, 1997), and as close as possible to the real world (Barfield & Hendrix, 1995; Bystrom, Barfield, & Hendrix, 1999), the higher will be the sense of presence (Schubert et al., 2001).

Studies conducted in navigation and more precisely wayfinding in the field of spatial cognition, have greatly benefited from the use of VR (Chen & Stanney, 1999; Darken & Peterson, 2014; Darken & Sibert, 1996; Lee & Wong, 2014; Steinicke, Visell, Campos, & Lécuyer, 2013, pp. 199–219; Viaud-Delmon, Ivanenko, Berthoz, & Jouvent, 1998). Besides, VR has been used to help reduce cognitive load associated to complex surgical procedures in otorhinolaryngology, by providing virtual training sessions to surgeons (Andersen, Konge, & Sørensen, 2018). To this date, not much has been done on the relationship between navigation and cognitive load analysis in virtual reality. This implies that before investigating complex situations in VR, we first need to validate its efficacy for cognitive load analysis during navigation.

In the present study, we aimed to set up two experiments. The first one was conducted in real-life condition, in an actual train station, and the second one was carried out in the VR version of the same tested train station. To reduce the effects of any covarying factors, we tested for known effects on cognitive load in our experiment. We evaluated the difference in expertise level of travelers: regular travelers as experts and occasional travelers as novices. This assumption was based on the fact that expertise improves mnemonic performance (for review: expertise of taxi drivers for street names, Kalakoski & Saariluoma, 2001; Memory of drawing experts, Perdreau & Cavanagh, 2015; Chess masters' performance compared to novices', Reingold, Charness, Pomplun, & Stampe, 2001). We predicted occasional travelers to show higher physiological activity than regular travelers, during information processing. Higher self-perceived mental load was also expected for occasional travelers. We also hypothesized regular travelers to demonstrate a higher mnemonic performance for relevant information on signage boards and information display screens than occasional travelers, for factual information and their corresponding contextual specifications. These known cognitive differences between experts and novices should be statistically the same for both real-life and virtual reality experiments.

2. Method

The real-life experiment was performed in a Parisian train station, Saint-Michel Notre Dame, line C. The virtual reality experiment was conducted in our University laboratory, in a fully equipped room.

2.1. Participants

112 participants of age 26 (± 5) years old, constituting a stratified sample of 50% male, were selected for the study, from 121 participants recruited in total. 56 participants were randomly assigned to the real-life condition, and the other 56 participants randomly to the VR condition. Half of the participants in each of these two conditions were female. The recruitment for this study was made among undergraduate students from Paris Descartes University, from the Faculty of Psychology. These participants were rewarded additional marks for one course module. Other undergraduate participants were recruited online from a site dedicated to cognitive science studies and they were rewarded too. A second group of participants was recruited directly from the investigated train station.

Participants' use of video games was monitored, since virtual reality has a common interface to that of video games, which could be subjected to expertise for regular gamers. Our sample consisted of participants who played video games at most once a month.

2.1.1. Expertise level

Participants' selection was also based on their expertise level of train station and particularly for the investigated one, using a survey. The latter informed us about the frequency of visits of Saint-Michel Notre Dame train station. This category contained three questions: question one investigated the frequency of utilization of Saint-Michel Notre Dame for line C, ranging from *every day* to *one to two times a year*, on a five segmented Likert scale. The second question dealt with the frequency of utilization of Saint-Michel Notre Dame globally (for all trains' lines except for line C), using the same scale as for the previous question. Information such as exit names, train's map, visual and spoken information, to name a few, were common among the different lines in the investigated train station. Thus, passengers could develop indirect knowledge of the investigated line that needed to be accounted for. The third question was based on utilization of SNCF train station around Paris and suburb (apart from Saint-Michel Notre Dame), with the same five segment Likert scale. Some features could also be common among different train stations from the same train operator, such as the charter used on information displays and signage boards, but also the instructional design used to present information. A threshold score was defined from these three questions to distinguish regular travelers (experts) from occasional travelers (novices). The scale used and the threshold estimated for participants' categorization in different expertise groups are illustrated in Table 1 below.

We chose these questions for our expertise survey since we expected to evaluate memory as an indicator of performance in cognitive load.

Table 1
Categorization of travelers' sample.

5	4	3	2	1	0
Everyday	At least 3 times a week	Between one and two times a week	Once or twice a month	Once or twice a year	Not at all

Threshold expert: Participant needs to obtain a score of 5 or4 in at least two of the three questions described above. One of which should be question one. The sum of answers for the three questions should be between 8 and 15 inclusive). Threshold novice: Sum of answers for the three questions should be between 0 and 4 inclusive, with a score of at most 1 for items 1 and 2, and at most 2 for item 3.

Besides, according to our experimental protocol, participants had to find a particular landmarks and regions of the train station that should be assessed rapidly if they had a prior knowledge of the train station, or schemas learned from other train stations of the same train operator. In our experimental sample, regular travelers were assimilated as *experts*, since they take suburban train every day to go to work or university, in Paris central. Besides, most of them have been living for their whole life in the same region of the suburb, and all of them have been taking the same train for more than 10 years, practically every day. This period of time corresponds to the estimated time required for a chess player beginner to reach expert level (Simon & Gilmartin, 1973). On the other hand, our novice sample consisted mostly of people living in Paris center and taking the investigated train station for at most once or twice a year.

Once the participants filled in our short survey for evaluation of their expertise level, they were assigned to 1 of 4 different groups of investigated profile (male novice, male expert, female novice, female expert). We continued our recruitment until we obtained 112 participants, with 28 participants per group. We assumed that the three items of the survey were sufficient for an expertise assessment of our 112 randomly distributed in the condition factor (*real-life* vs *virtual reality*). In order to test statistically if these items were good predictors of expertise, we conducted a multiple linear regression test. We used the following model for our regression and results of this analysis are presented in Table 2 below.

Results showed that the three investigated items are responsible for the prediction of expertise, $F(3,108) = 378.8$, $p < .001$, and account for 91.1% of the prediction. The associated residual error is 0.15. Analyses of t-values indicate that each variable contributed significantly to the prediction of expertise ($p < .001$).

2.1.2. Cognitive neurological assessments

To complete the inclusion process, we conducted a survey to make sure that our 112 participants were not under medication. We also made sure that they were not suffering from any particular neurological or psychiatric problems, and presented normal basic cognitive skills, by completing a set of neuropsychological tests at the beginning of the experiment. These tests comprised the short version of the Beck Depression Inventory (BDI), used to evaluate depression (Beck, Steer, & Brown, 1996; Collet & Cottraux, 1986), the revised Piper Fatigue scale, used to evaluate fatigue (Piper, Dibble, Dodd, & Weiss 1998), State-Trait Anxiety Inventory (STAI Y-A) used to evaluate state anxiety level (Spielberger, 1983), the Prospective and Retrospective Memory Questionnaire (PRMQ) used to evaluate subjective prospective and retrospective memory skills (PM and RM, respectively) in daily life (Smith, Del Sala, Logie, & Maylor, 2000) and the ZOO map test, used to evaluate planning skills (Wilson, Evans, Emslie, Alderman, & Burgess, 1998). These neurological tests served as inclusion criteria for participants with no depression, low fatigue, low anxiety, normal mental planning level and no apparent subjective complaints of prospective and retrospective memory. Six participants were not kept for the final experiment after these analyses. We thus replaced them after investigation of 15 additional profiles. A 2-way MANOVA (*context* vs *profile*) was conducted on these different variables and gave a non-significant effect on *context* (VR vs real-life), $F(3,108) = 1.58$, $p = .15$, $\eta^2 = 0.10$, and non-significant profile effect (*regular* vs *occasional*), $F(3,108) = 0.87$, $p = .06$, $\eta^2 = 0.06$. No interaction was noted between context and profile factors, $F(3,108) = 1.58$, $p = .15$, $\eta^2 = 0.10$. These non-significant differences between contexts and profiles allowed us to check that no bias could intervene as a secondary factor before investigating cognitive load. Table 3 below illustrates the mean scores, the lowest scores and the highest scores of our sample, together with threshold values for interpretation.

Table 2
Mean scores and Regression analyses of final 112 participants.

Variable	Mean ± SD for Novice	Mean ± SD for Expert	Estimate	Std Error	t-value	*p*	CI 2.5%	CI 97.5%
Utilization of investigated line	0.91 ± 0.84	3.98 ± 0.80	.101	.015	6.89	< .001***	.072	.130
Utilization of whole train station	0.95 ± 0.82	4.07 ± 0.81	.091	.015	5.91	< .001***	.060	.121
Utilization of other train station of the same train operator	1.05 ± 0.77	4.07 ± 0.81	.106	.015	7.03	< .001***	.076	.136

$\text{Expertise}_i = \beta_0 + \beta_1 \text{*Investigation train line}_i + \beta_2 \text{*Investigation train station}_i + \beta_3 \text{*Train stations of the same train operator}_i + \varepsilon_i$.
This gave the following regression model:
$\text{Expertise}_i = -0.746 + .101 \text{*Investigation train line}_i + 0.091 \text{*Investigation train station}_i + 0.106 \text{*Train stations of the same train operator}_i + 0.15$.

2.2. Experimental environments

2.2.1. Real-life environment

The selected environment was Saint-Michel Notre Dame, train station. This train station accommodates nearly 100 K travelers each day. This phenomenon represents a real case of *mass transit*, characterized by the continuous movement of thousands of people in space and time, which in this case is the limited space available in the train station. The highest level in passengers' flow can be reached during peak hours, usually at 8 a.m. and 5 p.m.; before and after office hours.

2.2.2. Virtual reality environment

In the laboratory, a VR multimodal environment was designed to simulate different everyday life scenarios at Saint-Michel Notre Dame train station, based on observations made in the actual train station. In order to create a virtual environment that could be used as a replacement of a real-life condition, according to Fuchs and Moreau (2006), several indications had to be followed: immersion, interaction, and interface. The immersive level of the VE puts in action the participant's senses, and mainly sight, hearing and touch. This was done by using the Head Mounted Display of HTC Vive. The interactive aspect depends on the relation with other avatar(s) and objects in the virtual train station. This represents a key procedure in the development of an ecological VE. The interface used should be easily understood and ergonomic (Fuchs, Moreau, & Guitton, 2011). Pictures taken from the actual train station, together with textures available on Unity® asset store, that matched objects from the train station, were used to create a realistic interface. Participants navigated in the VE using a controller in each hand, by moving them up and down. During the conception of the VE, multiple tests were conducted to make sure that physical effort invested by participants while moving the controllers, were proportional to the displacement of the participant's avatar in the VE. Other prior tests were conducted on angular displacement of the headset with slow and quick movement, so as to avoid a mismatch between the visual characteristics of the virtual scenes and what it would have been in the real world. These tests allowed us to give an ecological aspect to our VE and to reduce the risk of cybersickness and more precisely motion sickness (Bruck & Watters, 2011). Though we optimized as much as possible the VE, the risk of cybersickness could not be totally removed. This side effects of virtual technologies, characterized by nausea and head ache, was explained to the participants and the latter were asked to let the experimenter know if any of cybersickness symptoms arose, in order to stop the experiment. During our experiment, 3 participants suffered slightly from cybersickness and their session was canceled instantaneously. Necessary assistance was provided to theses participants and we made sure that they felt healthy again before leaving. To balance our sample three other participants with the same profile were later recruited.

For the purpose of our experiment we created a VE from the architectural blueprints of Saint-Michel Notre Dame line C. We used the 3D software Unity3D® to design the train station. We applied the rules of interaction, immersion and interface (Fuchs & Moreau, 2006) to design our 3D environment. The detailed procedures for the construction of our virtual train station are illustrated in the Appendix section of this paper. Some pictures obtained from the virtual train station (left side) are compared to pictures taken from the same region in the real train station (right side), as shown in Fig. 1 below.

The overall purpose of the experiment was clearly explained to the participants, without giving information that could bias the experiment. The participants agreed to the terms by signing an informed consent form, explaining all possible health issues related to our experimental setup and material. Ethical clearance was granted before running the experiment. Every possible risk associated to the use of virtual reality was explained, such as the limit of movement in the experimental room while wearing the virtual reality headset, to avoid knocking against the wall was prevented using a virtual limitation shown in the virtual environment itself. The participants were volunteers and could stop the experiment at any moment.

2.3. Navigation path

During the study, participants taking part in the virtual experiment were required to navigate in the train station using a head mounted VR headset HTC Vive to be immersed in the VE. Walking action was simulated using HTC controllers, by moving them up and down. The path was the same for both experimental conditions (real-life and virtual). The trip of the participant is illustrated in Fig. 2 below.

The starting point was on platform A and the participants had to move to the mezzanine on the upper floor, cross the bridge and find the stairs leading to platform B. Afterwards, they had to move on platform B to a specific exit, *Petit-Pont Notre Dame*. From there, they needed to meet a virtual avatar who handed them a book. Eventually, the participants had to go back to platform A, using the stairs and walked till the information display screen next to the ticket validating machines on platform A. During the experiment there was an expertise effect on the time taken by travelers to visit each section of the train station. The mean time taken by occasional travelers ($M = 4.57 \pm 0.17$ min) was significantly higher than that of regular travelers

Table 3
Table of scores to neuropsychological tests of final 112 participants.

	BDI	Fatigue scale	STAI Y-A	STAI Y–B	ZOO map test	PM	RM
Mean score	0.99	2.54	33.7	38.6	161	18.4	8.49
Lowest value	0	0.75	26	34	0	11	9
Highest value	4	4.33	35	45	338	30	27
Threshold values	No depression: 0-4	Low fatigue level: 0–4.5	Low anxiety: < 39	Low anxiety: < 39	± 2 standard deviations from mean value		

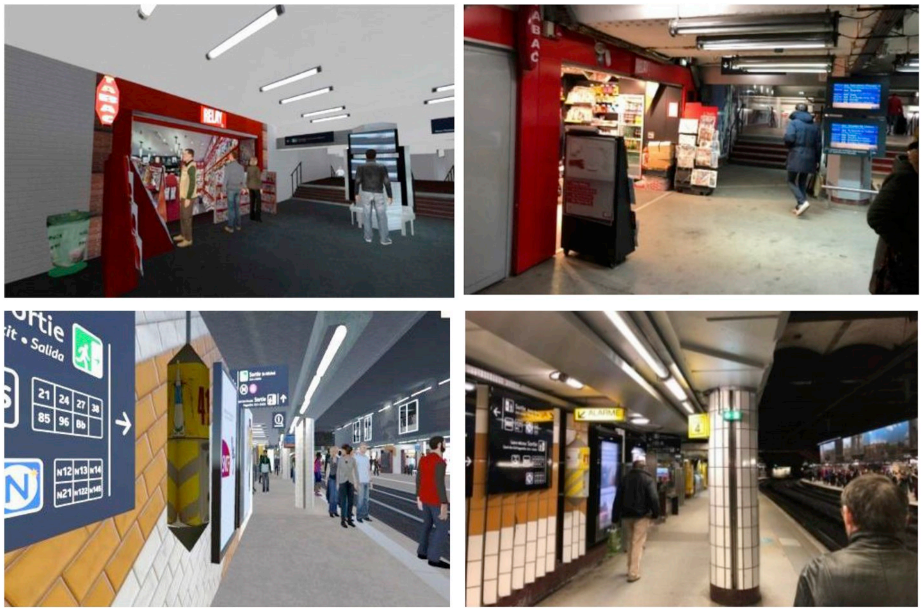


Fig. 1. irtual vs real life train station of Saint-Michel Notre Dame, line C.

(M = 3.74 ± 0 .04 min) for each experimental phase ($p < .001$). There was no environmental effect (real-life x virtual reality) ($p > .18$).

2.4. Indication

Before the experiment, indication was given to the participants in a storytelling format (for e.g. Bucher, 2017; Ibanez, Aylett, & Ruiz-Rodarte, 2003; Shin, 2018), where they had to find the next move for their journey in the train station. There was no time limit to end the experiment, but the participants were nevertheless encouraged to perform as quickly as possible, as it would be the case in real-life. The indication was chosen so that the participants cover over 90% of the train station, allowing us to evaluate the use of information in almost all the train station.

The indication was as follows: "*You are heading to Orsay museum where a group of friends* is *waiting for you. You have decided to take train line C at Saint-Michel Notre Dame. You arrived at the train station by a connection leading to platform A. Before going to find your train to Orsay museum, you decided to buy a magazine at the Relay shop in the train station, before heading to the exit Petit-Pont Notre Dame. There, a friend is waiting for you to hand you a book. Your task is to find the path to your final destination (Orsay museum), while accomplishing all the previous tasks.*"

2.5. Cognitive load measures and materials

Cognitive load was evaluated during and after the experiment. Three main methods were used: *physiological, subjective* and *behavioral*.

2.5.1. Physiological measure

Electrodermal activity (EDA) was chosen as a *physiological measure* during the experiment. Endosomatic EDA was measured from the inner wrist skin of the participants using a sensor which quantified this activity for later analyses. In both virtual and real-life conditions of our experiment, an E4 wristband sensor Empatica® was used to monitor the EDA level. This wristband contained two electrodes on the sensor's band for EDA monitoring and has a permanent internal memory of 60 h, with a synchronization resolution of 5 s. The data collected on the device can be streamed via bluetooth® on a smartphone. The sensor measured blood volume at a frequency of 64 Hz, with an electrodermal activity of 4 Hz, at an acceleration of 32 Hz and a skin temperature measured every 4 s. EDA data were measured in micro Siemens (μs) at a frequency of 4 Hz every 0.25 s.

Variables known to impact EDA recordings were supervised according to recommendations from previous studies (Boucsein, 2012).

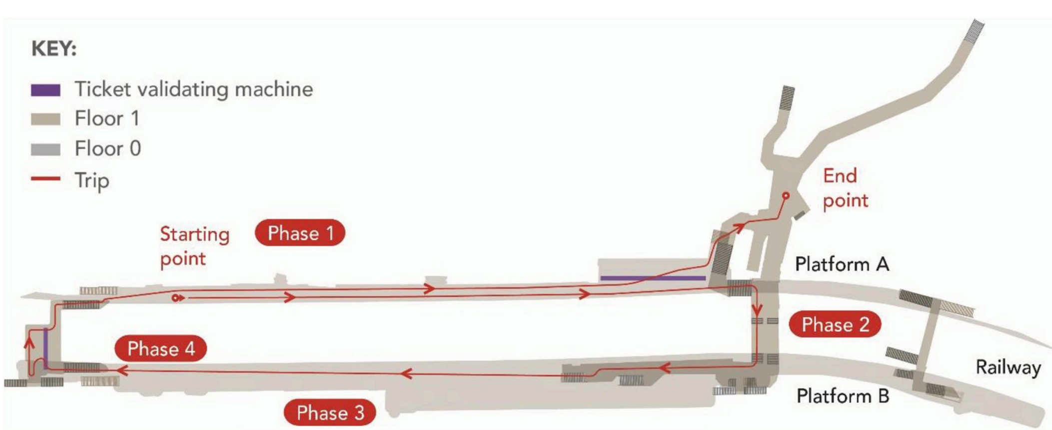


Fig. 2. Navigation path of travelers.

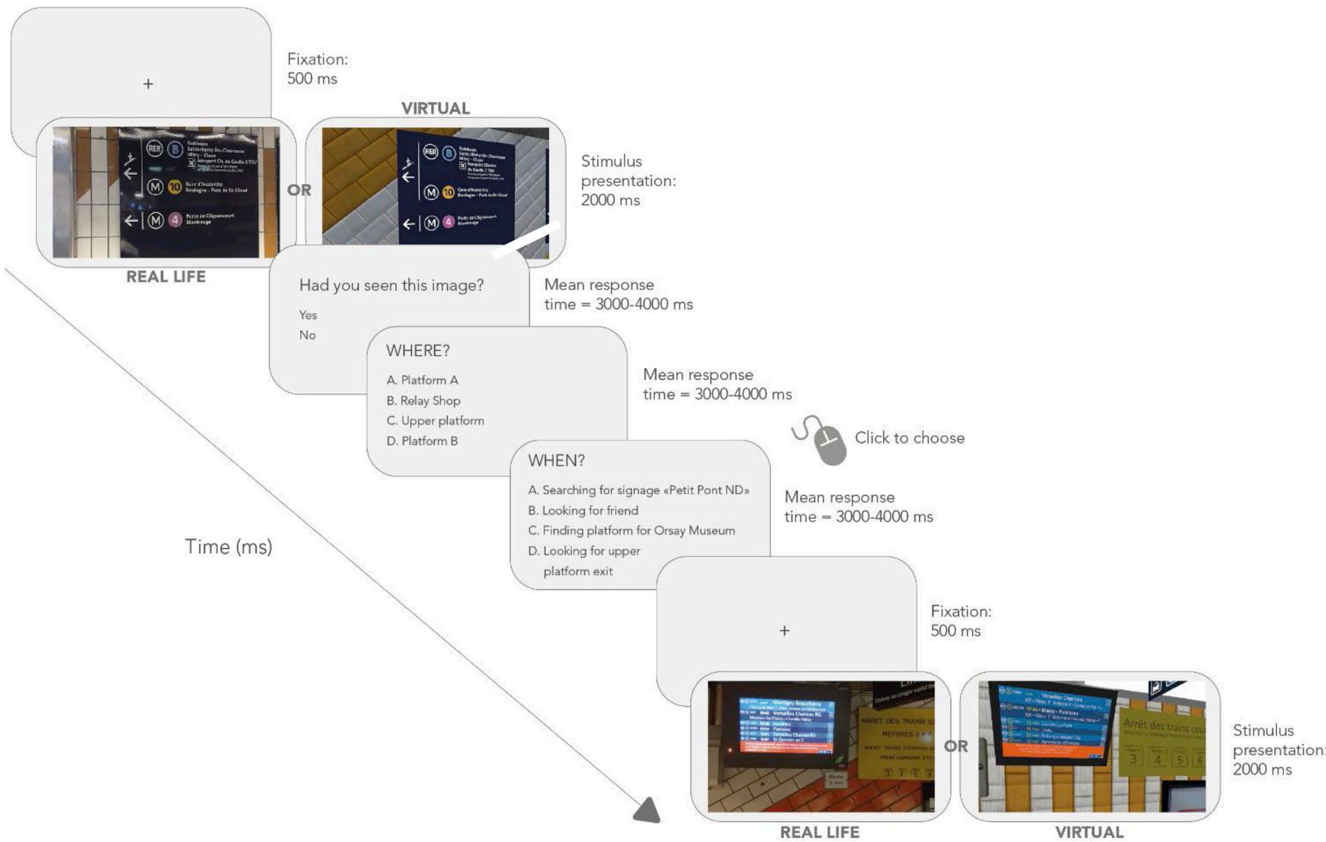


Fig. 3. Behavioral cognitive load measure - Memory test.

Variations in temperature were measured during the experiment, since EDA level is affected by it. Room temperature for virtual reality experiment was monitored and maintained at 20 °C using an air-conditioner, and the participants were asked to wear sports clothing, to be more comfortable and to prevent drastic rise in body temperature during the experiment due to sweating. The same condition was applied for real-life experimental condition, with the exception of the air-conditioned temperature. However, the average temperature of the air in both environments was 20.3 ± 0.2 °C with a non-significant difference between the two environmental conditions ($p = .76$). Data from electrodermal activity were collected at four different check points in the virtual environment. These were platform A, upper platform, platform B and the exit (Petit-Pont Notre Dame). We expected regular travelers to show lower EDA score than occasional travelers, with no statistical difference in EDA score between the two investigated environmental conditions.

2.5.2. Subjective measure

NASA- Task Load Index (NASA-TLX) (Hart & Staveland, 1988) was chosen to evaluate *subjective measure* during our experiment. This test gave an indication of self-perceived mental load by the participants. This measure was operationalized using weighted mean scores for 6 dimensions: mental demand, physical demand, temporal demand, performance, effort and frustration, perceived by the individuals with respect to the task.

We expected occasional travelers to express higher subjective mental load than regular travelers. No significant difference was predicted between real-life and virtual experiments.

2.5.3. Behavioral measure

In addition to the first two measures, a *behavioral measure* was conducted on the participants' performance by analyzing factual and contextual memory using a memory test, designed for this purpose. Our aim was to assess factual and contextual memory encoding performance, i.e. featuring binding. Binding refers to link a stimulus to its contextual information (Sperduti, Armougum, Makowski, Blondé, & Piolino, 2017). This test gave an indication of contextual information memorized by participants, linked to episodic memory (Johnson & Raye, 1981; Tulving, 2002).

The memory test was designed on OpenSesame Experiment and presented on a Hewlett-Packard HP EliteBook 820 G2, processor Intel® Core™ i5-5200U CPU @ 2.20 GHz, after the experiment in the train station (virtual or real). During this test, participants who took part in the virtual version of the experiment, were presented a series of pictures (N = 30 hits and n = 15 distractors) taken from the virtual train station. Participants who took part in the real-life experiment were presented 30 other pictures (hits) and 15 distractors, taken from the actual train station from the same angle. For both environmental conditions, the chosen pictures were those of signage boards and information displays with the train station's background as it was seen by the participants during the experiment. Names of exits, directions for different platforms and trains' destinations were modified for distractors. For each session, the 45 images were randomized to avoid any effect related to order of presentation or participant's fatigue. The questions asked during the test investigated the "*what*" of the object, corresponding to the recognition of the object and the ability to identify information among fake ones. The participants had to choose between *yes* or *no,* to say if the image has been presented during navigation or not. The "*where*" was also measured, which was related to the spatial localization of the stimulus in the train station, viewed during navigation. For this question, the participant had to choose among platform A, Relay shop, upper platform and platform B. Finally, the "*when*" was recorded, which informed on the temporal localization of the stimulus in the environment. Participants had to choose among: *when looking for Petit-Pont Notre Dame, when looking for friend, when finding platform for Orsay museum* and *when looking for upper platform exit.* Combining these three scores *(what, where and when)*, gave an indication of the binding score. The higher the binding score, the stronger is the link between the stimulus and the contextual information (for e.g.: Makowski, Sperduti, Nicolas, & Piolino, 2017; Meiser, Sattler, & Weiβer, 2008).

In order to have an adequate timing for stimuli presentation in our memory test, we pre-tested its different phases with 15 participants who were not part of our final experimental cohort. During the pre-test, the image identification difficulty was assessed according to the number of correct answers of the pre-tested participants. This allowed us to obtain a measure of the mean difficulty level of a list of pictures. For the final test, each image was presented for 2000 ms after which the participant had to choose if the image was presented or not, earlier in the train station (virtual or real-life condition). There was no time limit for the participants to decide: if the image was initially presented or not, in which temporal and spatial contexts it was presented. The participants were, however, encouraged to choose an answer as quick as possible. The mean time taken by the participants to answer each question was 3000 ms. The different steps for our memory test are illustrated in Fig. 3 below.

Higher binding scores (*what, when, where, total binding*) were expected for regular travelers than occasional travelers, with no environmental condition effect.

2.6. Procedure

The experiment started with a set of six neuropsychological tests (BDI, STAI-YA, Fatigue scale, PRMQ and Zoo map test) assigned to each participant. For the next steps, the procedures varied according to the experimental environment: real-life and VR. The real-life experiment took place between 10 and 12 a.m. at off-peak hours. The VE was designed to take place during this time interval. Only one participant took part in the experiment at a time. Participants were equipped with an E4 wristband sensor and the recordings of EDA signals were triggered. They were then asked to move around the train station for 10 min, checking their environment in a designated area, indicated by the experimenter, excluding the experimental perimeter of the train station. This was done to obtain a baseline of EDA signals for the participants in a non-experimental context. The data were measured online, with a feedback on our smartphone that allowed us to monitor variations in EDA level. The recorded EDA scores were accessible offline after the experiment. After the 10 min of baseline encoding, written and oral indications for the actual experiment were given to the participants in a storytelling format, as seen in section 2.4. Once the experiment started, another trigger was activated on the EDA wristband to mark the beginning of the experiment. During the study, the experimenter followed the participant at a distance of 2 meters, while avoiding any contact with the latter.

For the *virtual reality experiment*, 56 participants were also recruited. 28 of them were regular and 28 were occasional, with 50% of women in each category. Here also, one participant took part in the experiment at a time. Before starting the actual experiment in the virtual condition, the participants were able to navigate for 10 min in an empty version of the train station, without furniture, object, train, avatars and signage boards. There was only the building, the walls' texture and lights. This was done to help the participants accommodate to virtual reality and be at ease with the system. The participants tested all the settings of the controllers: walk slowly, walk quickly, accelerate, decelerate, walk upstairs quickly and slowly. Afterwards, the participants were equipped with an E4 wristband sensor, and the EDA baseline was recorded for 10 min in the empty virtual train station. Once this phase ended, written and oral indications for the experiment were given to the participants in a storytelling format, as seen in section 2.4. The virtual train station was then switched on and a trigger was activated on the EDA wristband to mark the start of the experiment, with a timer integrated in the app on which the data could be monitored online.

During both the virtual and real-life experiments, any unexpected behavior or reaction was noted by the experimenter. The participants were not supposed to ask any question to the experimenter. The experiment ended when the participants accomplished all the tasks mentioned in section 2.3. An additional neuropsychological test was then assigned to the participants: STAI Y–B to investigate anxiety level after the experiment. To evaluate the subjective mental load measure related to the experiment, the participants had to fill up the NASA-TLX test. A questionnaire assessing the sense of presence for an indirect subjective measure of immersion in virtual reality was filled up by participants who took part in the virtual train station experiment (Schubert et al., 2001).

Participants were then asked to take part in a memory test designed especially for this study as shown in Fig. 3. The purpose of the experiment was finally explained to the participants and their feedback were noted in order to explain possible deviant results and for potential upgrades of the study.

3. Data processing

During this experiment, we evaluated cognitive load of participants through EDA analyses for physiological dimension, NASA-TLX for subjective dimension and memory binding score for behavioral dimension.

Data collected were analyzed for two factors: travelers' profile (regular or occasional) and environmental situation (virtual reality or real-life) using the statistical computing and graphics software, R. Descriptive analyses were first performed using the package tidyverse (Wickham, 2017). The mean values and standard deviations for all collected data were hence obtained from these descriptive analyses. Normality and homoscedasticity tests were conducted using Car package (Fox et al., 2012). ANOVA was then calculated using the same package. Lsr (Navarro, 2015) was used for effect size calculation. Posthoc pairwise analyses were performed, using emmeans package (Estimated Marginal Means aka Least-Squares Means) (Lenth, 2016). LmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) and Psycho (Makowski, 2018) for Mixed Effects Models.

3.1. Electrodermal activity

EDA data were uploaded in Matlab-based software Ledalab, designed to analyze skin conductance data (Benedek & Kaernbach, 2010). The format of EDA E4 Empatica corresponds to "*Text Type 1*". The sampling frequency was 4 Hz, with threshold amplitude of 0.01 μs. The signal was first conditioned before storage. The uploaded data were preprocessed using low-pass Butterworth filter, down-sampled, cut, smoothened, and artifact-corrected. The data were then optimized .

A Continuous Decomposition Analysis (CDA) was then performed by analyzing and converting the raw Skin Conductance (SC) data into continuous tonic and phasic activities providing an unbiased quantification of SCR amplitudes and number of SCR peaks.

3.2. NASA-Task Load Index

Scores for NASA-TLX were calculated using a weighted mean score. This was done by comparing pairwise the 6 investigated dimensions, namely mental demand, physical demand, temporal demand, performance, effort and frustration level. The task of the participant was to choose which dimension between each pair was more important with respect to the demand of the task. For each comparison, a point was granted to the selected dimension. After each two by two comparisons, the weighted mean score was calculated by multiplying score assigned to each dimension out of 20, by the number of times that each dimension was chosen during pairwise comparisons.

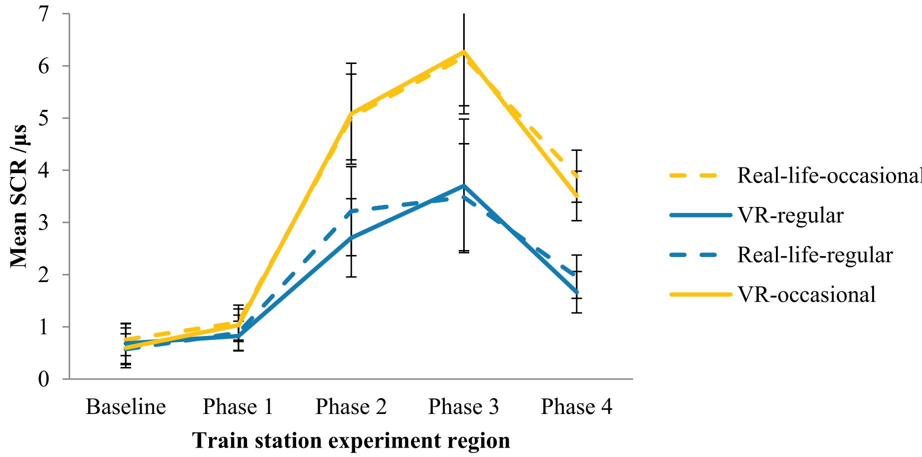


Fig. 4. Figure 4. EDA of regular and occasional travelers in virtual reality and real-life conditions, through mean SCR (CDA).

3.3. Binding score

Data obtained from Opensesame after memory test were automatically saved in a spreadsheet, obtained after each participant has taken part in the memory test. Each participant's spreadsheet contained the percentage of correct answers, wrong answers, spatial binding temporal binding and total binding for each stimulus presented. The time taken in milliseconds to respond to each question was also recorded. Additional data concerning keyboard letters pressed were also saved, to check if the participants' wrong answers were not due to wrong typing.

4. Results

4.1. Electrodermal activity

Data obtained by continuous deconvolution analysis showed a mean SCR score of .66 µs at baseline for the four groups of travelers (occasional travelers in real-life condition, occasional travelers in virtual reality, regular travelers in real-life condition and regular travelers in virtual reality condition). This score increased exponentially during phase 1 to phase 3 and decreased from phase 3 to phase 4. The mean SCR scores of occasional travelers increased more than those of regular travelers at any moment. The results obtained are illustrated in Fig. 4 below.

Since EDA was measured repeatedly throughout the experiment for the same participant, the independence of data was no more satisfied. A linear mixed model analysis was thus conducted on our two experimental factors (profile and condition), to test for random and fixed effects to account for spacing between each time points (Jaeger, 2008; Krueger & Tian, 2004), while taking time points (EDA_Phase) as error. Our model was as follows:

EDA_value = condition * profile + (1 | EDA_Phase)

Where,

EDA_value was skin conductance responses.
EDA_phase was the order of EDA measurements.

Analysis of fixed model gave no condition effect between *virtual reality* and *real-life* contexts $F(3,120) = .07$, $p = .79$. A profile effect was, however, detected between *regular* and *occasional* travelers, $F(3,120) = 111$, $p < .001$. No interaction was observed between condition and profile factors, $F(3,120) = .02$, $p = .89$. Analysis of the main effects of the linear mixed model is presented in Table 4 below.

The overall model predicting (EDA_value ~ condition* profile + (1 | EDA_Phase)) has a total explanatory power (conditional R^2) of 85.25%, in which the fixed effects explain 5.76% of the variance (marginal R^2). The model's intercept is at 1.75 (SE = .75, 95% CI [.12, 3.37]). Within this model, *condition* factor was not significant (β = −.02, SE = .09, 95% CI [-.19, .15], $t(552) = -.27$, $p > .1$) and can be considered as very small (std. β = −0.01, std. SE = 0.05). *Profile* factor, on the other hand, proved to be significant (β = .91, SE = .09, 95% CI [.74, 1.08], $t(552) = 10.54$, $p < .001$) and can be considered as medium (std. β = .53, std. SE = .05). No significant interaction was found between *condition* and *profile* factor (β = −.02, SE = .12, 95% CI [-.26, .22], $t(552) = -.13$, $p > .1$) and can be considered as very small (std. β = −.01, std. SE = .07).

4.2. NASA-Task Load Index

Cognitive load measured by the participant using NASA-TLX subjective measure of mental load was collected and illustrated in Fig. 5 below.

Globally, mental load dimensions were higher for occasional travelers than for regular travelers (except for temporal demand dimension). A 2 × 2 ANOVA was conducted on traveler's profile (regular vs occasional) and environmental condition (real-life vs virtual reality). The analysis of mental demand showed a significant effect for profile, $F(3,108) = 4201$, $p < .001$, $\eta_p^2 = .97$, higher for occasional travelers than for regular travelers. Analysis of environmental condition showed, $F(3,108) = 3.13$, $p = .08$, $\eta_p^2 = .03$, indicating no significant difference in virtual reality and real-life conditions. The interaction between environmental condition and profile was also non-significant: $F(3,108) = .02$, $p = .90$, $\eta_p^2 = .00$. These findings indicate a higher mental load level for occasional travelers compared to regular travelers. This difference is maintained in real-life and virtual environments. Similar analyses were conducted on physical demand, temporal demand, performance, effort and frustration level, and the results are resumed in Table 5 below.

For frustration level, the mean adjusted load ratings were higher for occasional travelers than for regular travelers. This difference was significant, $F(3,108) = 8495$, $p < .001$, $\eta_p^2 = .99$, indicating a travelers' profile effect. Adjusted load ratings were also higher for real-life condition than for VR condition, for occasional travelers. The data were the other way round for regular travelers (i.e. higher for VR condition than for real-life condition). This effect was significant, $F(3,108) = 8.87$, $p < .001$, $\eta_p^2 = .08$. Analysis of interaction between environmental condition X Travelers' profile gave, $F(3,108) = 37$, $p < .001$, $\eta_p^2 = .25$, indicating a significant interaction between our two factors. We thus performed a post hoc Tukey HSD test to verify pairwise interactions between each pair of data. The travelers' profile effect persisted ($p < .001$) when environmental condition was kept constant and when only the traveler's profile was modified (e.g. occasional travelers

Table 4
Linear mixed model for EDA of travelers in different train station contexts.

Variable	Coef	SE	t	df	p	Coef_std	SE_std	Effect_Size	CI lower	CI higher
(Intercept)	1.75	.75	2.31	4.04	.08	−.25	.44	small	.12	3.37
condition	−.02	.09	−.27	552	.79	−.01	.05	Very small	−.19	.15
profile	.91	.09	10.5	552	< .001***	.53	.05	medium	.74	1.08
condition * profile	−.02	.12	−.13	552	.89	−.01	.07	Very small	−.26	.22

in real-life condition vs regular travelers in real-life). However, the environmental condition effect persisted only for regular travelers ($p < .001$), when comparing regular travelers in real-life vs regular travelers in VR. No environmental condition was observed when comparing occasional travelers in real-life condition to occasional travelers in VR ($p = .13$). These results suggest that frustration level varies significantly in regular travelers with respect to the environmental condition.

4.3. Binding scores

Comparison of mean scores for correct answers, spatial binding, temporal binding and total binding are illustrated in Fig. 6 below. In each of these four dimensions, regular travelers showed a better performance than occasional travelers, reaching, however, the highest score of 43% in *correct answers* dimension.

A 2 × 2 ANOVA was conducted on travelers' profile (regular x occasional) and the environmental condition (virtual reality x real-life) on the four measured dimensions (*what, when, where and total binding*).

For analysis of *correct answers*, we obtained a significant difference for travelers' profile: $F(3,108) = 1032$, $p < .001$, $\eta_p^2 = .905$. This showed a significant difference in correct answers, based on travelers' profile, higher for regular travelers than for occasional travelers. In environmental condition, however, the analysis was not significant: $F(3,108) = .40$, $p = .53$, $\eta_p^2 = .004$. Analysis of interaction between environmental condition and travelers' profile indicated a non-significant interaction: $F(3,108) = .16$, $p = .69$, $\eta_p^2 = .002$. These findings showed that regular travelers expressed higher contextual and factual memories than occasional travelers, with similar results in real-life and virtual environments.

The same analyses performed on the other three dimensions showed the same significant p-values for different travelers' profile. No significant difference was obtained for environmental condition, for any memory performance. The results are summarized in Table 6 below.

4.4. Additional measures

Evaluation of travelers' subjective immersion level from virtual reality session was assessed using the *presence questionnaire* (Schubert et al., 2001). 4 dimensions were evaluated from 14 items, on a 7 divided scale, ranging from −3 to 3. These dimensions were *general experience, spatial presence, personal involvement and relevant realism*. Immersion level was high in each of the four dimensions and significantly different from the mean value of 0 (corresponding to a moderate sense of presence in the VE); general presence, $t(55) = 16.3$, $p < .001$; spatial presence, $t(55) = 46.5$, $p < .001$; involvement, $t(55) = 46.4$, $p < .001$ and experienced realism, $t(55) = 51.9$, $p < .001$.

When comparing the two levels of profile in a 1-way MANOVA, we found that regular travelers expressed no significant difference compared to occasional travelers, $F(1,54) = 1.23$; $p = .311$, $\eta^2 = .09$. A tendential significance was, however, noted in spatial presence dimension, indicating a difference between the two investigated profiles of travelers, being higher for regular travelers, $F(1,54) = 4.15$, $p = .05$, $\eta_p^2 = .07$. Results of descriptive analysis and full one-way MANOVA on each of the dimensions are presented in Table 7 below.

5. Discussion

Information processing generates cognitive load that has a direct impact on information acquisition and learning (Sweller, 1988) and the amount of information recalled (Kirsh, 2000). In our experiment, we took the example of train travelers who are exposed to numerous amounts of information that help them get to a desired point in a train station. Our experiment was conducted in real-life and in virtual reality (condition factor), with occasional and regular travelers (expertise factor), comparing their cognitive load level via physiological,

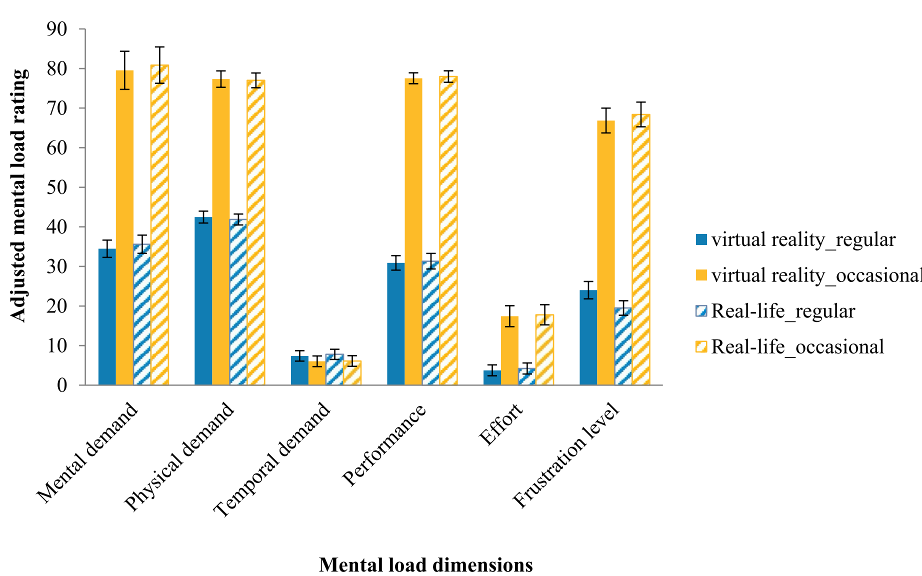


Fig. 5. NASA-TLX Mental load results distribution.

Table 5
Analyses for subjective cognitive load (adjusted load rating) for environmental situation factor (real life vs virtual reality) and travelers' profile factor (regular vs occasional).

Subjective cognitive score	Environmental condition	Travelers' profile	Environmental condition* Travelers' profile
Mental demand	F (3,108) = 3.13, p = .08, η_p^2 = .03	F (3,108) = 4201, p < .001, η_p^2 = .97	F (3,108) = .02, p = .90, η_p^2 = .00
Physical demand	F (3,108) = 2.02, p = .16, η_p^2 = .02	F (3,108) = 11486, p < .001, η_p^2 = .99	F (3,108) = .19, p = .66, η_p^2 = .00
Temporal demand	F (3,108) = .86, p = .36, η_p^2 = .01	F (3,108) = 37, p < .001, η_p^2 = .25	F (3,108) = .41, p = .52, η_p^2 = .17
Performance	F (3,108) = 1.84, p = .18, η_p^2 = .02	F (3,108) = 21733, p < .001, η_p^2 = .99	F (3,108) = .00, p = 1.00, η_p^2 = .00
Effort	F (3,108) = 1.09, p = .30, η_p^2 = .01	F (3,108) = 1201, p < .001, η_p^2 = .92	F (3,108) = .02, p = .89, η_p^2 = .43
Frustration level	F (3,108) = 8.87, p < .001, η_p^2 = .08	F (3,108) = 8495, p < .001, η_p^2 = .99	F (3,108) = 37, p < .001, η_p^2 = .25

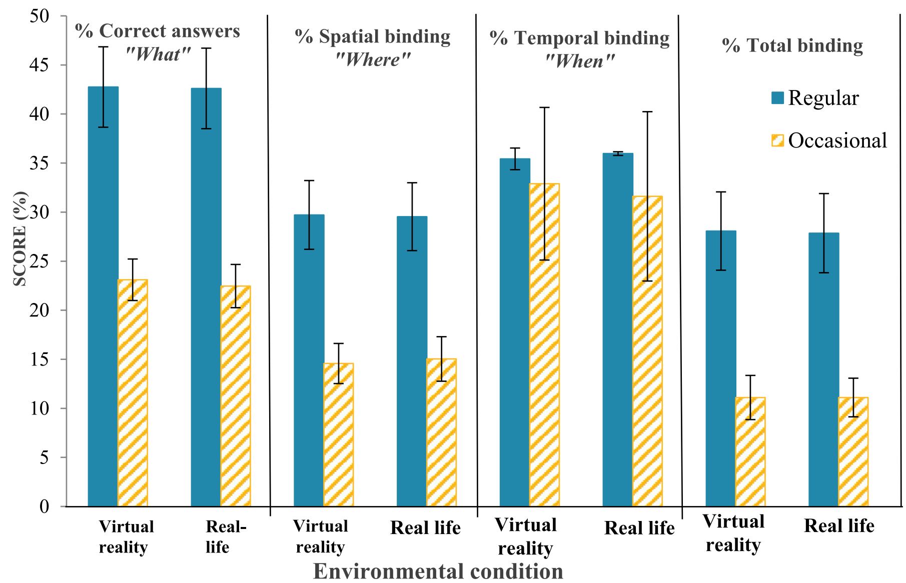


Fig. 6. NASA-TLX subjective measure of mental load.

subjective and behavioral responses. The findings showed no environmental condition effect between virtual and real-life conditions. EDA, NASA-TLX (except for subjective level of frustration in regular travelers), and binding memory scores were not statistically different for a given level of expertise, when compared between two experimental conditions – expert travelers had same cognitive load in virtual and real-life conditions, and so are the results for novice travelers. In our study, we tested for expertise effect as a confirmatory effect in order to evaluate variations in cognitive load, while contrasting real-life and virtual environments. Results showed higher cognitive load for novice travelers than for expert travelers, in both experimental conditions.

5.1. Modulations of cognitive load components

Cognitive load varies in real-life activities, due to modulations in cognitive resources (Chanquoy et al., 2007). These modulations can be analyzed through different pools of resources associated to intrinsic (Sweller, 1994), extraneous (Sweller & Chandler, 1991) and germane cognitive load (Sweller et al., 1998). In this study we investigated cognitive load as a whole, while considering the three different components of cognitive load during the experimental set-up.

Intrinsic cognitive load in our study was related to the basic degree of knowledge of the train station's charter to be learned, in order to understand information conveyed to travelers. This would allow travelers to encode information and use it to find their way through the train station. Travelers with knowledge of the investigated train station or those who had visited another train station using same charter in terms of information display, have a greater expertise level than those less exposed to these stimuli. Creating two groups of expertise levels, helped to monitor this aspect of cognitive load in the current study.

On the other hand, extraneous cognitive load depends on additional information, associated to the nomenclature of information itself during learning process. Since in our experiment we used actual train station's displays without any modification, this secondary factor was thus

Table 6
Analyses of binding scores for environmental situation factor (real life vs virtual reality) and travelers' profile factor (regular vs occasional).

Score	Environmental condition	Travelers' profile	Environmental condition* Travelers' profile
Correct answers	F (3,108) = .40, p = .53, η_p^2 = .004	F (3,108) = 1032, p < .001, η_p^2 = .905	F (3,108) = .16, p = .69, η_p^2 = .002
Spatial binding	F (3,108) = .07, p = .80, η_p^2 = .001	F (3,108) = 733, p < .001, η_p^2 = .872	F (3,108) = .34, p = .56, η_p^2 = .003
Temporal binding	F (3,108) = .12, p = .73, η_p^2 = .001	F (3,108) = 9.77, p < .001, η_p^2 = .083	F (3,108) = .68, p = .41, η_p^2 = .006
Total binding	F (3,108) = .55, p = .46, η_p^2 = .000	F (3,108) = 278, p < .001, η_p^2 = .878	F (3,108) = .82, p = .37, η_p^2 = .000
Response time	F (3,108) = .55, p = .46, η_p^2 = .005	F (3,108) = 278, p < .001, η_p^2 = .721	F (3,108) = .82, p = .37, η_p^2 = .008

Table 7
Descriptive analysis and MANOVA for immersion dimensions in VR.

Profile	General_Presence	Spatial_Presence	Involvement	Experienced_Realism
Occasional	2.11 ± .92	2.16 ± .34	2.21 ± .38	2.14 ± .34
Regular	1.93 ± .94	2.34 ± .37	2.29 ± .35	2.09 ± .27
MANOVA	$F(1,54) = .52$, $p = .47$, $\eta_p^2 = .01$	$F(1,54) = 4.15$, $p = .05$, $\eta_p^2 = .07$	$F(1,54) = .53$, $p = .47$, $\eta_p^2 = .01$	$F(1,54) = .22$, $p = .64$, $\eta_p^2 = .00$

controlled. Extraneous cognitive load depends on the complexity in the presentation of information to learners, that is train travelers in our study.

Germane cognitive load corresponds to the relevant information recalled by the travelers related to their journey in the train station and varies oppositely to extraneous cognitive load. For this purpose, we investigated the behavioral aspect of cognitive load through performance analysis, using memory binding score. The use of memory binding score as a behavioral measure in cognitive load analysis has never been used, as it is the case in the present study. One main advantage of using this behavioral measure was to take into account interindividual differences in information processing. This is due to the fact that participants, irrespective of their expertise level, have a unique interpretation of information and pay attention to different details in their surroundings. This would be a great advantage for the analysis of cognitive load in complex situations, where interindividual differences are much important. Memory binding score proved to be a convenient and promising behavioral measure for cognitive load.

5.2. Experimental condition factor: real-life or virtual reality environment

The absence of experimental condition effect suggests that travelers can experience the same cognitive load level, irrespective of the format of presentation of information (virtual reality and real-life). This observation is in line with findings obtained from previous studies on 3D modeling of virtual environment to investigate cognitive processes, initially assessed in real-life condition (Andersen et al., 2018; Chen & Stanney, 1999; Crescentini et al., 2016; Darken & Peterson, 2014; Darken & Sibert, 1996; Freeman et al., 2008; Klinger et al., 2005; Steinicke et al., 2013, pp. 199–219; Viaud-Delmon et al., 1998). We extended the findings by revealing similar cognitive load measurements based on physiological (Gevins et al., 1998; Paas et al., 2003; Sweller et al., 2011), subjective (Vidulich & Tsang, 2012) and behavioral (Brunken et al., 2003; Paas & Van Merriënboer, 1994) responses. Several reasons have been spotted for the success of the validation of virtual reality for cognitive load analysis. Two main reasons could be the design of the virtual environment itself and the use of storytelling format for giving indications to participants.

Storytelling has proven its efficacy in highly rated video games experience, where players felt more invested and immersed in a virtual universe. Storytelling also provides a mnemonic technique to help participants understand and learn indication in a more natural and contextualized way, through the use of a story rather than listing a set of tasks to perform. Our study is in line with former studies, which have proven the rise in mnemonic performance using storytelling format (Bucher, 2017; Ibanez et al., 2003; Shin, 2018). It would be interesting to investigate the use of storytelling, mostly in experiment with elaborate indications.

5.3. Expertise effect: a cognitive load indicator for virtual environment validation

Our findings confirmed that expertise effect, where experts have higher performances than novices, has an important impact on cognitive load (Kalakoski & Saariluoma, 2001; Perdreau & Cavanagh, 2015; Reingold et al., 2001). Moreover, whenever all secondary factors were kept constant, experts showed same performance in both environmental contexts. This phenomenon was used as an indicator to validate our virtual environment.

When traveling in big cities, with a mass transit context such as Paris, it is essential to think about all the complex situations that may arise each day in a train station, and especially how these situations may impact on the travelers' cognitive load. Interestingly, we found only one difference regarding subjective measures, since the experts felt more frustration in virtual environment than in the real environment. It suggests that experts could feel annoyed as a result of being unable to achieve their navigation as they used to do in the real-life, i.e. alone and without being observed by an experimenter. Accordingly, the experts seemed more attached and sensitive to their habits than the novices and seemed to be more vulnerable to social pressure. It would thus be interesting to investigate cognitive load in more complex and unpredicted situations, in order to see how expert effect would vary in these conditions. Previous studies had shown the existence of expertise reversal effect in non-optimal situations, where experts showed significantly lower performance than novices (Kalyuga, 2009; Sweller, Ayres, Kalyuga, & Chandler, 2003). For examples, delayed or canceled trains, engineering work on railway, traveler's accident and strike of employees, to name a few. Results from cognitive load analyses of travelers in these complex situations could help us optimize information given to travelers, in terms of instructional design. These concern the format (redundancy effect: Chandler & Sweller, 1991), the modality (Mayer, 2002) and the attentional levels (split attention effect: Chandler & Sweller, 1992; Mayer & Moreno, 1998) of information.

5.4. Strength, limitations, and future avenue of research

In our study we assessed young adults, with a mean age of 26 (± 5) years old. This is, however, not representative of all travelers in Saint-Michel Notre Dame train station. It would be interesting to extend our sample to higher age groups and evaluate cognitive load of elder travelers. For this to be efficient, it would be necessary to adapt our training session for our VE, by extending the training sessions and by creating additional ones. The participant would for instance perform some particular exercises in a virtual environment such as walking up and down stairs, moving sideways to avoid obstacles, picking up objects. These activities, however, can contribute to amplify one main side effect of VR: cybersickness. This main limitation in virtual reality technologies requires further studies to find practical solutions to facilitate its reduction. Besides from optimization of the virtual environment, one main method that proved its efficacy against cybersickness, is to provide repeated training sessions with a gradual increase in difficulty.

Other than extending and diversifying training sessions, it would also be important to think about a common threshold to define expertise, since expert travelers from the elder population would have a greater time exposure than younger experts, in the tested train station (e.g. expertise among chess players: Chase & Simon, 1973). It would be interesting to investigate expert memory models (Gobet, 1998) in relation with cognitive load theory.

When choosing the train station for our experiment, we look for different criteria such as localization, passengers' flow, travelers' profiles, size and volume of travelers per area of the train station (mass transit characteristics) and number of anomalies reported by travelers to local agents. This investigation led us to the conclusion that train stations, though located in the same city, are significantly different from each other. We can thus ask ourselves if our results obtained from Saint-Michel Notre Dame could be applied to other mass transit train stations in Paris city. It would be interesting to design different train stations and perform the same analyses on cognitive load, to contrast our results. These data obtained on cognitive load could be implemented in a supervised learning function of machine learning for diagnostics classification. We would hence be able to predict cognitive load of travelers in a given train station, using data collected from other train stations. This perspective is in line with past studies in machine learning prediction (for cancer prognosis and prediction: Kourou, Exarchos, Exarchos, Karamouzis, & Fotiadis, 2015; classification of mental tasks from EEG: Liang, Saratchandran, Huang, & Sundararajan, 2006).

6. Conclusion

This study validates the use of virtual reality as a technique for investigating cognitive load in daily life situations, using expert effect as an indicator of learning process. Different techniques that accounted for cognitive load in our study proved to be practical and nevertheless promising for studies involving navigation. A whole new field of work in non-static cognitive load investigation, involving non-optimal environmental scenarios can now be tested. These modulations in environmental conditions could give rise to learning effects, explained by the Cognitive Load Theory, and would represent a great opportunity to study these effects in applied experiments.

Funding

This work was supported by ANRT (Association Nationale de la Recherche et de la Technologie) for a CIFRE doctorate program between SNCF Mobility–Transilien and Paris Descartes University. Grant number 216/0860.

Acknowledgements

Thanks to Lab Mass Transit designer team for graphical assistance and artwork.

Appendix B. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jenvp.2019.101338.

Appendix

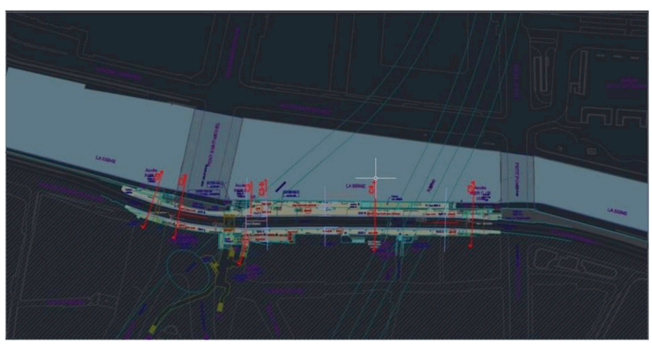


Fig. 7. Top view of Saint-Michel Notre Dame RERC train blueprints station.

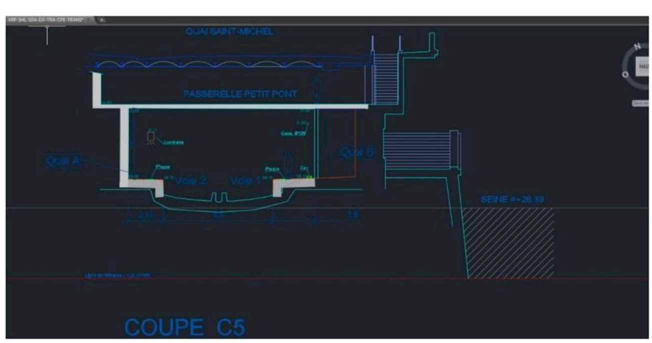


Fig. 8. Side view of Saint-Michel RERC train station blueprints.2

We chose to use the multiplatform game interface Unity 3D® for the development of the train station and for integration of 3D elements in the environment. For 3D construction of the walls and buildings in VR, we used the integrated Unity plug-in Pro-Builder. This function allowed us to easily create buildings as in Fig. 9 below.

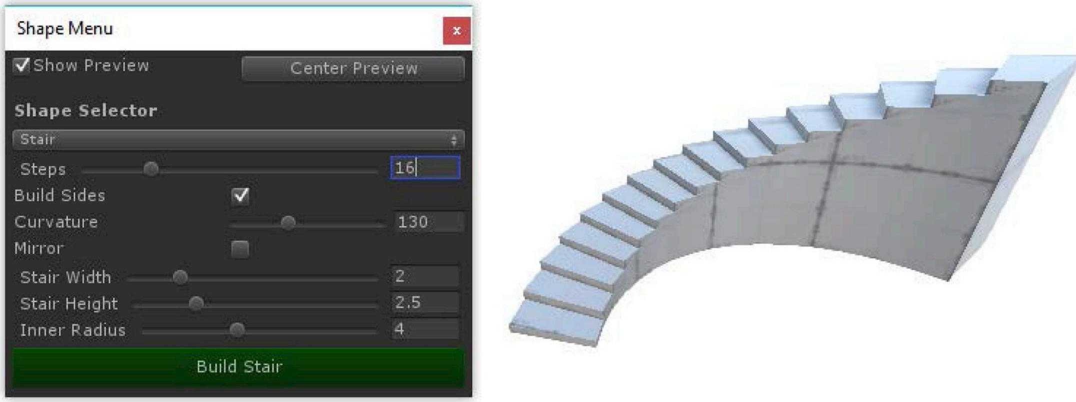
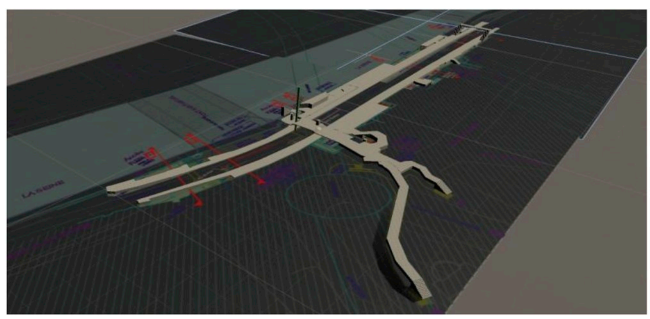
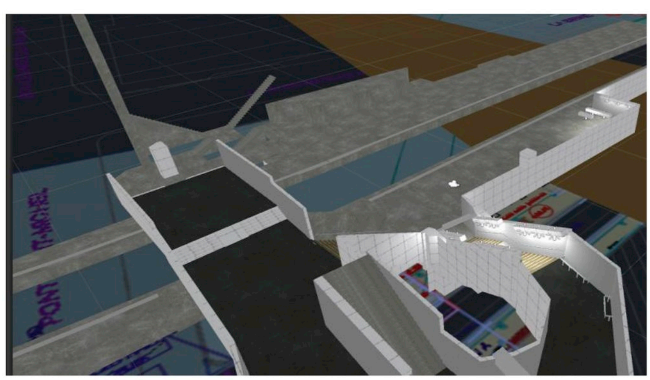


Fig. 9. Creation of stairs in Pro-Builder.3

We started by designing the train station floor with the designated metrics from the architecture plans before implementing the walls with the proportioned heights and altitudes. From pictures and videos made directly in the train station we have spotted the position of the light system and had integrated a complex light system with adjusted brightness and contrast as shown in Fig. 10i and ii below.



i



ii

Fig. 10. i. Creation of train station's floor from the architecture plans. ii. Creation of train station's walls and lights from the architecture plans.4

The different objects and furniture in the train station were designed using videos, pictures and observations made directly in the train station. Some objects were taken from the assets store of Unity and from Sketchup (Sample of train). Some objects' textures were modified using 3DSMAX and animation of characters and objects were made using integrated functions from unity3D. The acoustic environment was designed in the virtual environment using recordings made in the actual train station. These recordings were, the sound of travelers walking in the train station, a train arriving and departing from the train station, conversations heard in the train station and the acoustic recording for each region in the train station.

The contents of signage board and display screens designed in the virtual environment were obtained from passengers' information department of SNCF. We synchronized the time displayed on dynamic screens throughout the train station for a better immersion of participants in the train station. Passengers' flow was designed using integrated Unity plug-in for flow management, compatible with the avatars created to represent the travelers in the train station as shown in Fig. 11 below.

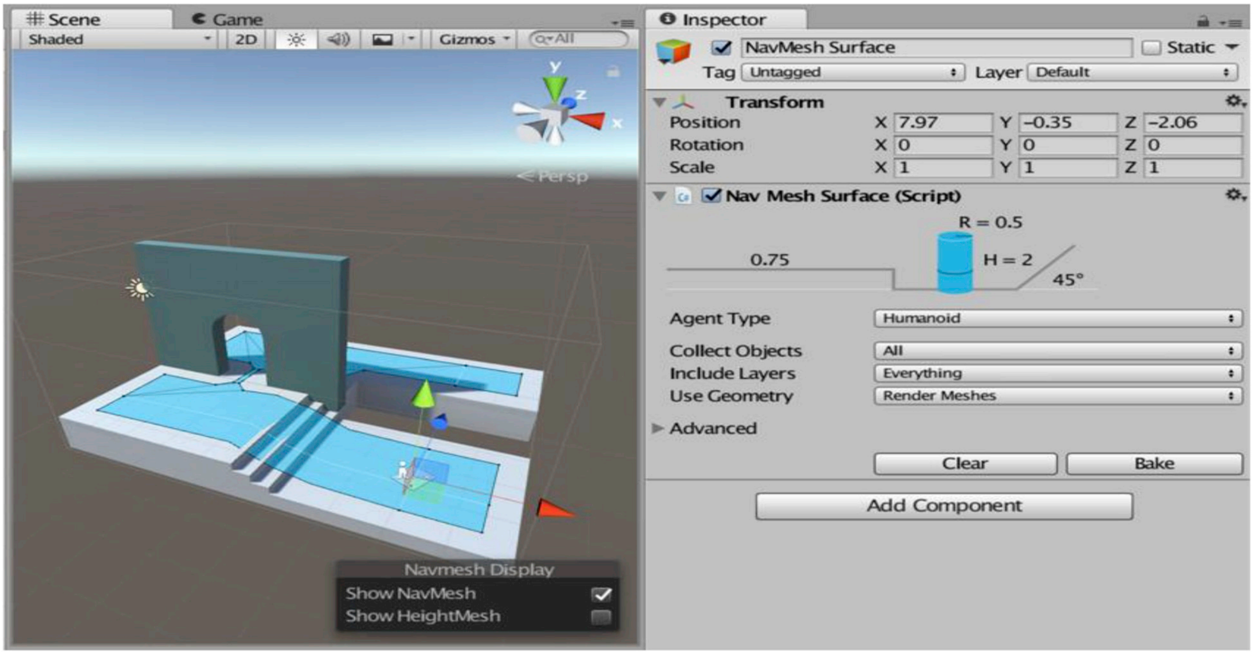


Fig. 11. Navigation Mesh surface for crowd simulation.5

Some scenarios were created for the crowd movement strategy, based on observation made in the actual train station. Particular scenarios at strategic points in the train stations were triggered by the experimenter using the keyboard. This was for instance, activating passengers' flow to another train's line after an announcement was made about a canceled train. Modifications were made on the virtual environment until it was close to the actual train station.

References

Andersen, S. A. W., Konge, L., & Sørensen, M. S. (2018). The effect of distributed virtual reality simulation training on cognitive load during subsequent dissection training. *Medical Teacher,* 1–6.

Baddeley, A., Logie, R., Bressi, S., Sala, S. D., & Spinnler, H. (1986). Dementia and working memory. *The Quarterly Journal of Experimental Psychology Section A, 38*(4), 603–618.

Barfield, W., & Hendrix, C. (1995). The effect of update rate on the sense of presence within virtual environments. *Virtual Reality, 1*(1), 3–15. https://doi.org/10.1007/BF02009709.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck depression inventory-II. *San Antonio, 78*(2), 490–498.

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods, 190*(1), 80–91.

Boucsein, W. (2012). *Electrodermal activity.* Wuppertal, Germany: Springer Science & Business Media.

Bruck, S., & Watters, P. A. (2011). The factor structure of cybersickness. *Displays, 32*(4), 153–158. https://doi.org/10.1016/J.DISPLA.2011.07.002.

Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 53–61.

Bucher, J. (2017). *Storytelling for virtual reality: Methods and principles for crafting immersive narratives.* New York & London: Focal press.

Bystrom, K.-E., Barfield, W., & Hendrix, C. (1999). A conceptual model of the sense of presence in virtual environments. *Presence: Teleoperators and Virtual Environments, 8*(2), 241–244. https://doi.org/10.1162/105474699566107.

Cain, B. (2007). *A review of the mental workload literature.* Toronto, Canada: Defence Research And Development.

Carrozzino, M., & Bergamasco, M. (2010). Beyond virtual museums: Experiencing immersive virtual reality in real museums. *Journal of Cultural Heritage, 11*(4), 452–458.

Cegarra, J., & Chevalier, A. (2008). The use of Tholos software for combining measures of mental workload: Toward theoretical and methodological improvements. *Behavior Research Methods, 40*(4), 988–1000.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293–332.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*(2), 233–246.

Chanquoy, L., Tricot, A., & Sweller, J. (2007). *La charge cognitive: Théorie et applications. Paris: France.* Armand Colin.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*(1), 55–81.

Chen, J. L., & Stanney, K. M. (1999). A theoretical model of wayfinding in virtual environments: Proposed strategies for navigational aiding. *Presence, 8*(6), 671–685.

Chen, F., Zhou, J., Wang, Y., Yu, K., Arshad, S. Z., Khawaji, A., et al. (2016). Galvanic skin response-based measures. *Robust multimodal cognitive load measurement*Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-31700-7_5.

Collet, L., & Cottraux, J. (1986). The shortened Beck depression inventory (13 items). Study of the concurrent validity with the Hamilton scale and Widlöcher's retardation scale. *L'Encephale, 12*(2), 77–79.

Cook, M. P. (2006). Visual representations in science education: The influence of prior knowledge and cognitive load theory on instructional design principles. *Science Education, 90*(6), 1073–1091.

Crescentini, C., Chittaro, L., Capurso, V., Sioni, R., & Fabbro, F. (2016). Psychological and physiological responses to stressful situations in immersive virtual reality: Differences between users who practice mindfulness meditation and controls. *Computers in Human Behavior, 59*, 304–316.

Darken, R. P., & Peterson, B. (2014). Spatial orientation, wayfinding, and representation: Handbook of virtual environment technology. In K. Stanney (Ed.). Monterey, California.

Darken, R. P., & Sibert, J. L. (1996). Navigating large virtual spaces. *International Journal of Human-Computer Interaction, 8*(1), 49–71.

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., & Heiberger, R. (2012). *Package 'car'.* Vienna: R Foundation for Statistical Computing.

Freeman, D., Pugh, K., Antley, A., Slater, M., Bebbington, P., Gittins, M., et al. (2008). Virtual reality study of paranoid thinking in the general population. *The British Journal of Psychiatry, 192*(4), 258–263.

Fuchs, P., & Moreau, G. (2006). *Le traité de la réalité virtuelle, Vol 2.* Paris, France: Presses des MINES.

Fuchs, P., Moreau, G., & Guitton, P. (2011). *Virtual reality: Concepts and technologies.* London, UK: CRC Press.

Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., et al. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors, 40*(1), 79–91.

Gobet, F. (1998). Expert memory: A comparison of four theories. *Cognition, 66*(2), 115–152.

Hart, S. G., & Staveland, L. E. (1988). *Development of NASA-TLX (task load index): Results of empirical and theoretical research. Advances in psychology, Vol 52*, 139–183 North-Holland.

Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems, 6*(2), 156–166.

Ibanez, J., Aylett, R., & Ruiz-Rodarte, R. (2003). Storytelling in virtual environments from a virtual guide perspective. *Virtual Reality, 7*(1), 30–42.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4),

434–446. https://doi.org/10.1016/J.JML.2007.11.007.
Jebara, N., Orriols, E., Zaoui, M., Berthoz, A., & Piolino, P. (2014). Effects of enactment in episodic memory: A pilot virtual reality study with young and elderly adults. *Frontiers in Aging Neuroscience, 6*, 338.
Jex, S. M. (1998). *Stress and job performance: Theory, research, and implications for managerial practice.* Sage Publications Ltd.
Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*(1), 67. https://doi.org/10.1037/0033-295X.88.1.67.
Kalakoski, V., & Saariluoma, P. (2001). Taxi drivers' exceptional memory of street names. *Memory & Cognition, 29*(4), 634–638.
Kalyuga, S. (2009). The expertise reversal effect. *Managing cognitive load in adaptive multimedia learning* (pp. 58–80). IGI Global.
Ke, Y., Qi, H., He, F., Liu, S., Zhao, X., Zhou, P., et al. (2014). An EEG-based mental workload estimator trained on working memory task can work well under simulated multi-attribute task. *Frontiers in Human Neuroscience, 8*, 703.
Kirsh, D. (2000). A few thoughts on cognitive overload. *Intellectica, 1*(30), 19–51.
Klinger, E., Bouchard, S., Légeron, P., Roy, S., Lauer, F., Chemin, I., et al. (2005). Virtual reality therapy versus cognitive behavior therapy for social phobia: A preliminary controlled study. *CyberPsychology and Behavior, 8*(1), 76–88.
Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal, 13*, 8–17.
Koutsoudis, A., Arnaoutoglou, F., & Chamzas, C. (2007). On 3D reconstruction of the old city of Xanthi. A minimum budget approach to virtual touring based on photogrammetry. *Journal of Cultural Heritage, 8*(1), 26–31.
Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. *Psychophysiological Metrics Of Mental Workload, 279*(12).
Krueger, C., & Tian, L. (2004). A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. *Biological Research for Nursing, 6*(2), 151–157. https://doi.org/10.1177/1099800404267682.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), https://doi.org/10.18637/jss.v082.i13.
Lee, E. A. L., & Wong, K. W. (2014). Learning with desktop virtual reality: Low spatial ability learners are more positively affected. *Computers & Education, 79*, 49–58.
Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *Journal of Statistical Software, 69*(1), 1–33 progress. Multiple-task performance, 279-328.
Liang, N. Y., Saratchandran, P., Huang, G. B., & Sundararajan, N. (2006). Classification of mental tasks from EEG signals using extreme learning machine. *International Journal of Neural Systems, 16*(01), 29–38.
Makowski (2018). *The psycho package: An efficient and publishing-oriented workflow for*.
Makowski, D., Sperduti, M., Nicolas, S., & Piolino, P. (2017). "Being there" and it, remembering it: Presence improves memory encoding. *Consciousness and Cognition, 53*, 194–202.
Mayer, R. E. (2002). Multimedia learning. *Psychology of learning and motivation. Vol 41. Psychology of learning and motivation* (pp. 85–139). Academic Press.
Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312.
Meiser, T., Sattler, C., & Weißer, K. (2008). Binding of multidimensional context information as a distinctive characteristic of remember judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(1), 32.
Murakami, H. (2009). *1Q84.* Japan, Kyoto: Shinchosha.
Murray, C. D., & Gordon, M. S. (2001). Changes in bodily awareness induced by immersive virtual reality. *CyberPsychology and Behavior, 4*(3), 365–371.
Musse, S. R., & Thalmann, D. (2001). Hierarchical model for real time simulation of virtual human crowds. *IEEE Transactions on Visualization and Computer Graphics, 7*(2), 152–164.
Navarro, D. (2015). *Learning statistics with R.* Adelaide, South Australia: lulu.
Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429.
Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1–4.
Paas, F. G., & Van Merriënboer, J. J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*(1), 122.
Perdreau, F., & Cavanagh, P. (2015). Drawing experts have better visual memory while drawing. *Journal of Vision, 15*(5) 5-5.
Picard, R. W. (2015). March). Recognizing stress, engagement, and positive emotion. *Proceedings of the 20th international conference on intelligent user interfaces* (pp. 3–4). ACM.
Piper, B. F., Dibble, S. L., Dodd, M. J., Weiss, M. C., Slaughter, R. E., & Paul, S. M. (1998). The revised piper fatigue scale: Psychometric evaluation in women with breast cancer. *Oncology nursing forum.* Oncology Nursing Society.
Plancher, G., Barra, J., Orriols, E., & Piolino, P. (2013). The influence of action on episodic memory: A virtual reality study. *The Quarterly Journal of Experimental Psychology, 66*(5), 895–909.
Plancher, G., Gyselinck, V., Nicolas, S., & Piolino, P. (2010). Age effect on components of episodic memory and feature binding: A virtual reality study. *Neuropsychology, 24*(3), 379.
Plancher, G., Gyselinck, V., & Piolino, P. (2018). The integration of realistic episodic memories relies on different working memory processes: Evidence from virtual navigation. *Frontiers in Psychology, 9*, 47.
Plancher, G., Tirard, A., Gyselinck, V., Nicolas, S., & Piolino, P. (2012). Using virtual reality to characterize episodic memory profiles in amnestic mild cognitive impairment and alzheimer's disease: Influence of active and passive encoding. *Neuropsychologia, 50*(5), 592–602.
Reingold, E. M., Charness, N., Pomplun, M., & Stampe, D. M. (2001). Visual span in expert chess players: Evidence from eye movements. *Psychological Science, 12*(1), 48–55.
Riva, G. (2005). Virtual reality in psychotherapy. *CyberPsychology and Behavior, 8*(3), 220–230.
Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments, 10*(3), 266–281.
Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., & Ehlert, U. (2009). Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on Information Technology in Biomedicine, 14*(2), 410–417. https://doi.org/10.1109/TITB.2009.2036164.
Shin, D. (2018). Empathy and embodied experience in virtual environment: To what extent can virtual reality stimulate empathy and embodied experience? *Computers in Human Behavior, 78*, 64–73.
Shi, Y., Ruiz, N., Taib, R., Choi, E., & Chen, F. (2007). *April). Galvanic skin response (GSR) as an index of cognitive load. CHI'07 extended abstracts on Human factors in computing systems*https://doi.org/10.1145/1240866.1241057.
Simon, H. A., & Gilmartin, K. (1973). A simulation of memory for chess positions. *Cognitive Psychology, 5*(1), 29–46. https://doi.org/10.1016/0010-0285(73)90024-8.
Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments, 6*(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603.
Smith, G., Del Sala, S., Logie, R. H., & Maylor, E. A. (2000). Prospective and retrospective memory in normal ageing and dementia: A questionnaire study. *Memory, 8*(5), 311–321.
Sperduti, M., Armougum, A., Makowski, D., Blondé, P., & Piolino, P. (2017). Interaction between attentional systems and episodic memory encoding: The impact of conflict on binding of information. *Experimental Brain Research, 235*(12), 3553–3560.
Spielberger, C. D. (1983). *STAI state-trait anxiety inventory for adults form Y: Review set; manual, test, scoring key.* Mind Garden.
Steinicke, F., Visell, Y., Campos, J., & Lécuyer, A. (2013). *Human walking in virtual environments.* New York, NY, USA: Springer.
Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285.
Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312.
Sweller, J., Ayres, P., & Kalyuga, S. (2011). Measuring cognitive load. *Cognitive load theory* (pp. 71–85). New York, NY: Springer.
Sweller, J., Ayres, P. L., Kalyuga, S., & Chandler, P. (2003). *The expertise reversal effect*.
Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and Instruction, 8*(4), 351–362.
Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.
Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53*(1), 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114.
Ulicny, B., & Thalmann, D. (2001). Crowd simulation for interactive virtual environments and VR training systems. *Computer animation and simulation 2001* (pp. 163–170). Vienna: Springer.
Van Merriënboer, J. J., & Sweller, J. (2010). Cognitive load theory in health professional education: Design principles and strategies. *Medical Education, 44*(1), 85–93.
Van Veen, H. A., Distler, H. K., Braun, S. J., & Bülthoff, H. H. (1998). Navigating through a virtual city: Using virtual reality technology to study human action and perception. *Future Generation Computer Systems, 14*(3–4), 231–242.
Viaud-Delmon, I., Ivanenko, Y. P., Berthoz, A., & Jouvent, R. (1998). Sex, lies and virtual reality. *Nature Neuroscience, 1*(1), 15.
Vidulich, M. A., & Tsang, P. S. (2012). Mental workload and situation awareness. *Handbook of human factors and ergonomics: 4*, (pp. 243–273).
Whyte, J., Bouchlaghem, N., Thorpe, A., & McCaffer, R. (2000). From CAD to virtual reality: Modelling approaches, data exchange and interactive 3D building design tools. *Automation in Construction, 10*(1), 43–55.
Wickham, H. (2017). *The tidyverse. R package ver. 1.1. 1*.
Wilson, B. A., Evans, J. J., Emslie, H., Alderman, N., & Burgess, P. (1998). The development of an ecologically valid test for assessing patients with a dysexecutive syndrome. *Neuropsychological Rehabilitation, 8*(3), 213–228.