

Full length article

Cognitive characteristics in firefighter wayfinding Tasks: An Eye-Tracking analysis

Yang Ye^a, Yangming Shi^b, Pengxiang Xia^a, John Kang^c, Oshin Tyagi^c, Ranjana K. Mehta^c, Jing Du^{d,*}^a Engineering School of Sustainable Infrastructure & Environment, University of Florida, 1949 Stadium Road 365 Weil Hall, Gainesville, FL 32611, United States^b Department of Civil, Construction, and Environmental Engineering, University of Alabama, 261 Hardaway Hall, Tuscaloosa, AL 35406, United States^c Department of Industrial & Systems Engineering, Texas A&M University, 2005C ETB Building, College Station, TX 77843, United States^d Engineering School of Sustainable Infrastructure & Environment, University of Florida, 1949 Stadium Road 460F Weil Hall, Gainesville, FL 32611, United States

ARTICLE INFO

Keywords:

Firefighter

Wayfinding

Virtual Reality

Gaze analysis

Cognitive analysis

ABSTRACT

During search and rescue, firefighters need to find paths in an unfamiliar space with minimum time and information available. The effective memorization and retrieval of critical spatial information can help reduce risks and increase mission efficiency. Although evidence has shown that different formats of wayfinding information, including landmarks, routes, and surveys, can impact search and rescue performance in different manners, a deeper understanding of the characteristics of firefighters' cognitive processes related to the varying wayfinding information formats is less explored. To evaluate firefighters' performance and cognitive characteristics in search and rescue, a firefighter experiment in Virtual Reality (VR) was conducted. Firefighters ($n = 40$) were recruited to participate in the simulated rescue task. After reviewing the spatial information in different formats, firefighters were requested to find three victims inside a VR maze as quickly as possible. Task performance was evaluated by the number of victims found and the time spent. Firefighters' gaze patterns were analyzed to evaluate their cognitive status. The result showed that although the cognitive load under the survey and route conditions was significantly higher than under the landmark condition ($p < 0.001$), the decision-making involved a more effective cognitive process related to choosing the right path at critical waypoints such as where a turning decision must be made. Thus, the perceived workload and fatigue levels of the two conditions were lower, and the wayfinding performance was better. In contrast, with landmark information, the cognitive load levels were consistently high, along with increased mental fatigue.

The findings reveal a series of cognitive features related to a more effective spatial decision-making in search and rescue. In the future, it is expected that these cognitive features can be used to develop real-time monitoring and prediction models for wayfinding performance.

1. Introduction

Firefighters perform arduous wayfinding tasks in hostile, chaotic, and unfamiliar environments, which can lead to significant safety risks [1]. According to the U.S. Fire Administration (USFA), a total of 23,825 (39%) firefighters were injured on the fire ground and 62 lost their lives in the line of duty in 2019 [2]. Firefighting and rescue tasks impose extensive physical and mental pressures on firefighters, and hence, wayfinding during the emergency tasks can be extremely difficult. Firefighters need to orient themselves in complex buildings, locate

victims, and retreat from safe exits. This process requires firefighters to possess good spatial knowledge of building layouts. However, it is nearly impossible to familiarize firefighters with the building layout comprehensively given the complexity of contemporary buildings and the limited time for preparation when facing an emergency. It is critical to gain a deeper understanding of the cognitive process of firefighters for developing and retrieving spatial knowledge that is used to optimize the pre-rescue review sessions for a better understanding of the task, and eventually to scale down the task risks.

As for the theoretical framework for understanding such a cognitive

* Corresponding author.

E-mail addresses: ye.yang@ufl.edu (Y. Ye), shiyangming@ua.edu (Y. Shi), xia.p@ufl.edu (P. Xia), jjkang612@tamu.edu (J. Kang), oshin_tyagi@tamu.edu (O. Tyagi), rmehta@tamu.edu (R.K. Mehta), eric.du@essie.ufl.edu (J. Du).<https://doi.org/10.1016/j.aei.2022.101668>

Received 17 February 2022; Received in revised form 24 May 2022; Accepted 6 June 2022

Available online 13 June 2022

1474-0346/© 2022 Elsevier Ltd. All rights reserved.

process, we rely on the Landmark-Route-Survey (LRS) model [3] proposed by Siegel and White's in 1975. The LRS model describes the formation of spatial knowledge as a synergy of three forms of spatial information: landmark, route, and survey [4,5]. The model suggests that when a person is developing the spatial understanding of any new spaces, the spatial knowledge can be stored and processed in the above three forms. Specifically, spatial knowledge developed from landmarks uses distinctive objects to mark the spatial positions and orientations. Route-based spatial knowledge concentrates on the sequence of locations or orientations that leads to a certain destination. Survey-based spatial knowledge is usually represented as maps, referring to a more comprehensive spatial connection between locations. Although normally different forms of spatial knowledge are leveraged and developed in parallel [6], it is infeasible for firefighters to develop such a comprehensive level of spatial knowledge given the limited time in search and rescue tasks. Instead, providing firefighters with structured spatial knowledge (i.e., a proper combination of different formats of spatial knowledge) can accelerate the spatial knowledge acquisition dramatically [7–9]. In addition, evidence shows that different forms of spatial knowledge are associated with distinct neural processes of working memory [8], which can also impact the retrieval efficiency. This is rather important in emergency wayfinding tasks because the high physical and mental demands during the mission can further impose burdens on the spatial knowledge retrieval. There is a pressing need to understand the unique cognitive implications of different forms of spatial knowledge for firefighters to plan out emergency wayfinding tasks. The lessons learned can help design a more effective pre-mission review strategy.

To fill the knowledge gap, we simulated a fire rescue wayfinding task in Virtual Reality (VR) mazes with two phases: the information reviewing phase and the wayfinding performance phase. To focus on the individual cognitive patterns related to the forms of spatial knowledge, we provided participants with distinct information in the forms of landmarks, routes, and maps during the reviewing phase. After reviewing the information, participants were asked to conduct a wayfinding task in VR. Eye-tracking was utilized in both phases to measure the dynamic cognitive implications of different information. A total of 40 frontline firefighters were recruited to participate in this study. The theoretical background, experiment design, measures, and findings are explained in the rest of this paper.

2. Literature review

2.1. Spatial knowledge development and cognitive processes

To answer the question of how humans map or model the spatial environment, Siegel and White [3] proposed the LRS model. This model describes spatial knowledge as a mental representation of the environment that is constructed by temporal integration of successive perceptions over the environment. Further behavioral [3,4,10] and neurological [11,12] studies have reached a consensus that the spatial knowledge consists of fragmented and distorted reference pieces of the space. The basic elements of such projection pieces are landmark representations, route representations, and survey representations, which are developed in parallel during a typical spatial knowledge acquisition process [6,13]. Different forms of spatial knowledge (landmark, route, and survey) can be acquired in synergies (prone to real-life-like navigation-observe-learning progress) or separation (prone to a training process with a certain spatial knowledge delivery method) [14]. In a more common and ideal scenario, a person would rely on a natural spatial knowledge acquisition process to build a comprehensive understanding of the space, including the gradual development of landmark, route, and survey knowledge that can take time. However, for emergency wayfinding, it would require the delivery of spatial knowledge with only a limited time. It can be infeasible for a firefighter to acquire spatial knowledge in a built-up manner prior to the task. In contrast,

rapid pre-mission briefings are usually conducted based on limited information, such as oral communications of the route. As such, the spatial knowledge acquisition process is often incomplete and separated.

Although the extent of interaction and integration among landmark, route, and survey knowledge can gradually increase along with the learning process, existing studies indicate that different acquisition strategies can correspond with largely different durations for comprehension and varying cognitive implications [8]. By analyzing the comprehension time for spatial knowledge in controlled experiments, Lee and Tversky [8] found that landmark knowledge and route/survey knowledge represented separate cognitive processes, and the format in which the spatial knowledge was conveyed could impact the comprehension process significantly. Nori and Giusberti [15] found that spatial learning with landmark and route knowledge had different directional cognitive mechanisms compared to that with survey knowledge [10]. In addition to the evidence from behavioral data, neurological studies also suggested that the underlying neural paths were different when learning spatial knowledge with different strategies. Epstein and Kanwisher [16] discovered that the activation patterns in the navigation and visual memory region of a human brain (parahippocampal place area) were different when exposed to different formats of spatial knowledge. Neuroimaging analysis by Maguire et al. [17] further verified that the neural substrates' pattern of topographical memory retrieval shared different patterns when the given information was changed.

In summary, behavioral and neurological evidence has generally proven the theoretical validity of the cognitive implications of spatial information forms during both the spatial knowledge encoding and decoding, which could lead to substantially different wayfinding behaviors. More evidence is needed to examine how the unique cognitive processes, driven by the three wayfinding knowledge forms, correlate with the wayfinding performance. Especially, understanding how such correlations are manifested in extreme situations, such as during the emergency wayfinding, can contribute to the design of effective pre-rescue briefings for emergency wayfinding.

2.2. Pupillometry as a cognitive measure

Measuring cognitive status is well-established area of research. In general, there are three types of cognitive measurements: performance measurements, subjective assessments, and physiological sensing [18]. Performance measurements and subjective assessments are typically collected after a task is completed to posteriorly measure the overall cognitive status. When the goal is to analyze the continuous and temporal status changes of the dynamic cognitive process, the posterior measurements are not appropriate. Physiological sensing techniques like Electroencephalogram (EEG) [19], among many others, require a complicated setup process, leading to usually uncomfortable user experiences and a potentially limited motion range.

With the advent of wearable eye-tracking devices, pupillometry has been widely used and recognized as an effective indicator of mental state [20–22]. Pupillary diameter and eye blink activities are generally recognized pupillometry measures indicating cognitive activities [20]. Behavioral [22–24] and neurological [25] studies consistently agreed that pupillary diameter and eye blink rate would be considered closely related to cognitive load and mental fatigue levels respectively. According to recent neurofunctional studies, the underlying principle for the effectiveness of these measures is that activities in the neuro-modulatory brain systems can trigger sympathetic and parasympathetic branches of the nerve system, and further trigger sphincter and dilator muscles that control the pupillary dilation [26,27]. The association between pupillary size and cognitive load has been examined in various domains, such as solving mathematics [18,23], driving behaviors [28], and speech perception [29] etc. However, noises in data caused by position/motor artifacts, blinks, and hardware tracking errors during the pupillary measurement must be processed via a standard and delicate process [30], including the outlier removal, missing value

interpolations, baseline corrections, and latency corrections [30,31]. After the data processing, tasks with higher difficulty levels are proven to be associated with a larger pupillary dilation [32,33] and more active pupillary oscillations [34,35]. In addition to the pupillary dilation, eye blinks are considered closely tied to central dopaminergic functions which can indicate different psychological conditions [36]. Studies showed that eye blink rate was an effective indicator of fatigue and mental loads [24,37,38]. Eye blink and pupillary dilation together are indicative of complementary information but are demonstrated in mutually exclusive biological paths [36]. Specifically, the eye blink rate is an indicator more pertaining to measuring the sensory processing of the cognitive process [39]. Thus, eye blink rate is more indicative of mental fatigue, rather than a more neutral measure of cognitive load changes [37,39]. In contrast, pupillary dilation often occurs during the sustained information processing phase and is more associated with the increased cognitive load [40]. In sum, both pupillary dilation and eye blinks can be effective metrics for mental activities with different focuses on cognitive load and mental fatigue respectively.

3. Methodology

3.1. VR maze for wayfinding simulation

Our objective is to examine the behavioral and cognitive characteristics of firefighters in wayfinding tasks affected by different types of spatial knowledge. To maintain controlled and risk-free conditions, we built a set of office mazes in VR to simulate firefighter wayfinding tasks with Unity 3D-5.6.3 fl. The basic unit of the VR mazes was a 10 m by 10 m office room. We modulated the rooms with a varying number of entrances/exits and connected the rooms to form different mazes. Smoke and fire effects were added to simulate fire rescue scenes in an immersive way. We placed three virtual victims in the mazes and showed the

locations of victims with different forms of information. Firefighters were asked to find these victims in the adverse virtual environment within the given time.

A within-participant experiment design was selected to focus on the effect of different conditions on the same subject, i.e., any participant was requested to engage in all conditions. To reduce the within-participant learning effect interaction [41], we randomly generated homogeneous maze layouts in Unity with a maze generation algorithm [42] for each condition. The office maze designs followed the Bowyer-Watson algorithm [43] to ensure the consistency of the total travel distance needed to find target victims and the same number of turns among the randomly generated paths. Thus, the mazes under different conditions could have a consistent level of complexity, which maintained the common ground for comparison. Fig. 1 shows the generated office maze layouts:

The navigation in this virtual rescue task was designed to be as intuitive as possible. We adopted a hand-waving algorithm to control the marching-forward motion in VR while the navigation orientation followed the facing direction of a participant. The virtual environment was delivered via a VR headset (HTC VIVE Head Mounted Display, USA). The moving trajectories of firefighters were logged in VR at the frequency of 90 Hz and were exported as CSV files at the end of each trial for further analysis. Tobii Pro eye tracker was integrated with a VR headset [44] to track eye movement and pupillary data.

3.2. Experiment procedure

As shown in Fig. 2, the first stage was pre-experiment briefings. After signing the informed consent, participants were asked to complete a background survey on their age, gender, body-mass index, work experience, and VR experience. A paper-based cube test provided by the Educational Testing Services (ETS) was performed to examine the

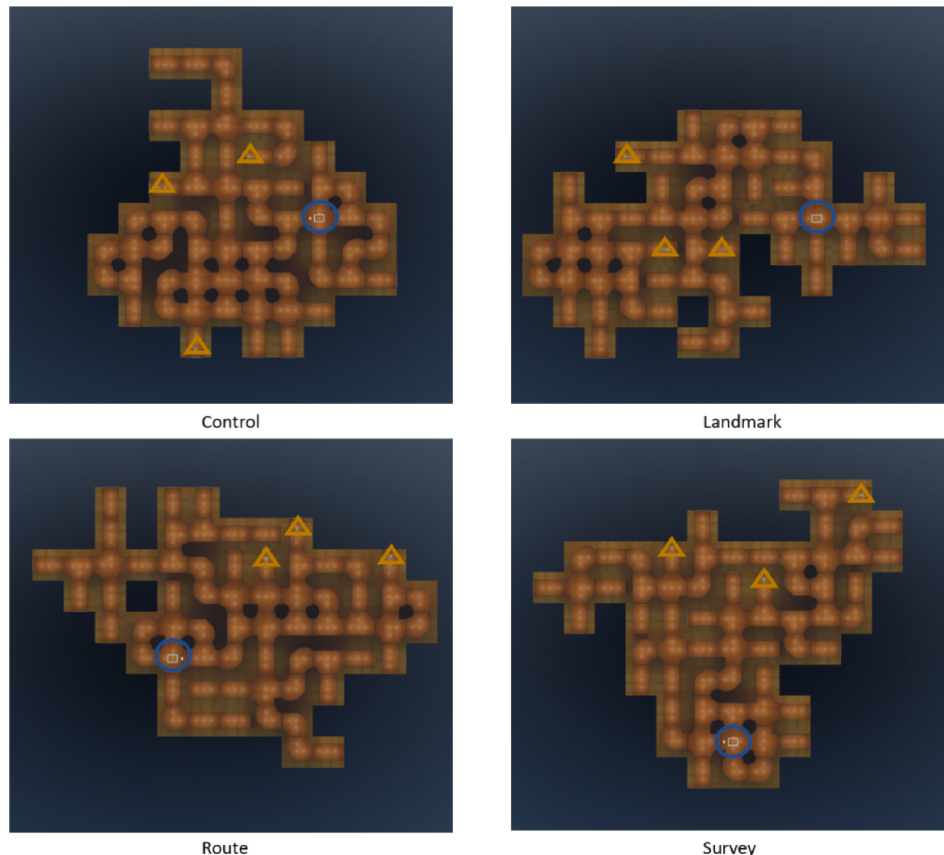


Fig. 1. Randomly generated maze layouts; Blue circles denote the starting position; Yellow triangles denote the location of victims.

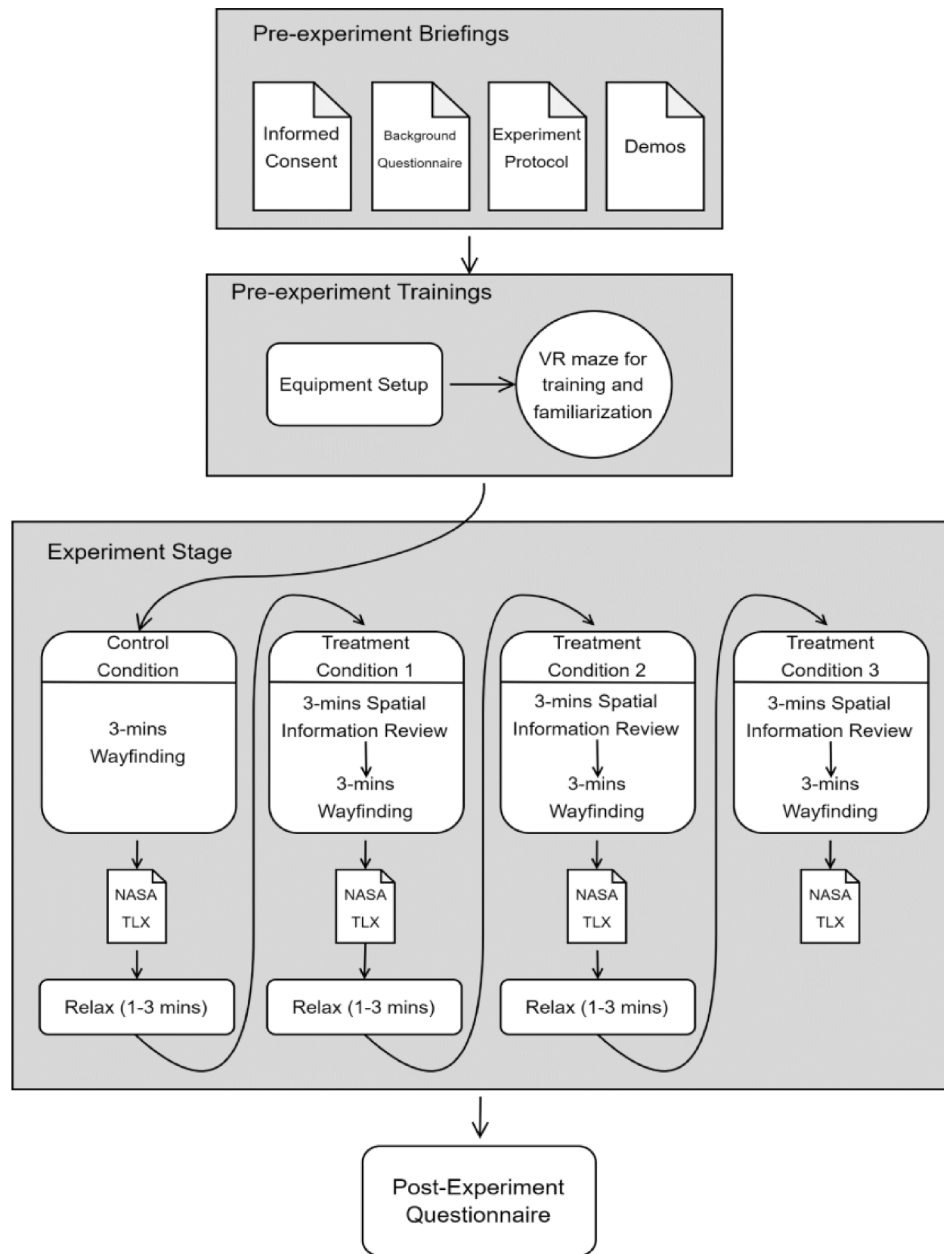


Fig. 2. Experiment Procedure.

participants' baseline spatial ability [45]. Participants with visual diseases or severe VR sickness were excluded from the experiment. We then went through the experiment protocol with the participants. Instructions and demonstrations on the VR headsets and navigation control methods were given to participants. After briefings, we set up the equipment for participants. We prepared a training scene for participants to get familiar with the VR environment and the navigation methods in the VR mazes. The training scene was also a VR maze (with a unique layout that was different from the maze layouts shown in Fig. 1) in which participants could practice navigation and wayfinding actions.

After confirming that participants were familiar with the experiment setup and that no additional familiarization burden existed, participants were guided to begin the experiment with the control condition. The control condition did not have the information review phase, simulating a case in which no prior knowledge of the building or victims could be provided. The participants were asked to navigate freely in the maze (free exploration) and find the three victims within 3 min. The control condition served as a performance baseline and further familiarized the

participants with the experiment task. Then each participant went through all the treatment conditions in the within-subject experiment. The participants were randomly assigned to experiment blocks in which review-and-perform experiments with a certain sequence were conducted. Each block represented a sequence of experiment conditions. There were three blocks: landmark-route-survey, survey-landmark-route, and route-survey-landmark. This incomplete block design can minimize the influence of sequencing [46].

The three types of spatial knowledge are usually leveraged by a person at the same time in wayfinding tasks, according to the original Siegel and White framework. Nonetheless, given the fundamental difference among the three types of spatial knowledge, distinct cognitive processes may be involved for each of them, leading to varying wayfinding performance. The scope of this research is to investigate the unique cognitive characteristics (e.g., cognitive load and decision-making) driven by each type of spatial knowledge. As a result, a controlled experiment should be ensured to rule out the combined effects of different types of spatial knowledge. The design of treatment

conditions aims to separate the landmark, route, and survey knowledge from each other and reduce the interference among them. Although these different formats of knowledge are entangled with each other in a typical spatial learning process, the existing literature shows the need for exploring the implications of each of them in a separate manner, such as Nys et al. [47], Kim and Bock [6]. Inspired by existing studies, we decoupled the spatial knowledge format as shown in Fig. 3. The landmark condition visually showed participants with key landmarks (chairs, shelves) that could guide them to the three victims. To simulate the real-world emergency search and rescue conditions, most landmarks must be approached close enough to be visible. The route condition instructed the optimal path that participants shall take to find victims. Although it appeared to be a textual information memorization task, the content provided by the route information was indeed related to the spatial guidance. The survey condition showed the map of the maze, the starting location, and the victims. Landmark pictures, route texts, and survey maps are distinct but effective ways to provide navigation guidance, although people may have largely different preferences. To understand the features of using different spatial knowledge in

firefighter wayfinding tasks, we intentionally simulated the low visibility conditions. Different types of information under different conditions stimulated the development of corresponding formats of spatial knowledge.

Each experiment condition consisted of two phases: the review phase and the performance phase. In the review phase, participants were given 3 min to review the information in the corresponding format, developing corresponding specific forms of spatial knowledge. The performance phase started immediately after the review phase, in which the participants were given a maximum time of 3 min to find the three victims in the virtual maze as accurately and quickly as possible. We chose three minutes as the review time because the pilot test showed that three minutes were adequate for all participants to memorize the given information. Similarly, three-minute tasks were used for the performance phase because our pilot data showed that it was the time needed to navigate through the entire maze. As such, the use of three minutes was deemed appropriate for both the review phase and the performance phase. After experiment tasks, participants were asked to finish a NASA Task Load Index (NASA TLX) [48] questionnaire. Then the participants

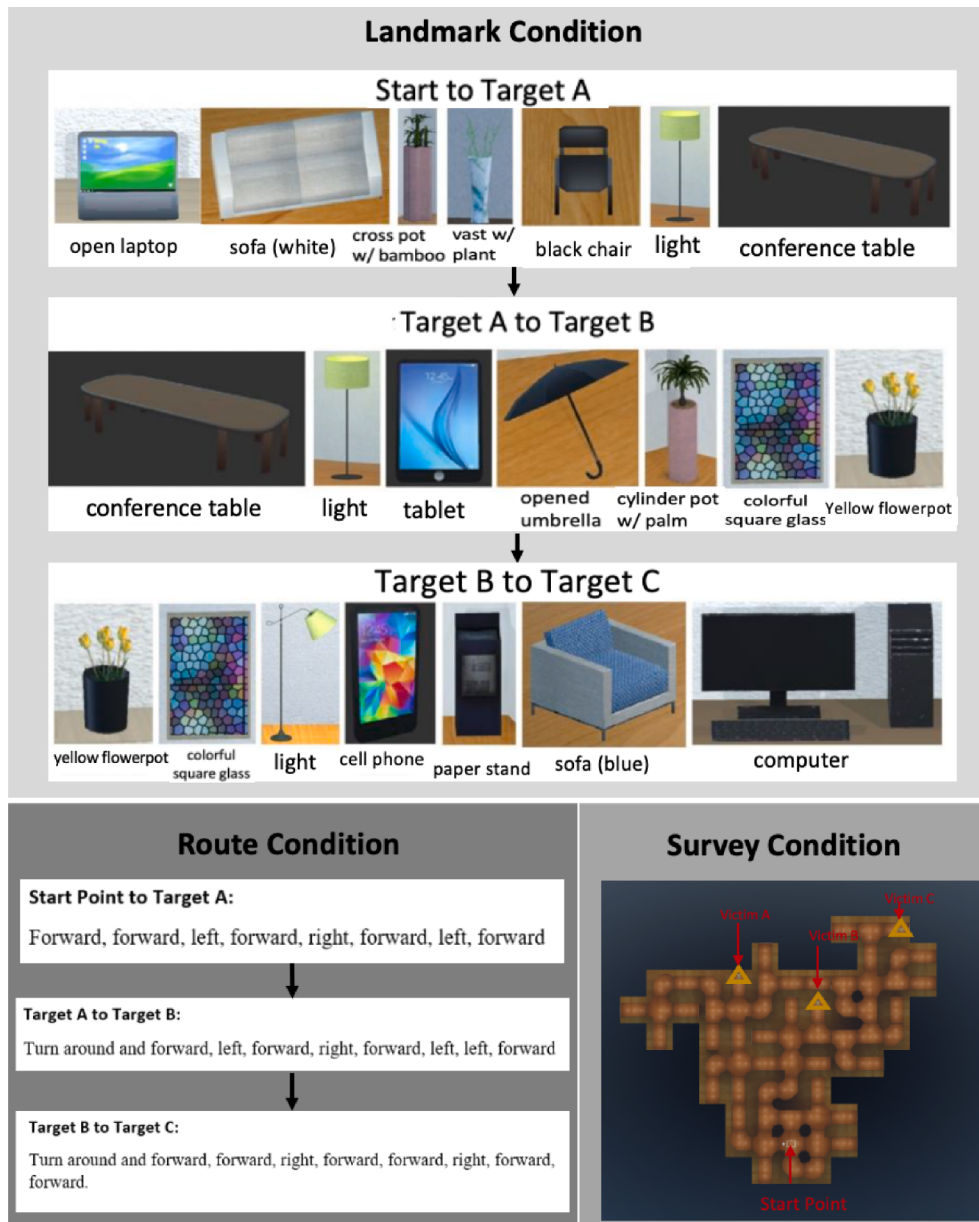


Fig. 3. The information reviewed by participants under different experiment conditions.

were given 2 min of break time before starting the review phase of the next experiment condition.

3.3. Data collection and evaluation

3.3.1. Task performance

Participants' position, orientation, and task completion time were recorded throughout the experiment performance phase. The core task was to find the victims in the shortest time but not to exceed the 3-minute time limit. We quantified the task performance by the number of victims found and the time spent. We coupled the two metrics by adopting the term 'wayfinding score' (Q). Wayfinding score is defined as the Eq. (1):

$$Q = \sum_{i=1}^3 \frac{T - t_i}{T} \quad (1)$$

where T is the fixed total time (3 min), t_i is the time spent for finding the victim i . If victim i has not been found, $t_i = T$. Using this evaluation method, the wayfinding task would have the Q value ranging between 0 (the worst performance, in which no victim was found) and 3 (the best performance, in which all the victims were found immediately). A higher Q value refers to a better wayfinding performance. This Q value was used as the main wayfinding performance indicator.

3.3.2. Cognitive load metrics

We measured the cognitive load with both subjective measures (NASA TLX) and objective measures (eye-tracking). NASA TLX questionnaire was filled once after each experiment condition to provide an overall subjective assessment of the performance and subjective evaluation of the mental load. The mental load index reported by the participants was collected for the post-experiment analysis. We scaled the NASA TLX index to a range of 0–100.

Tobii eye tracker [44] was integrated with an HTC VIVE headset to collect gaze and pupillary changes that indicated cognitive status. The Tobii eye tracker can record gaze focus and pupillary size at a frequency of 120 Hz and an estimated accuracy of 0.5°. Literature showed that gaze focus data can be used to calculate gaze stationary entropy [49], fixations [50], or saccades [51] to estimate the attention distribution patterns. The pupillary size was recorded throughout the information review phases and task performance phases at a frequency of 90 Hz. Although the frequency was lower than many pupillometry studies [52], it was high enough to capture pupillary diameter changes and blink activities [53] to indicate cognitive status. The higher frequency data could be useful if oscillating eye movements needed to be analyzed, which is out of the scope of this research [53].

The first set of metrics we extracted was the eye blink rate. Eye blink duration typically ranges from 50 ms to 300 ms [54]. The eye tracker in our experiment captured pupillary change at a frequency of 90 Hz which represented a blank interval of 11 ms and thus could accurately capture an eye blink event. The eye tracker recorded eye closure events as invalid pupillary size values. Thus, we filtered the invalid pupillary size data entries for both eyes to capture the eye closures. One blink activity was defined by a continuous period in which a spontaneous eye closure of both left and right eyes was detected. The blink rate was calculated by dividing the number of blinks by the time.

We also analyzed cognitive load by measuring the pupillary dilation. Extracting cognitive activities from pupillary size data needs several steps of data processing. The first denoise procedure is to remove light impacts. Light intensity is the dominant factor for pupillary size changes while cognitive changes impose relatively minor impacts [26]. As a result, we maintained the baseline luminance of the VR maze model consistent across all experiment conditions. Then, following a pupillary size correction method previously developed by us [31], we estimated the luminance changes induced by the VR headset and subtracted the pupillary size changes caused by light changes from the overall pupillary

size data. The light luminance received by both eyes was estimated from the screen color channels (red, green, and blue) value as given by [55], which was further used to estimate the pupillary size change induced by luminance [56]. Pupillary changes driven by cognitive activities were estimated by subtracting the light effects from overall pupillary changes. Then we followed a widely used pupillary size preparation method to calculate the residual pupillary changes purely due to cognitive status changes [18,23,28]. We performed a 1-min pupillary size baseline measurement session for each participant in which participants were sitting calmly. The average pupillary size during this session was used as the individual pupillary baseline. Then we extracted the pupillary dilation from the overall pupillary response. The widely recognized pupillary index, Percentage of dilation (PD) [18,26,31,33], was applied to indicate cognitive load. PD was computed following the Eq. (2):

$$PD = \frac{P_i - r}{r} \quad (2)$$

where P_i is the pupillary dilation at the time i , and r is the pupillary size baseline for this individual.

3.3.3. Decisional pattern analysis

In addition to aggregated measures that focus on the final outcome or performance, we also wanted to analyze the pattern changes at important discrete decision points during the task, e.g., turning points. As shown in Fig. 1, the randomly generated VR mazes had a high degree of freedom in navigation: rooms were connected in a complex way and could lead to many dead ends. Not all the rooms were equally challenging for the participants to find their destinations. Specifically, each room had two features that defined the level of difficulty in the wayfinding task: the number of entrances/exits, and whether it was on a correct path. We propose that the number of entrances and exits in a particular room can affect the level of difficulty of a wayfinding path decision. As a result, separate analyses should be performed for different levels of difficulty among the rooms. In a one-exit or two-exit room, a participant only had one choice: returning from the entrance in a one-exit room or proceeding to another exit in a two-exit room. In contrast, a participant needed to decide which exit to choose in a three-exit room or a four-exit room. So, a general assumption we made was that one-exit and two-exit rooms were not as relevant as three-exit or four-exit rooms in terms of decision-making. However, the assumption did not hold when a participant entered the wrong room, i.e., a room that was not on any path leading to the correct destinations and could only lead to dead ends (without victims). When in a wrong room, the participants would need to re-route the navigation path. The complexity of such a re-routing decision was affected by how many possible new paths the current room could lead to. When a room had more possible paths, such as with more entrances and exits, the decisional process was more difficult. Thus, we could categorize the cognitive processes in different rooms according to the different complexity levels of decision-making. We quantified the decisional complexity of a specific room with the number of entrances/exits. The decisional complexity also depended on whether the present room was on the correct path to the destination. When it was on the wrong path, the decisional complexity was also higher to move back to the correct path. Table 1 summarizes the decisional complexity level of rooms with different features. Trivial/Decisional room refers to the room with relatively low/high complexity of

Table 1

Room Decisional Importance according to number of entrances/exits and correctness.

Number of Entrances/Exits	Room Correctness	Decisional Complexity
1 or 2	Correct	Trivial Room
1 or 2	Wrong	Decisional Room
3 or 4	Correct	Decisional Room
3 or 4	Wrong	Decisional Room

decision-making processes. Categorizing the rooms according to decisional complexity gave us a better insight into participants' responses when facing spatial decisions. Although with different designs and layouts, the individual maze rooms were all made within a 10 m by 10 m square space. So, the navigation trajectories could be marked by the sequence of visited rooms for data processing.

We proposed an indicator for the decisional analysis: Decision-over-Trivial Index (DTI). DTI is defined as the mental effort increment when facing decisions, calculated by comparing the cognitive load at decision points (i.e., important waypoints for turning decisions) over at the trivial points (i.e., waypoints or straight lines that do not require any turning decisions). DTI is given in Equation (3):

$$DTI = \frac{\sum_{i=1}^n D_i}{n} - \frac{\sum_{j=1}^m D_j}{m} \quad (3)$$

where m and n are the numbers of trivial and decision points that a participant comes across in a trial, D_i is the percentage of pupillary dilation in room i . The overall PD calculation regards the complete experiment trial as a single data point. In contrast, the decision point analysis breaks the continuous process into discrete wayfinding decision points. To be specific, DTI is an index indicating cognitive allocation. DTI value indicates to what extent a participant spends (effective) mental efforts on making spatial decisions. In our analysis, DTI quantifies the aggregated cognitive load a participant spends only on making decisions for the important waypoints, e.g., when in a room with more entrances/exits, or when trying to move back to the right path to the destination. Because DTI distinguishes the relative cognitive load spending on important spatial decision-making from all cognitive load data, a higher DTI value might indicate that the participant concentrated more on making important wayfinding decisions, and thus is considered effective for the wayfinding performance. Thus, in the setting of our experiment, a higher DTI is desired as it means a more efficient mental effort allocation strategy.

4. Data analysis

4.1. Overview

We recruited 40 frontline firefighters to participate in our study from the Bryan Fire Department in Bryan, Texas. Out of all, eight firefighters reported severe VR motion sickness and were unable to complete the experiment. As a result, data from 32 effective participants were used for the analysis. The remaining 32 effective participants were all healthy adult males, with a mean age of 31 years (ranging from 23 years old to 41 years old). Their working experience ranged from 8 months to 17 years, with a mean working experience of 7 years. No previous VR experience was reported from any of the participants.

4.2. Wayfinding performance

Our experiment was a simple repeated measures design, with the treatments being different spatial knowledge forms during the review phase. The impact of the different orders of conditions was eliminated by treatments on multiple levels: eliminating familiarization burden (pre-training sessions), eliminating residual effect (idling time between treatments), and randomizing experiment sequence orders. We performed a D'Agostino and Pearson's test [57] to examine the normality of the wayfinding score (and also the same test for other normality tests performed in this study). Threshold confidence interval was selected as 95% ($p = 0.05$) hereafter. The normality test result rejected the null hypothesis that the wayfinding scores were normally distributed (with $p < 0.001$) which failed to satisfy the pre-assumption of Analysis of variance (ANOVA). Thus, the Friedman test, a non-parametric substitution of repeated measures comparable with ANOVA, was performed to analyze the statistical difference among different knowledge formats.

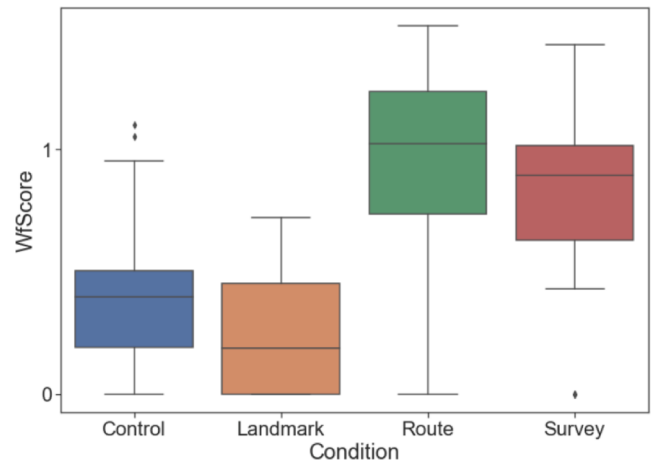


Fig. 4. Wayfinding Score under Different Conditions.

Fig. 4 shows the wayfinding performance results with different experimental conditions.

Friedman's test showed that the spatial knowledge format had a significant impact on the final task performance ($p < 0.001$). We performed Wilcoxon signed rank tests to analyze the pairwise differences. The result showed that the control group (given no prior knowledge of the maze spatial layouts) was significantly different from all other conditions ($p < 0.05$, 0.001, and 0.001 when compared with landmark, route, and survey respectively). Landmark group was significantly different from route group ($p < 0.001$) and survey group ($p < 0.001$). But building spatial knowledge developed from route information had no significant difference compared with survey information ($p = 0.304$) in terms of the final wayfinding score. The results indicated that the knowledge reviewing session was helpful for firefighters to perform wayfinding tasks in general, and firefighters performed better when reviewing route and survey knowledge than reviewing landmark knowledge. Cube test results showed no significant relationship between participants' baseline spatial abilities and their wayfinding performances.

4.3. Subjective workload evaluation

We performed a NASA TLX questionnaire survey to evaluate the perceived task workload after finishing the wayfinding task in each condition. We followed a standard NASA TLX data collection and processing pipeline [58] to calculate the perceived workload level of the wayfinding task under different conditions. The result is shown in Fig. 5.

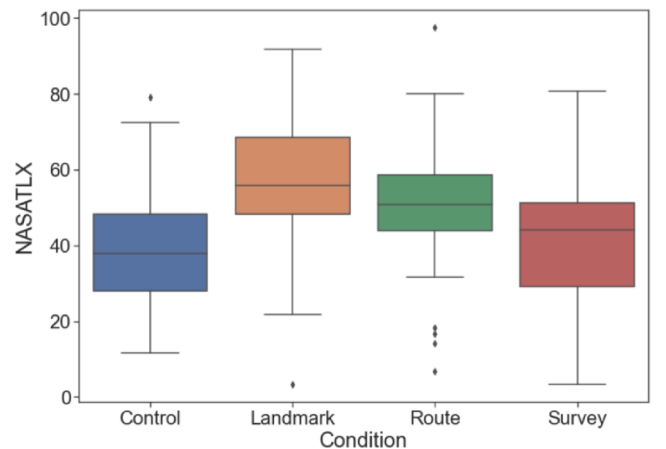


Fig. 5. Subjective Workload (NASA TLX) in Different Conditions.

Similar to the patterns shown in the wayfinding score, the results were not normally distributed. Thus, we followed the same statistical analysis pipeline as in the wayfinding score analysis. The results showed that the perceived workload was significantly different between landmark condition and route condition ($p = 0.014$), between landmark and survey condition ($p < 0.001$), between route and survey condition ($p < 0.001$), between control and landmark condition ($p < 0.001$), and between control and route condition ($p < 0.001$). No significant difference was found between the control and survey conditions ($p = 0.888$). Subjective workload results indicated that firefighters generally reported the highest workload level when reviewing the information in landmark format and the lowest workload level when reviewing the information in survey format or review no information (control group).

4.4. Eye blink rate for measuring fatigue

As discussed earlier, eye blink rate (BR) can be used as an indicator of fatigue. We calculated BR according to the number of eye blinks across the total experiment time span (3 min). Data analysis showed that the BR distribution was non-normal and was significantly different across four conditions ($p < 0.001$). The pairwise test showed that all pairwise combinations were significantly different except between the control condition and survey condition ($p = 0.571$). The p -values were 0.001 between control and landmark, 0.0061 between control and route, 0.009 between landmark and route, 0.001 between landmark and survey, and 0.035 between route and survey. Fig. 6 shows the blink rate results. The blink rate pattern implicated that the control and survey conditions led to the lowest fatigue, while landmarks caused the highest fatigue level.

4.5. Pupil dilation for measuring cognitive load

4.5.1. Average Percentage of pupillary dilation (PD)

To further explore the impact of different conditions on the cognitive process, we collected and analyzed the pupillary size data. After performing the data cleaning as described in the methodology, we calculated the average Percentage of Dilation (PD) under different conditions, as shown in Fig. 7. The results showed that PD was not normally distributed and was significantly impacted by the form of spatial knowledge provided ($p < 0.001$). Similar to the wayfinding performance score result, the pairwise test showed that participants' PD was significantly different between the control condition and all other conditions ($p = 0.027$, $p < 0.001$, $p < 0.001$ for landmark, route, and survey respectively). A significant difference was found between landmark and route conditions ($p < 0.001$), and between landmark and survey conditions ($p < 0.001$). No significant difference was found between route and survey conditions ($p = 0.561$). The result indicated that the control

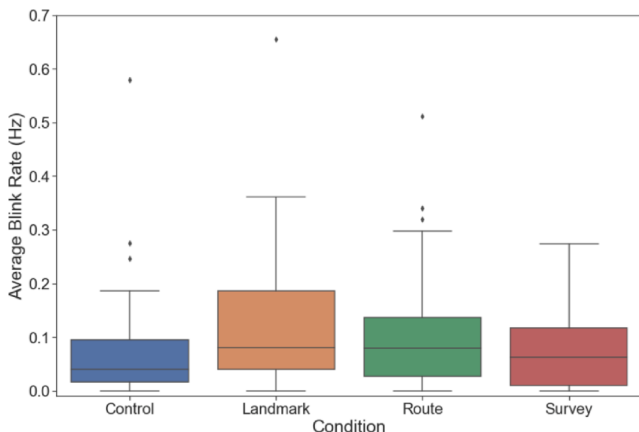


Fig. 6. Average Blink Rate in Different Conditions.

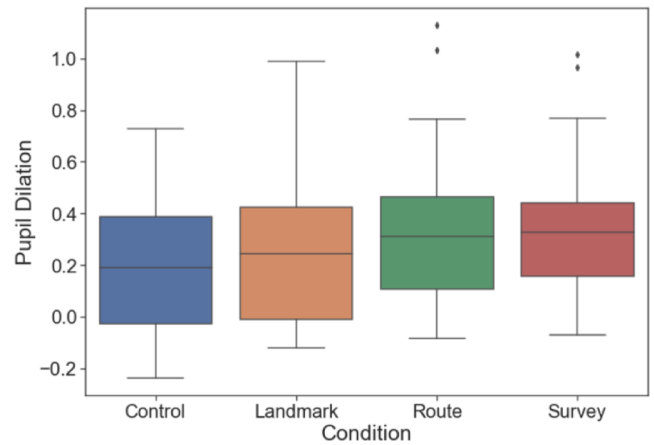


Fig. 7. Overall Average Percentage of Pupil Dilation.

condition imposed the least mental load, while the route and survey conditions imposed a relatively higher mental load.

4.5.2. Variation of pupillary dilation (PD Variation)

Considering that our study involved a continuous decisional process, we also wanted to analyze the pupillary response pattern over time, which could be measured as the variation when performing the task. Fig. 8 is an example that shows the PD curves of a randomly selected participant. It showed that the variability of PD was quite different among different conditions for the same person. Specifically, under route and survey conditions, not only the overall levels of PD increased, but also the ranges between the maximum and minimum PD values were also bigger. It indicated that this participant experienced a more drastic change in terms of cognitive load under route and survey conditions, such as switching from a low level of mental load to using a lot of mental loads in a very short period of time. We compared the variance of PD among different conditions. A higher variance means that PD changes more drastically from the mean, implicating a higher magnitude of pupillary response to stimuli (e.g., when trying to memorize the layout or making an important turning decision). Fig. 9 shows the variance of PD under different conditions. The statistical result indicated that the control condition showed no significant difference from the landmark condition ($p = 0.078$), and the route condition had no significant difference from the survey condition ($p = 0.073$). All other pairs were significantly different. The results suggested that not only the average pupillary dilation was different, but also the magnitude of pupillary response was more drastic in route and survey conditions compared to control and landmark condition.

4.5.3. Decision-over-Trivial index (DTI) Analysis

To further examine the spatial variation patterns, we performed the Decision-over-Trivial Index (DTI) analysis for pupillary dilation. The result is shown in Fig. 10. The data was not normally distributed. Wilcoxon test showed that the control condition had no significant difference from the landmark condition ($p = 0.445$). The p -value between the control condition and route group was 0.055, which was not significant but still very small. The control condition was significantly different from the survey condition ($p < 0.001$). All the treatment conditions were significantly different: between landmark and route ($p = 0.036$), between landmark and survey ($p < 0.001$), and between route and survey ($p = 0.028$). This result suggested that when reviewing information in survey format, firefighters showed the highest DTI while reviewing in landmark format led to the lowest DTI.

5. Discussion

The experiment revealed some potential behavioral and cognitive

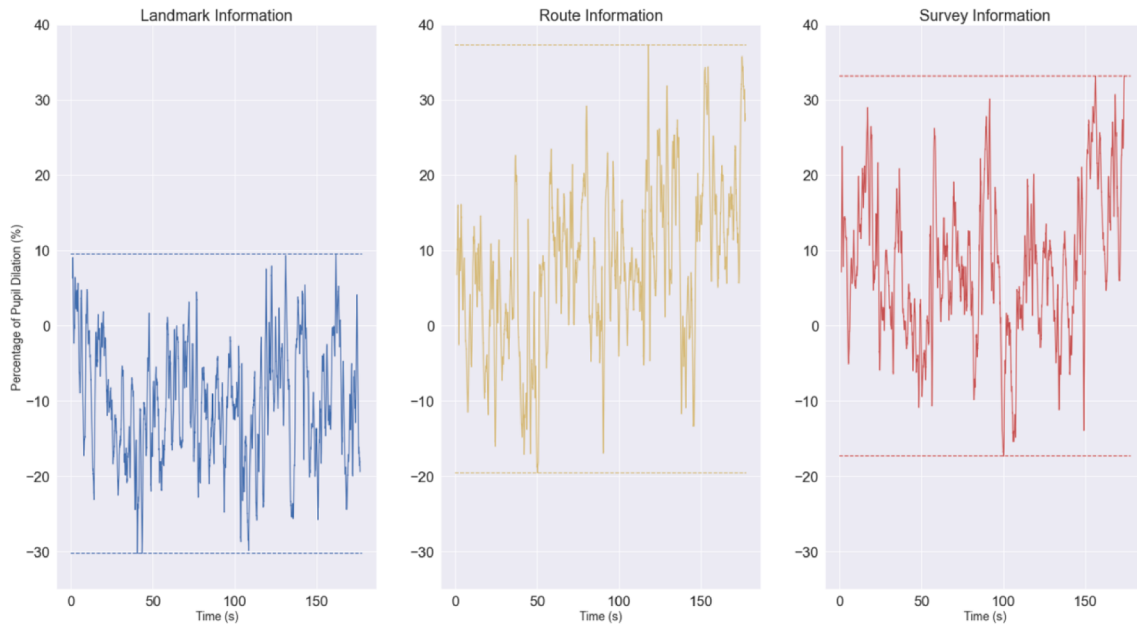


Fig. 8. Percentage of Pupil Dilation during the Experiment from a Randomly Selected Participant.

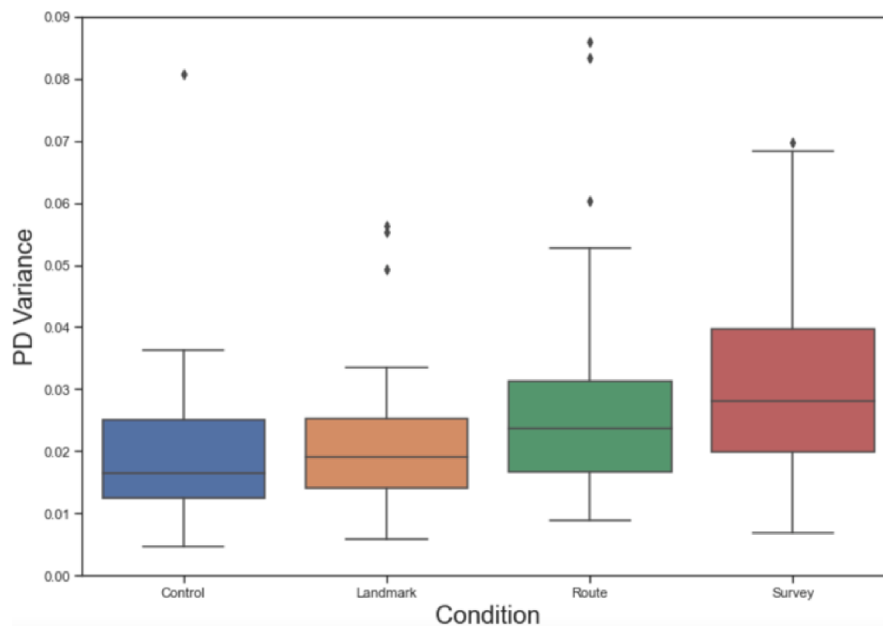


Fig. 9. Variance of Percentage of Pupil Dilation.

implications after reviewing different forms of spatial information. The results suggested that reviewing building layout information in route and survey formats could contribute to more effective use of spatial cognition. Route and survey-based spatial knowledge helped participants understand the space better and thus led to a better wayfinding performance (finding more victims in a shorter time).

In addition to task performance, we were also concerned about the cognitive process that may have caused the difference in wayfinding performance. We estimated cognitive load with the percentage of pupil dilation. The results showed that survey and route conditions were associated with a higher cognitive load. This is an interesting finding as most literature would consider a higher cognitive load to be related to worse performance. We found that the cognitive load in this wayfinding task should be considered a neutral measure, which was related to the mental efforts that participants were spending on making critical

wayfinding decisions. In other words, given the route and survey formats, a participant generally spent more mental effort (thinking harder) when performing the wayfinding task, which led to better performance.

To validate this interpretation, we analyzed the spatial distribution features of the PD in all experiment trials. An interesting finding was that PD variability was significantly different among all conditions. PD variance described the spatial distribution of PD across the continuous task. The source of PD variance was the participant's mental response to stimuli. More specifically speaking, in this continuous cognitive-driven wayfinding task, if a participant was constantly thinking under a consistent level of mental load, the PD variance should be small. In contrast, if the participant was thinking harder under some circumstances and was more relaxed under other circumstances, the PD variance would be larger. Thus, the PD variance analysis showed that when facing stimuli, such as making important wayfinding decisions, route

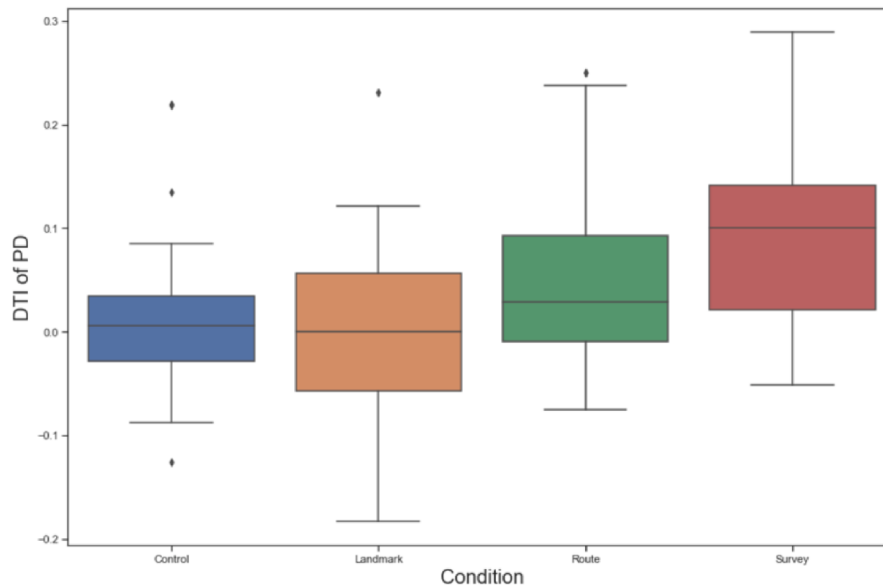


Fig. 10. Decision-over-Trivial Index of Percentage of Pupil Dilation.

and survey conditions imposed a higher magnitude of impact on participants while control and landmark conditions imposed a lower impact. To find out whether wayfinding decision-making was the stimuli (or part of the stimulus) that caused the drastic increase of cognitive load for the route and survey conditions, we adopted a decisional feature analysis. As described in the methodology, we divided the continuous process into discrete steps according to the maze rooms. We adopted a parameter, Decisional-over-Trivial Index (DTI), to analyze the PD difference between decisional points and trivial points. The result showed that DTI was higher in the survey condition, followed by the route condition, and least in the landmark and control conditions. This result echoed the variance analysis findings, in a sense that survey and route conditions showed a higher PD variance when making important wayfinding decisions. It may suggest that after reviewing spatial information in the forms of survey and route, participants tended to be more relaxed when no decisions were needed to be made. When facing decision-making points, the mental load raised much faster and higher accordingly to solve the decision and spatial memory retrieval problem. DTI values also echoed the subjective workload estimates (NASA TLX) and fatigue estimates (blink rate). It indicates that mental effort allocation strategy may also be related to the accumulation of fatigue, i.e., when people are more relaxed at trivial points, fatigue is less likely to happen.

As mentioned, we also examined the perceived workload under different conditions. NASA TLX results indicated that participants generally perceived the lowest workload in the survey condition, followed by the route conditions. Landmark caused the highest perceived workload. The objective fatigue index, eye blink rate, was consistent with this result. Subjective and objective evidence both agreed that it led to the least mental fatigue to perform the VR maze wayfinding using the survey information, which implied that the survey knowledge might be easier to perceive and retrieve. In contrast, relying on the landmark information caused a higher level of mental fatigue. Our interpretation of this result is that relying on landmarks for navigation in an emergency is naturally difficult in two ways. First, firefighters need to convert visuospatial information (i.e., the images of the landmark objects) into phonological information in their cognitive process (such as their semantics). It requires additional time and causes an additional cognitive burden to process the information during the task. Second, our experiment focuses on the emergency wayfinding where visibility is often affected. In our VR environment, visibility was lowered to simulate real-world emergency wayfinding conditions. Participating firefighters must go close enough to an object to be able to recognize it. It indeed

increased the level of difficulty of relying on landmark information. We designed this condition to capture the real-world challenges firefighters would likely face.

Compared with the cognitive load result indicated by PD, the TLX and eye blink rate results showed that route and survey knowledge led to a higher cognitive load but a lower perceived workload and lower accumulated fatigue. This finding inferred that perceived workload and fatigue may not be directly associated with cognitive load. The results echo previous findings [59–61] that fatigue shares a different neural path with mental load and may change differently with the mental load. Thus, it is likely that the cognitive load and fatigue assessments are measuring different aspects or processes during the wayfinding decision-making. For instance, the cognitive load may be related to measuring the effort that a participant uses to retrieve spatial knowledge, while fatigue and subjective workload measure the difficulty level of mapping the mental spatial knowledge with the reality. In addition, it is also possible that the cognitive load during this experiment was below the threshold of cognitive overload, and thus the increased cognitive load measures were more of a positive indicator of the effective use of mental effort in making wayfinding decisions, rather than a negative indicator of the mental fatigue.

Apart from the treatment groups, another interesting finding from the workload and fatigue analysis was that, if no prior information about the maze spatial layout was given (control condition), participants reported similar workload and fatigue levels as those reviewing survey information. This makes sense because when having no prior knowledge about the maze layout, a participant may randomly choose a navigation strategy without worrying about retrieving memories, and thus is not as tired as under the landmark condition. This partially answers why the control condition showed a slightly better performance than the landmark condition.

This study decoupled landmark, route, and survey to observe individual patterns. However, the decoupling method can be potentially improved. In addition, combining the information in a structured way might be more intuitive, practical, and effective for wayfinding tasks in real life. It is the future agenda to determine the effective method of spatial information delivery by fine-tuning the combinations of different formats of spatial information, for instance, marking the route on a survey map, and showing directions based on landmark objects. Meanwhile, this study only recruited male firefighters for participation due to the limited number of female frontline firefighters. Thus, whether gender difference could play a role in the spatial cognition process

remains a question unanswered. In addition, this study aimed to explore the spatial cognitive process in an emergency wayfinding task, thus only well-trained firefighters were recruited. Whether our findings can be extended to general wayfinding tasks still needs further validation. It is suggested that in future wayfinding studies, participants with more diverse backgrounds can be recruited.

6. Conclusion

This study investigated different spatial knowledge acquisition strategies prior to wayfinding in the context of emergency fire rescue. Especially, the behavioral and cognitive implications using different forms of spatial information were analyzed. A total of 32 frontline firefighters completed this study. We found that when acquiring the spatial knowledge by utilizing the route and survey (map) information, participants tended to achieve a better wayfinding performance. We further elaborated the findings by looking into the corresponding cognitive characteristics. We found that although route and survey knowledge could lead to a higher cognitive load in general, a lower workload was perceived, and less fatigue was detected when performing the task. This pattern indicated two possibilities: cognitive load may be a neutral indicator than fatigue and perceived workload indicators, and cognitive overload did not occur in our experiment.

We also found an interesting pattern that the survey-driven spatial cognitive process led to a higher variance in mental effort distribution (spatial distribution), in which a lower cognitive load was reported at trivial points and a much higher cognitive load was reported at critical decisional points. In contrast, the landmark-driven spatial cognitive process led to a more evenly distributed mental effort: participants' mental effort increments at decisional points were not as high as in the other conditions. This may indicate a difference in attention allocation strategies under different spatial knowledge acquisition strategies.

In conclusion, this paper analyzes and discusses several behavioral and cognitive implications of firefighter wayfinding tasks under different spatial knowledge acquisition strategies. It is expected to contribute to a better understanding of the performance in emergency wayfinding tasks and the cognitive causes thus contributing to the design of effective spatial knowledge delivery methods. Furthermore, the findings reveal a series of cognitive features related to a more effective spatial decision-making in search and rescue. In the future, we can track the pupillary pattern and other ergonomics features with portable devices to provide real-time measures of users' cognitive status, as well as estimate potential impacts on task performance. If the cognitive overload is observed, commands on retreat, additional help, or cover can be made. Thus, this study might inspire a real-time monitoring and prediction system for possible cognitive or performance issues in the emergency wayfinding.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank all the firefighters for their valuable time to participate in this study. This material is supported by the National Science Foundation (NSF) under Grant 1937878, as well as the National Institute of Standards and Technology (NIST) under Grant 60NANB18D152. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not reflect the views of the NSF and NIST.

References

- [1] D.L. Smith, Firefighter fitness: improving performance and preventing injuries and fatalities, *Curr. Sports Med. Rep.* 10 (3) (2011) 167–172, <https://doi.org/10.1249/JSR.0b013e31821a9fec>.
- [2] U.S. Fire Administration (2020). Statistical Reports on the U.S. Fire Problem. [Online] Available: <https://www.usfa.fema.gov/data/statistics/reports/>.
- [3] A.W. Siegel, S.H. White, The Development of Spatial Representations of Large-Scale Environments, in: *Advances in Child Development and Behavior*, vol. 10, H. W. Reese Ed.: JAI, 1975, pp. 9–55.
- [4] B. Lapeyre, S. Hourlier, X. Servantie, B. N'Kaoua, H. Sauzéon, Using the landmark–route–survey framework to evaluate spatial knowledge obtained from synthetic vision systems, *Hum. Factors* 53 (6) (2011) 647–661, <https://doi.org/10.1177/0018720811421171>.
- [5] S. Werner, B. Krieg-Brückner, H. A. Mallot, K. Schweizer, C. Freksa, Spatial Cognition: The Role of Landmark, Route, and Survey Knowledge in Human and Robot Navigation1, Berlin, Heidelberg, 1997: Springer Berlin Heidelberg, in *Informatik '97 Informatik als Innovationsmotor*, pp. 41–50.
- [6] K. Kim, O. Bock, Acquisition of landmark, route, and survey knowledge in a wayfinding task: in stages or in parallel? *Psychol. Res.* 85 (5) (2021) 2098–2106, <https://doi.org/10.1007/s00426-020-01384-3>.
- [7] A. Schwering, J. Krukar, M. Galvao, W. Schick, H. Löwen, Wayfinding through orientation: schematizing landmark, route and survey information in a single map, *Spatial Cognit. Computation* 17 (4) (2017) 1–31, <https://doi.org/10.1080/13875868.2017.1322597>.
- [8] P.U. Lee, B. Tversky, Interplay between visual and spatial: the effect of landmark descriptions on comprehension of route/survey spatial descriptions, *Spatial Cognit. Computation* 5 (2–3) (2005) 163–185, <https://doi.org/10.1080/13875868.2005.9683802>.
- [9] T.T. Brunyé, H.A. Taylor, Extended experience benefits spatial mental model development with route but not survey descriptions, *Acta Psychol.* 127 (2) (2008) 340–354.
- [10] R. Nori, S. Grandicelli, F. Giusberti, Alignment effect: primary–secondary learning and cognitive styles, *Perception* 35 (9) (2006) 1233–1249.
- [11] E.R. Chastil, Neural evidence supports a novel framework for spatial navigation, *Psychon. Bull. Rev.* 20 (2) (2013) 208–227.
- [12] J. Sluzenski, N.S. Newcombe, E. Satlow, Knowing where things are in the second year of life: implications for hippocampal development, *J. Cognit. Neurosci.* 16 (8) (2004) 1443–1451.
- [13] D.R. Montello, A new framework for understanding the acquisition of spatial knowledge in large-scale environments, in: *Spatial and temporal reasoning in geographic information systems*, pp. 143–154, 1998.
- [14] D. Caduff, S. Timpf, On the assessment of landmark salience for human navigation, *Cogn. Process.* 9 (4) (2008) 249–267.
- [15] R. Nori, F. Giusberti, Cognitive styles: errors in directional judgments, *Perception* 32 (3) (2003) 307–320.
- [16] R. Epstein, N. Kanwisher, A cortical representation of the local visual environment, *Nature* 392 (6676) (1998), 598–601, <https://doi.org/10.1038/33402>.
- [17] E.A. Maguire, R.S. Frackowiak, C.D. Frith, Recalling routes around London: activation of the right hippocampus in taxi drivers, *J. Neurosci.* 17 (18) (1997) 7103–7110.
- [18] R. Gavas, D. Chatterjee, A. Sinha, Estimation of cognitive load based on the pupil size dilation, in: *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2017: IEEE, pp. 1499–1504.
- [19] R. Gavas, R. Das, P. Das, D. Chatterjee, A. Sinha, Inactive-state recognition from EEG signals and its application in cognitive load computation, in: *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016: IEEE, pp. 003606–003611.
- [20] S.P. Marshall, Identifying cognitive state from eye metrics, *Aviat. Space Environ. Med.* 78 (5) (2007) B165–B175.
- [21] K.S. Chiew, T.S. Braver, Temporal dynamics of motivation-cognitive control interactions revealed by high-resolution pupillometry, *Front. Psychol.* 4 (2013) 15.
- [22] K.S. Chiew, T.S. Braver, Dissociable influences of reward motivation and positive emotion on cognitive control, *Cogn. Affective Behav. Neurosci.* 14 (2) (2014) 509–529.
- [23] V. Peysakhovich, F. Vachon, F. Dehais, The impact of luminance on tonic and phasic pupillary responses to sustained cognitive load, *Int. J. Psychophysiol.* 112 (2017) 40–45.
- [24] A.D. Souchet, S. Philippe, D. Lourdeaux, L. Leroy, Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: a review, *Int. J. Hum.-Computer Interaction* 38 (9) (2022) 801–824, <https://doi.org/10.1080/10447318.2021.1976509>.
- [25] M.K. Eckstein, B. Guerra-Carrillo, A.T. Miller Singley, S.A. Bunge, Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cogn. Neurosci.* 25 (2017) 69–91.
- [26] P. van der Wel, H. van Steenbergen, Pupil dilation as an index of effort in cognitive control tasks: a review, *Psychon. Bull. Rev.* 25 (6) (2018) 2005–2015.
- [27] T. Verguts, E. Vassena, M. Silvetti, Adaptive effort investment in cognitive and physical tasks: a neurocomputational model, *Front. Behav. Neurosci.* 9 (2015) 57.
- [28] O. Palinko, A.L. Kun, A. Shyrovok, P. Heeman, Estimating cognitive load using remote eye tracking in a driving simulator, in: *Proceedings of the 2010 symposium on eye-tracking research & applications*, 2010, pp. 141–144.
- [29] A.A. Zekveld, S.E. Kramer, J.M. Festen, Cognitive load during speech perception in noise: the influence of age, hearing loss, and cognition on the pupil response, *Ear Hear.* 32 (4) (2011) 498–510.

- [30] S. Mathôt, J. Fabius, E. Van Heusden, S. Van der Stigchel, Safe and sensible preprocessing and baseline correction of pupil-size data, *Behav. Res. Methods* 50 (1) (2018) 94–106.
- [31] Y. Shi, J. Du, D.A. Worthy, The impact of engineering information formats on learning and execution of construction operations: a virtual reality pipe maintenance experiment, *Autom. Constr.* 119 (2020), 103367.
- [32] P. Kiefer, I. Giannopoulos, A. Duchowski, M. Raubal, Measuring cognitive load for map tasks through pupil diameter, in: *The Annual International Conference on Geographic Information Science*, Springer, 2016, pp. 323–337.
- [33] D.T. Payne, M.E. Parry, S.J. Harasymiw, Percentage of pupillary dilation as a measure of item difficulty, *Perception Psychophys.* 4 (3) (1968) 139–143.
- [34] S.P. Marshall, The index of cognitive activity: Measuring cognitive workload, in: *Proceedings of the IEEE 7th conference on Human Factors and Power Plants*, 2002: IEEE, pp. 7–7.
- [35] A.T. Duchowski, et al., The index of pupillary activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation, in: *Proceedings of the 2018 CHI conference on human factors in computing systems*, 2018, pp. 1–13.
- [36] D. Hoppe, S. Helfmann, C.A. Rothkopf, Humans quickly learn to blink strategically in response to environmental task demands, *Proc. Natl. Acad. Sci.* 115 (9) (2018) 2246–2251.
- [37] J.A. Stern, D. Boyer, D. Schroeder, Blink rate: a possible measure of fatigue, *Hum. Factors* 36 (2) (1994) 285–297.
- [38] R. Martins, J. Carvalho, Eye blinking as an indicator of fatigue and mental load—a systematic review, *Occupational Saf. Hygiene* 10 (2015) 231–235.
- [39] G.J. Siegle, N. Ichikawa, S. Steinhauer, Blink before and after you think: blinks occur prior to and following cognitive load indexed by pupillary responses, *Psychophysiology* 45 (5) (2008) 679–687.
- [40] M. Pedrotti, M.A. Mirzaei, A. Tedesco, J.-R. Chardonnet, F. Mérienne, S. Benedetto, T. Baccino, Automatic stress classification with pupil diameter analysis, *Int. J. Hum.-Computer Interaction* 30 (3) (2014) 220–236.
- [41] G.R. Loftus, M.E.J. Masson, Using confidence intervals in within-subject designs, *Psychon. Bull. Rev.* 1 (4) (1994) 476–490.
- [42] Unity 3D. “Maze Generator.” <https://assetstore.unity.com/packages/tools/modeling/maze-generator-41853> (accessed 2022).
- [43] S. Rebay, Efficient Unstructured Mesh Generation by Means of Delaunay Triangulation and Bowyer-Watson Algorithm, *J. Comput. Phys.* 106 (1) (1993/05/01/ 1993,) 125–138, <https://doi.org/10.1006/jcph.1993.1097>.
- [44] TobiiPro. “Tobii Pro VR Integration.” <https://www.tobiiipro.com/product-listing/vr-integration/> (accessed 2022).
- [45] G.B. Dadi, P.M. Goodrum, T.R. Taylor, C.M. Carswell, Cognitive workload demands using 2D and 3D spatial engineering information formats, *J. Construct. Eng. Manage.* 140 (5) (2014) 04014001.
- [46] P.W.M. John, An application of a balanced incomplete block design, *Technometrics* 3 (1) (1961) 51–54.
- [47] M. Nys, V. Gyselinck, E. Orriols, M. Hickmann, Landmark and route knowledge in children’s spatial representation of a virtual environment, *Front. Psychol.* 5 (2015), <https://doi.org/10.3389/fpsyg.2014.01522>.
- [48] S.G. Hart, NASA-task load index (NASA-TLX); 20 years later, in: *Proceedings of the human factors and ergonomics society annual meeting*, 2006, vol. 50, no. 9: Sage publications Sage CA: Los Angeles, CA, pp. 904–908.
- [49] B.A. Shiferaw, L.A. Downey, J. Westlake, B. Stevens, S.M.W. Rajaratnam, D. J. Berlowitz, P. Swann, M.E. Howard, Stationary gaze entropy predicts lane departure events in sleep-deprived drivers, *Sci. Rep.* 8 (1) (2018), <https://doi.org/10.1038/s41598-018-20588-7>.
- [50] P. Jiménez, L.M. Bergasa, J. Nuevo, N. Hernández, I.G. Daza, Gaze fixation system for the evaluation of driver distractions induced by IVIS, *IEEE Trans. Intell. Transp. Syst.* 13 (3) (2012) 1167–1178.
- [51] R. Tomlinson, Combined eye-head gaze shifts in the primate. III. Contributions to the accuracy of gaze saccades, *J. Neurophysiol.* 64 (6) (1990) 1873–1891.
- [52] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer, *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford, 2011.
- [53] R. Andersson, M. Nyström, K. Holmqvist, Sampling frequency and eye-tracking measures: how speed affects durations, latencies, and more, *J. Eye Movement Res.* 3 (3) (2010).
- [54] Y. Wang, S.S. Toor, R. Gautam, D.B. Henson, Blink Frequency and Duration during Perimetry and Their Relationship to Test-Retest Threshold Variability, *Invest. Ophthalmol. Vis. Sci.* 52 (7) (2011) 4546–4550, <https://doi.org/10.1167/iovs.10-6553>.
- [55] W.C.W. Draft. “Working Draft-techniques for Accessibility Evaluation And Repair Tools.” <https://www.w3.org/TR/AERT/> (accessed 2022).
- [56] C.A. Blackie, H.C. Howland, An extension of an accommodation and convergence model of emmetropization to include the effects of illumination intensity, *Ophthalmic Physiol. Opt.* 19 (2) (1999) 112–125.
- [57] R. DiAgostino, An omnibus test of normality for moderate and large sample sizes, *Biometrika* 58 (34) (1971) 1–348.
- [58] S.G. Hart, L.E. Staveland, Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, *Adv. Psychol.* 52 (1988) 139–183.
- [59] J. DeLuca, Fatigue, cognition, and mental effort, *Fatigue Window Brain* 37 (2005).
- [60] M. Milyavskaya, B. Galla, M. Inzlicht, A. Duckworth, More effort, less fatigue: How interest increases effort and reduces mental fatigue, 2018.
- [61] B.W. Hornsby, The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands, *Ear Hear.* 34 (5) (2013) 523–534.